

CHALMERS



GÖTEBORGS UNIVERSITET

Modellering av nervmönster med spatiala punktprocesser

*Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Kandidatarbete inom civilingenjörsutbildningen vid Chalmers*

Christian Källgren
Hampus Lane
Simon Eriksson

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2017

Modellering av nervmönster med spatiala punktprocesser

Examensarbete för kandidatexamen i matematisk statistik inom matematikprogrammet vid Göteborgs universitet

Christian Källgren
Simon Eriksson

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers

Hampus Lane

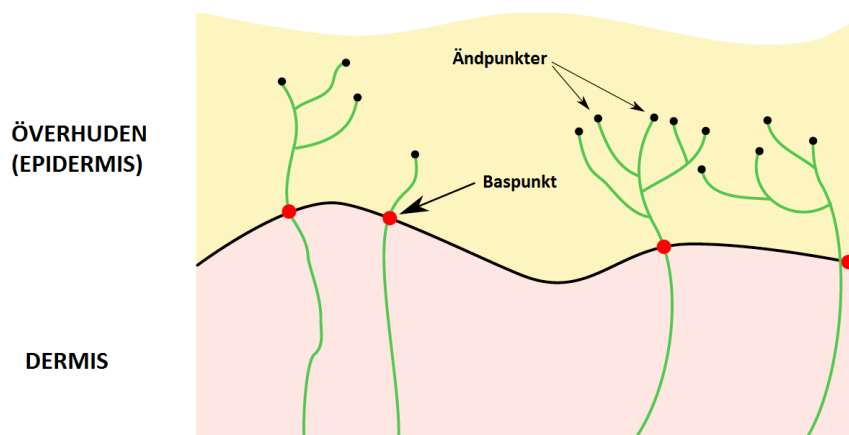
Handledare: Aila Särkkä & Claes Andersson
Examinatorer: Maria Roginskaya & Marina Axelson-Fisk

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2017

Populärvetenskaplig presentation

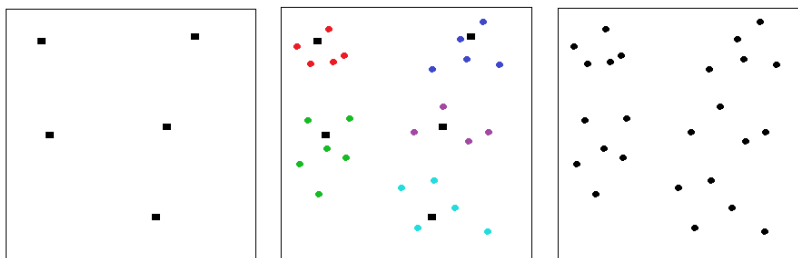
Går det att förutspå en nervsjukdom?

Diabetespatienter riskerar att utveckla en nervsjukdom som kallas diabetesneuropati som bland annat kan orsaka känselbortfall och smärtor. Känseln påverkas av hur tätt nervändarna ligger i huden. Utifrån detta är det intressant att studera det mönster som bildas av nervändarna. Nervmönstret bildas genom att ett antal nervtrådar tränger in i överhuden, var och en av dessa förgrenar sig sedan och ger upphov till många ändpunkter ovanför, se Figur 1.



Figur 1: (Claes Andersson, 2017) Bilden ovan visar ett tvärsnitt av hur nervtrådar tränger in i överhuden och förgrenar sig i ändpunkter.

För att kunna beskriva nervmönstren behövs en modell. Ett exempel på en sådan modell som kan återskapa hur nervändarna är fördelade i huden är en s.k. klustrad spatial punktprocess. Modellen går ut på att man slumpar fram ett antal centrum för att sedan vid varje centrum slumpa fram ett antal ändpunkter nära varandra. Denna modell fungerar bra då gruppcentrum agerar som baspunkt och alla de färgade punkterna inom samma grupp visar ändpunkterna (nervändarna) som kommer från samma nerv. Detta illustreras i mittenbilden av Figur 2, där nervändarna från samma nerv har samma färg. För en sådan modell behöver man uppskatta hur många baspunkter mönstret har (parameter 1), hur många nervändar det finns i varje grupp (parameter 2) och hur dessa placerar sig i relation till baspunkterna (parameter 3). Det är dessa parametrar man vill se om de är olika hos patienter med eller utan diabetesneuropati.



Figur 2: I bilden till vänster kan man se hur gruppernas centrum är placerade helt slumpmässigt, dessa representerar var nervtrådarna kommer in i överhuden. I bilden i mitten slumpar vi sedan ut ett antal punkter runt varje mittpunkt, dessa representerar nervändarna som är förgrenade från den inkommande baspunkten. I den högra bilden är baspunkterna borttagna och vi har ett simulerat nervmönster.

När man undersöker nervmönster behövs en metod för att uppskatta de här parametrarna utifrån uppsättningen av ändpunkter. I den här studien har vi undersökt en sådan metod vid

namn 'minimum contrast', eller MCM. I MCM behöver man välja ett värde. Vi har studerat hur valet av detta värde påverkar parametrarna i klustermodellen.

Det visar sig att det är viktigt att välja ett värde som är åtminstone lika stort som spridningen av ändpunkterna i varje grupp. För att förhindra att en av parametrarna blir för högt uppskattad väljer man ett värde som är så litet som möjligt, men det har inte lika stor inverkan.

Problemet är att man i regel inte känner till spridningen innan man utför MCM. Därför har vi utvecklat nya metoder för att ta fram värdet utan att behöva känna till spridningen. Den ena metoden bygger på att man först genomför MCM med några valda startvärden. Därefter genomförs MCM igen med ett nytt värde baserat på de parametrar man fick ut från MCM. Den andra metoden går ut på att man medvetet väljer ett för stort värde och sedan går man nedåt tills dess att uppskattningen av parametrarna blir för avvikande och man väljer då det senaste värdet.

Uppskattningen av parametrarna med MCM blev marginellt bättre av att använda värden framtagna utav våra metoder jämfört med att bara använda standardvärden. Våra metoder ger instabilare uppskattningar av parametrarna jämfört med att använda standardvärden. Detta beror delvis på att standardvärdet till MCM är ganska väl valt.

Vi applicerade MCM med värden från våra metoder på nervmönster från såväl friska personer samt personer med mild diabetesneuropati. Nervmönstren från personer med diabetesneuropati hade i genomsnitt lite lägre antal nervändar och mindre spridning på dessa. Detta är lovande för att potentiellt kunna diagnostisera diabetesneuropati genom att enbart kontrollera nervmönstret, dock behöver vidare studier göras för att säkerställa att sambandet gäller.

Sammanfattning

Diabetiker kan ibland utveckla nervsjukdomen diabetesneuropati som bland annat kan orsaka känselbortfall och smärtor. Det har observerats att nervändarna i överhuden hos patienter med diabetesneuropati verkar vara mer klustrade än hos friska patienter [7]. Försök har gjorts i hopp om att i en tidig fas kunna urskilja patienter med risk att utveckla diabetesneuropati.

De nervmönster som nervändarna utgör kan på ett naturligt sätt beskrivas med spatiala punktprocesser där parametrarna till en sådan process kan erhållas med hjälp av 'minimum contrast'-metoden (MCM). Projektets huvudsakliga uppgift är att undersöka hur val av integrationsgräns i MCM påverkar parameterskattningarna.

Till denna studie valdes Thomasprocessen där vi utöver integrationsgränsen i MCM undersöker om denna process är ett lämpligt val för att beskriva nervmönster från patienter ur två grupper. Den ena gruppen utgör en kontrollgrupp av friska patienter utan symptom av diabetesneuropati. Den andra gruppen innehåller diabetespatienter med milda symptom av sjukdomen. Undersökningarna sker med hjälp av simuleringsstudier och R-paketet `spatstat` [1].

I det här arbetet visas att valet av integrationsgräns i MCM är viktigt och att integrationsgränsen måste vara över ett kritiskt värde för att undvika alltför stor varians hos parameterskattningarna. Dessutom finns det påföljder med att välja ett för stort värde av integrationsgränsen men det har inte lika stor betydelse. Slutligen visas att vi inte kan förkasta att Thomasprocessen beskriver nervmönstren från de två grupperna.

Abstract

People with diabetes can sometimes develop diabetic neuropathy, a nerve disease which can infuse numbness and pain. It has been observed that the end points of nerve fibers in the epidermis, which is the outermost living layer of the skin, appear more clustered on patients with diabetic neuropathy [7]. Attempts to model the fiber patterns have been made in the hope that at an early stage distinguish patients with risk of developing diabetic neuropathy.

These nerve fibers can be described by spatial point processes. The unknown parameters of such a process can be estimated using the minimum contrast method (MCM) where the main goal of this project is to investigate how the choice of upper integration limit in MCM affects the parameter estimates.

The Thomas process was chosen for this project where in addition to the examination of the integration limit in MCM we also examine if this process is a suitable choice to describe nerve fibers from patients associated with two groups. The first acts as a control group with healthy subjects without symptoms of diabetic neuropathy. The second group consist of subjects with diabetes and mild symptoms of the disease. This is done by means of simulation and the R-package `spatstat` [1].

It turns out that the choice of integration limit is important and that there are some guidelines of how to make this choice to avoid the variance of the parameter estimations getting too large. Likewise, complications in selecting too large of an integration limit are present although their importance are not as prominent. Lastly we show that we can not reject that the Thomas process describes nerve fibers from the two groups.

Innehåll

1	Inledning	1
1.1	Syfte	2
1.2	Problem	2
1.3	Avgränsningar	2
2	Teori	3
2.1	Spatiala punktprocesser	3
2.1.1	Typer av spatiala punktprocesser	3
2.1.2	Sammanfattande statistikor	4
2.2	Thomasprocessen	6
2.3	K -funktionen för Thomasprocessen	6
2.4	Uppskattning av K -funktionen	8
2.5	'Minimum contrast'-metoden	10
2.6	Anpassningstest	10
3	Utförande	12
3.1	Simuleringsarbete	12
3.1.1	Överskattning av $\hat{\kappa}$	14
3.1.2	Val av t_0	15
3.2	Nervdatan	16
3.2.1	Parameterskattningar på nervmönster	16
3.2.2	Anpassningstest på nervdatan	17
4	Diskussion	19
A	Appendix A	22
A.1	Utplaningspunkt	22
A.2	Maximum likelihood estimation av σ^2	22
A.3	χ^2 -fördelning med två frihetsgrader	23
B	Appendix B	23
B.1	Svepande t_0	23
B.2	Iterativt t_0	25

Förord

Vi vill tacka Aila Särkkä och Claes Andersson som har varit våra handledare för deras vägledning genom hela projektet. Vi är tacksamma för deras noggrannhet, för deras förslag och för all hjälp. Tillsammans har vi skapat ett projekt som vi i gruppen är väldigt stolta över.

Det har förts en dagbok som har sammanfattat veckans arbete och en loggbok för gruppens möten samt varje gruppmedlems nedlagda tid på projektet. Alla gruppmedlemmar har redigerat alla delar i projektet, men följande medlemmar tillskrivs som författare av

Christian Källgren:

Avsnitt 1, 2.1, 2.1.1, 2.1.2, 2.3, 2.4, 2.5, 2.6, 3.2.2, 4, populärvetenskaplig presentation och Appendix A.3

Hampus Lane:

Avsnitt 1, 2.1.1, 2.1.2, 3.1, 3.1.1, 3.1.2, 4, populärvetenskaplig presentation, Appendix A.1 och Appendix B.

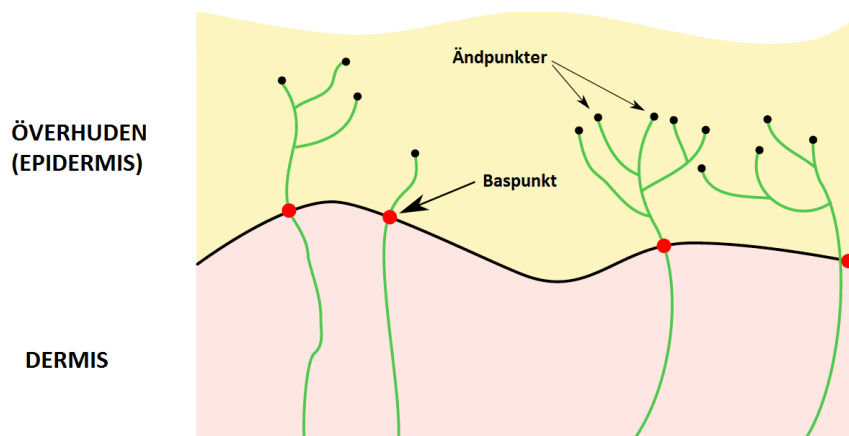
Simon Eriksson:

Avsnitt 1, 2.1.1, 2.1.2, 2.2, 2.5, 3.1, 3.1.1, 3.1.2, 3.2, 3.2.1, 4, sammanfattning och Appendix A.2

1 Inledning

Diabetesneuropati är en nervsjukdom som kan utvecklas hos diabetespatienter som bland annat kan orsaka känselbortfall och smärtor. Då bilder från provtagningar undersökts med mikroskop har det observerats att nervfibrer i överhuden hos patienter med diabetesneuropati verkar vara mer klustrade än hos friska patienter [7]. Försök att modellera dessa nervmönster har därför gjorts med ambition att på ett tidigt stadium kunna urskilja patienter med anlag att utveckla diabetesneuropati.

Speciellt är det av intresse att undersöka nervträdarnas ändpunkter och uppbyggnad då det är dem som känner av till exempel värme och smärta. För att undersöka dessa betraktas nervträdarnas uppbyggnad i överhuden. Då nervträd bildas kommer de så småningom tränga in i överhuden. Dessa inträdespunkter kallar vi baspunkter där valet av namn får sin förklaring av att nervträdarna därefter förgrenar sig i ytterhuden för att jämnt fördela känselkapaciteten över huden. Baspunkterna utgör alltså en bas för nya nervträd. De förgrenade nervträdarna avtar så småningom där ändarna då utgör ett punktmönster över huden.



Figur 3: (Claes Andersson, 2017) Nervträd som tränger in i överhuden

Då ändpunkterna observeras som punkter på en yta blir spatials punktprocesser ett naturligt alternativ att beskriva dessa nervmönster med. Spatials punktprocesser är stokastiska processer vars utfall är punkter i ett rum och kan kategoriseras utifrån hur de distribuerar sig. Processer där punkter tenderar att ligga nära varandra kallas klustrade medan processer där punkterna hamnar regelbundet utspridda kallas reguljära. Ett typiskt exempel på spatials punktprocesser är positioneringen hos träd. Träd vars frön inte sprider sig långt från moderplantan blir en realisation av en klusterprocess, planterade träd blir en realisation av en reguljär process, medan de träd som sprider sina frön med vinden kan positioneras helt slumpmässigt.

Ett sätt att jämföra de observerade nervmönstren mellan patienter med, respektive utan, diabetesneuropati är att anpassa en punktprocessmodell efter mönstren och jämföra de uppskattade parametervärdena. Vid modellenpassning uppskattas vanligtvis modellens parametrar med maximum likelihood-metoden (ML-metoden) men för spatials punktprocesser kan dock täthetsfunktionen och likelihoodfunktionen vara svårhanterliga. Ett möjligt alternativ är då 'minimum contrast'-metoden (MCM) som baseras på följande integral (1.1) där modellparametrarna θ väljs för att minimera integralen. $S(t; \theta)$ är en funktion som sammanfattar processen till modellen och $\hat{S}(t)$ dess uppskattning.

$$D(\theta) = \int_{\nu}^{t_0} w(t) \left(\hat{S}(t)^c - S(t; \theta)^c \right)^p dt \quad (1.1)$$

Ett problem med den här metoden är att det finns viss godtycklighet i valet av t_0 , c och $w(t)$ och det är därför intressant att veta hur stor inverkan dessa val har på resultatet.

Metoden och olika val av $S(t; \theta)$ beskrivs mer i nästa kapitel.

1.1 Syfte

Det huvudsakliga syftet i detta projekt är att med hjälp av simuleringar studera hur valet av integrationsgränsen t_0 i MCM påverkar parameterskattningarna, samt applicera resultatet på nervmönster från patienter ur två olika grupper. En kontrollgrupp bestående av friska patienter och en grupp bestående av diabetiker med milda symptom av diabetesneuropati. Vi refererar till samlingen av samtliga nervmönster som nervdatan.

1.2 Problem

Eftersom de observerade nervmönstren tycks vara grupperade väljs en klusterprocess som vi antar är ett rimligt val för att beskriva nervmönstren. Thomasprocessen är en sådan och valdes därför att dess konstruktion påminner om nervmönstrens uppbyggnad med bas- och ändpunkter.

Simuleringsstudien fokuserar främst på att undersöka hur valet av integrationsgränsen i MCM påverkar parameterskattningarna och hur den kan väljas på ett bra sätt. Om vi vet vilken inverkan valet av integrationsgränsen har kan det hjälpa oss när vi undersöker nervmönstren där de sanna parametrarna är okända.

Med parameterskattningarna kan vi slutligen undersöka hur bra Thomasprocessen beskriver nervmönstren. Även om det visar sig att Thomasprocessen beskriver nervmönstren på ett bra sätt kan det bli aktuellt att undersöka om något annat val av process beskriver nervmönstren bättre.

1.3 Avgränsningar

Vi begränsade arbetet till att endast studera en typ av punktprocess och att endast fokusera på integrationsgränsens inverkan på MCM. Förslag till utökning av projektet diskuteras längre fram.

2 Teori

I detta avsnitt introduceras den teori om spatiala punktprocesser som berör projektet. En allmän presentation av spatiala punktprocesser ges och komponenter som använts under arbetet diskuteras. Bland dessa ingår Thomasprocessen, Ripleys K-funktion och speciellt MCM som utgör kärnan av projektet. Litteraturen som används för att sammanställa projektets teori är Diggle [2], Illian [4] och Møller [5].

2.1 Spatiala punktprocesser

En slumpmässig mekanism vars realisationer är punkter i något rum S , kallar vi för en spatial punktprocess. Utfallen kan vara oregelbundna och sakna tydlig struktur, det vill säga helt slumpmässiga. Alternativt är utfallen grupperade, vilket vi kallar klustrade, eller så har utfallen en mer reguljär struktur. Vi benämner realisationer av spatiala punktprocesser som spatiala punktmönster. Utfallsrummet för en spatial punktprocess är rummet av konfigurationer i S . Även då processen är definierad på hela S görs m observationer, x_1, \dots, x_m , enbart i en begränsad delmängd A av S . Allmänt för de flesta tillämpningar är $A \subset S = \mathbb{R}^n$. Vi ger här exempel på spatiala punktmönster och möjligtvis naturligt tillhörande antaganden av A .

Exempel 2.1. Ett klassiskt problem är att förstå fördelningen av någon specifik trädart i ett område, t.ex. askar inom en kvadratkilometer. I detta fallet är antagandet $A = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. Ett ytterligare exempel kan vara att förstå fördelningen av stormar eller allvarliga oväder på jorden och dess förflyttning i tiden. Då kan antagandet vara $A = \mathcal{S}^2 \times \mathbb{R} \subset \mathbb{R}^4$.

Vi kommer endast studera punktprocesser som är stationära och isotropa. En spatial punktprocess sägs vara stationär om antalet punkter i $A \subset S$ bara beror på arean eller volymen $|A|$ och inte på dess position i S . En spatial punktprocess är isotrop om den är invariant under rotation, det vill säga att fördelningen av punkter runt varje godtycklig punkt är densamma i alla riktningar.

2.1.1 Typer av spatiala punktprocesser

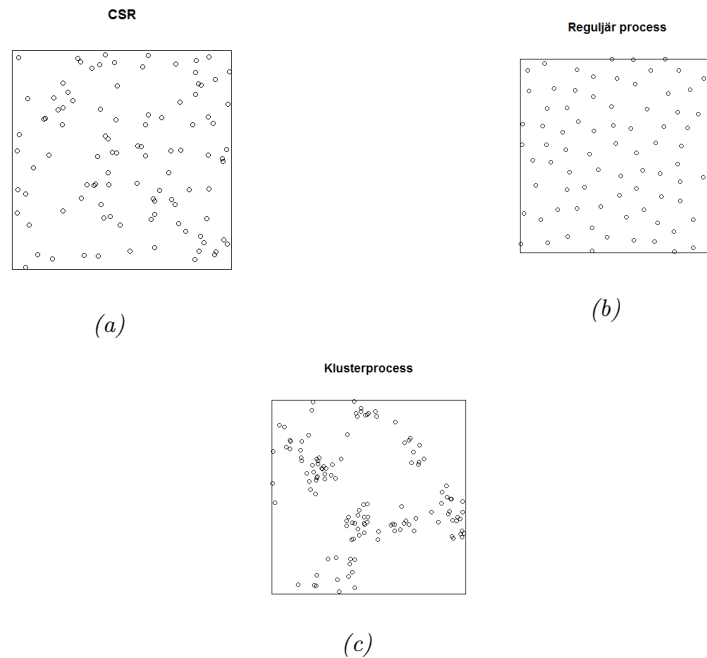
Den enklaste typen av spatiala punktprocesser är en homogen spatial Poissonprocess, även kallad Complete Spatial Randomness (CSR). CSR definieras genom följande två kriterier.

Definition 2.1. Vi definierar en spatial punktprocess med intensitet λ per enhetsyta som CSR om följande gäller.

- i. Antalet punkter n i ett område $A \subset S$ är Poissonfördelat med väntevärdet $\lambda|A|$.
- ii. Givet n punkter i ett område A , så är dessa oberoende och likformigt fördelade på A .

Då punkterna är oberoende och likformigt fördelade över området A är CSR processer stationära och isotropa.

En realisation av en Poissonprocess är typiskt ett spatialt punktmönster som saknar tydlig struktur. Detta illustreras i figur 4a. Om mönstret ser ut att ha en struktur finns formellt två andra typer, reguljärt och klustrat, vilket illustreras i figur 4b respektive 4c.



Figur 4: Exempel på tre typer av punktmönster. CSR (a), Reguljärt (b) och klustrat (c). För samtliga mönster gäller $|A| = 1$

Även om dessa tre klassificeringar kan ses restriktiva kan samma spatiala punktmönster ha olika struktur på olika skalor. Vi försöker illustrera detta genom följande enkla exempel.

Exempel 2.2. Säg att man skulle hålla ut 20 tärningar på ett kvadratisk bord med area 1, och betrakta prickarna på tärningarna som ett spatialt punktmönster. Då finns det lokalt, på varje enskild tärning, ett reguljärt mönster men på mer global skala kan det till exempel vara klustrat. Här är antagandet $A = [0, 1] \times [0, 1] \subset S = \mathbb{R}^2$.

2.1.2 Sammanfattande statistikor

När en samling data presenteras kan det ibland vara svårt att få en bra överblick över vilka egenskaper eller karaktärer datan har. Idén med sammanfattande statistikor är att hitta en metod som sammanfattar observationer på ett sätt som förmedlar så mycket information som möjligt om datan. Ett vanligt exempel är medelvärdet av en samling tal.

För spatiala punktmönster brukar sammanfattande statistikor istället vara funktioner som beskriver avstånd mellan utfall. I det här stycket följer en beskrivning av de tre vanligaste sammanfattande statistikorna för spatiala punktprocesser.

2.1.2.1 G-funktionen

G-funktionen beskriver fördelningen av avståndet från ett godtyckligt utfall till det närmsta utfallet. Normalt finns ingen sluten form av $G(t)$ och uppskattningar av funktionen får användas istället. Om x_i är ett utfall i A och t_i är avståndet till dess närmaste granne x_j kan vi uppskatta G-funktionen genom att räkna andelen punkter vars närmaste granne är inom avstånd t (Se Figur 5):

$$\hat{G}(t) = n^{-1} \sum_{i=1}^n \mathbf{1}\{t_i \leq t\} \quad (2.1)$$

där n är antalet utfall i A och $\mathbf{1}\{\cdot\}$ är indikatorfunktionen som är 1 då argumentet i parentesen är uppfyllt och 0 annars. Under Poissonantagandet är det teoretiska värdet för G följande:

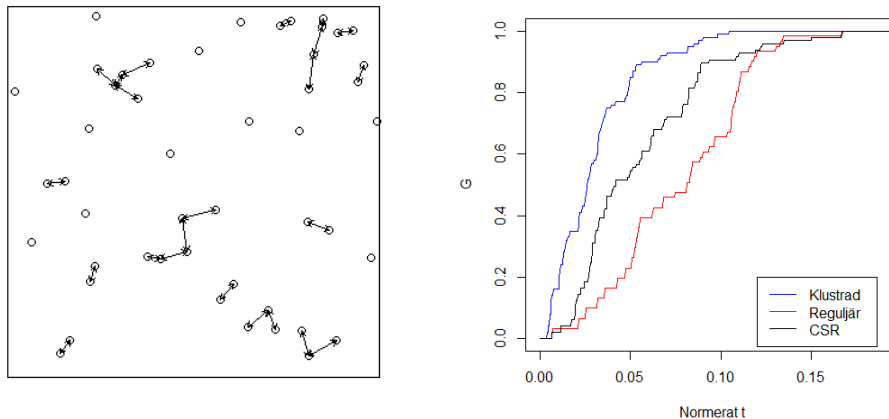
$$G(t) = 1 - (P(\#t_i < t = 0)). \quad (2.2)$$

Då alla är utfall är oberoende och likformigt fördelade blir det

$$G(t) = 1 - e^{-\lambda\pi t^2}, \quad \hat{\lambda} = n|A|^{-1}. \quad (2.3)$$

Notera att dessa funktioner bara stämmer om man studerar en oändlig yta. För att få G på t.ex. en enhetskvadrat måste man således lägga till kantkorrigering för att ta hänsyn till de saknade punkterna utanför randen.

För klusterprocesser är det genomsnittliga avståndet till närmsta grannen lägre, således ligger dess G -funktion över G -funktionen för CSR med motsvarande intensitet. På samma sätt ligger G -funktionen under CSR för reguljära processer. Se Figur 5



Figur 5: Till vänster ser vi alla punkter vars närmsta granne är inom avstånd 0.1 markerade med ett streck till sin närmsta granne. Således uppskattas $G(0.1)$ av andelen av alla punkter som är förbundna med ett streck. Till höger ser vi $\hat{G}(t)$, för en kluster-, reguljär- respektive Poissonprocess.

2.1.2.2 F-funktionen

Det andra alternativet sammanställer avståndet t_i från en godtycklig punkt i A till det närmsta utfallet x_i . Ett beprövat tillvägagångssätt vid uppskattning av $F(t)$ är att slumpa de godtyckliga punkterna från punkter i ett $[k \times k]$ -rutnät. Ett tillräckligt stort k bör väljas för att uppskattningen ska bli tillräckligt bra. Däremot begränsas uppskattningsprecisionen av antalet utfallspunkter n så alltför stora k kommer inte ha någon inverkan.

Likt G -funktionen kan vi uppskatta $F(t)$ med

$$\hat{F}(t) = m^{-1} \sum_{i=1}^m \mathbf{1}\{t_i \leq t\} \quad (2.4)$$

där m är antalet godtyckliga punkter. Under Poissonantagandet gäller återigen

$$F(t) = 1 - e^{-\pi\lambda t^2}, \quad \hat{\lambda} = n|A|^{-1} \quad (2.5)$$

I en klusterprocess ligger de flesta utfallen koncentrerade, därför kommer närmsta utfall från en godtycklig punkt vara längre bort än för CSR. Således är F -funktionen under CSR i fallet för klusterprocesser. På samma sätt är F -funktionen över CSR för reguljära processer.

2.1.2.3 K-funktionen

K -funktionen är relaterat till antalet utfall som ligger inom en radie r från ett godtyckligt utfall. För detta projekt valde vi främst att använda oss av detta alternativ och beskrivs därför mer utförligt i avsnitt 2.3.

Ripleys K -funktion definieras av

$$\lambda K(r) = \mathbb{E}_0[N_X(B_0)] \quad (2.6)$$

där B_0 är en boll centrerad i godtycklig punkt x_i av processen med radie r och $N_X(B_0)$ är antalet ytterligare utfall från processen inom ett avstånd r från ett godtyckligt utfall. Processens intensitet är λ , det vill säga antalet utfall per areaenhet. Det utfall som radien utgår från ingår inte.

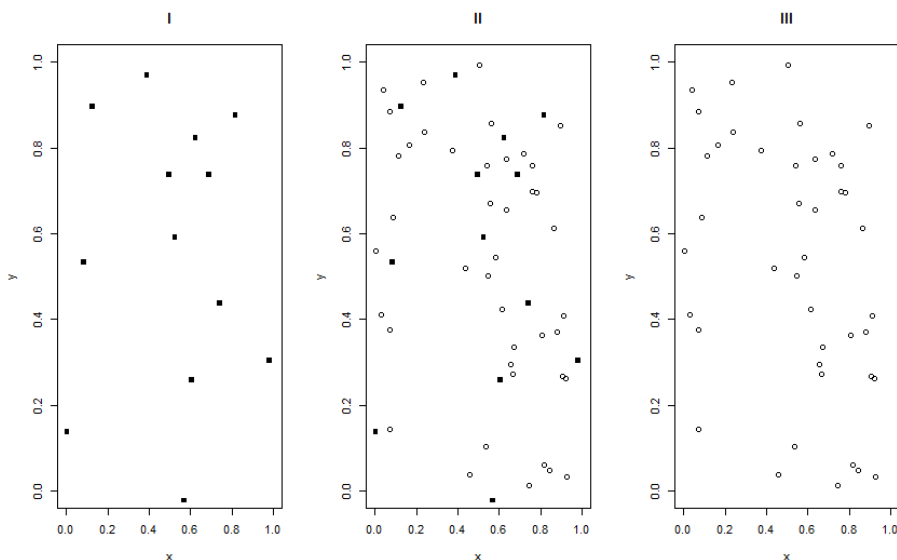
K-funktionen ligger över Poissonfallet för klusterprocesser och under Poisson för reguljära. Om vi erinrar exempel 2.2 kan vi notera att K-funktionen kommer ligga under CSR upp till tärningsstorleken, då de på denna skala är reguljära och därefter ligga över CSR då de på alla större skalor är klustrade.

2.2 Thomasprocessen

Thomasprocessen kan illustreras utifrån tre steg. Först utgår vi från en homogen Poisson-process som genererar så kallade föräldrapunkter med intensitet κ i ett område A . Dessa punkter genererar i sin tur dotterpunkter som till antalet är Poissonfördelade med intensitet μ . Dotterpunkterna ersätter sedan sina föräldrapunkter. Dotterpunkternas positionering kring respektive föräldrapunkt är oberoende och normalfördelade $N(0, \sigma^2\mathbf{I})$. Detta leder till att punkterna fördelar sig så att de är invarianta under translation och rotation kring origo, en egenskap som benämns stationaritet respektive isotropi.

Denna konstruktion påminner om hur kroppens nervtrådar tränger in i överhuden för att sedan förgrena sig, vilket beskrevs i avsnitt 1. Detta är en orsak till att Thomasprocessen valts till projektet.

Vi betecknar Thomasprocessen med $T(\kappa, \mu, \sigma)$. Processens intensitet ges av $\lambda = \kappa\mu$.



Figur 6: **(I)** Genererade föräldrapunkter. **(II)** Genererade föräldrapunkter och dotterpunkter **(III)** Observerat punktmönster.

2.3 K-funktionen för Thomasprocessen

För att härleda det teoretiska uttrycket för Ripleys K -funktion för Thomasprocessen använder vi ekvation (2.6). Antag vidare att vi har en punkt x_i placerad i origo tillhörande kluster c . Vi börjar med att betrakta väntevärdet $\mathbb{E}_0[N_X(B_0)]$ som en summa av två väntevärden,

$$\mathbb{E}_0[N_X(B_0)] = \mathbb{E}_0[N_{X_c}(B_0)] + \mathbb{E}[N_{X-c}(B_0)]. \quad (2.7)$$

Den ena är det förväntade antalet ytterligare punkter inom radie r från en godtycklig punkt inom samma kluster c . Detta betecknas $\mathbb{E}_0[N_{X_c}(B_0)]$. Den andra termen är det förväntade antalet ytterligare punkter inom radie r , men tillhörande ett annat kluster, betecknat $\mathbb{E}[N_{X-c}(B_0)]$. Detta väntevärde är inte specifikt till något kluster utan är samma över hela

mönstret X . Alltså är det enbart arean av B_0 , vilket är πr^2 , multiplicerat med den totala intensitet för processen, $\lambda = \kappa\mu$. Således är $\mathbb{E}[N_{X_c}(B_0)] = \pi r^2 \kappa\mu$. Den första termen i högerledet i (2.7) kräver lite utförligare uträkning. Vi använder lagen om sammanlagd förväntan

$$\mathbb{E}_0[N_{X_c}(B_0)] = \sum_{k=2}^{\infty} \mathbb{E}_0[N_{X_c}(B_0) \mid K_{X_c} = k] P(K_{X_c} = k). \quad (2.8)$$

Summeringen är över väntevärden av $N_{X_c}(B_0)$ betingat på totala storleken på klustret, betecknat K_{X_c} , multiplicerat med sannolikheten att slumpmässigt välja ett kluster av just den storleken. Vi börjar summeringen med $k = 2$ istället för $k = 0$, eftersom ett kluster med noll punkter saknar mening. Ett kluster med $k = 1$ ger $\mathbb{E}_0[N_{X_c}(B_0) \mid K_{X_c} = 1] = 0$, då vi enbart räknar de ytterligare punkterna i klustret. Den första faktorn i högerledet ur (2.8), $N_{X_c}(B_0) \mid K_{X_c} = k$, är antalet ytterligare punkter tillhörande kluster c inom radie r , givet totalt k punkter tillhörande klustret. De $k - 1$ ytterligare punkterna i klustret ligger inom radie r med en sannolikhet $F(r)$ där $F(r)$ är fördelningsfunktionen för avståndet mellan två punkter tillhörande samma kluster, $F(r)$ är således fördelningsfunktionen för avståndet mellan två oberoende och normalfördelade variabler. Slumpvariabeln $N_{X_c}(B_0) \mid K_{X_c} = k$ är därför binomialfördelad med $k - 1$ som antal utfall och $F(r)$ är sannolikheten för varje utfall, $N_{X_c}(B_0) \mid K_{X_c} = k \sim \text{Bin}(k - 1, F(r))$. Väntevärdet av en binomialfördelad variabel är antalet utfall multiplicerat med sannolikheten,

$$\mathbb{E}_0[N_{X_c}(B_0) \mid K_{X_c} = k] = (k - 1)F(r). \quad (2.9)$$

För att få ett enklare uttryck för den andra faktorn i 2.8, $P(K_{X_c} = k)$, utgår vi ifrån appendix A.2 i [3]. Det vill säga vi observerar att sannolikheten att slumpmässigt välja ett kluster av storlek k , $P(K_{X_c} = k)$, är proportionellt med k multiplicerat med sannolikheten att ett kluster är av storlek k , $P(D_c = k)$. Men för att $P(K_{X_c} = k)$ ska vara en sannolikhetsfunktion måste $\sum_{k=1}^{\infty} P(K_{X_c} = k) = 1$. Vi åstadkommer det genom att dela varje $P(K_{X_c} = k)$, för $k = 1, 2, 3, \dots$, med $\sum_{l=1}^{\infty} lP(D_c = l)$, eftersom $\frac{\sum_{k=1}^{\infty} kP(D_c = k)}{\sum_{l=1}^{\infty} lP(D_c = l)} = 1$. Men $\sum_{l=1}^{\infty} lP(D_c = l) = \mathbb{E}[D_c] = \mu$. Alltså har vi

$$P(K_{X_c} = k) = \frac{kP(D_c = k)}{\mu}. \quad (2.10)$$

Utifrån (2.9), (2.10) och att $D_c \sim \text{Poisson}(\mu)$ blir (2.8)

$$\begin{aligned} \mathbb{E}_0[N_{X_c}(B_0)] &= \sum_{k=2}^{\infty} \mathbb{E}_0[N_{X_c}(B_0) \mid K_{X_c} = k] P(K_{X_c} = k) \\ &= \sum_{k=2}^{\infty} (k - 1)F(r) \frac{kP(D_c = k)}{\mu} \\ &= \frac{F(r)}{\mu} \sum_{k=2}^{\infty} (k - 1)kP(D_c = k) \\ &= \frac{F(r)}{\mu} \mathbb{E}[D_c(D_c - 1)] \\ &= \frac{F(r)}{\mu} (\mu^2 + \mu - \mu) \\ &= \mu F(r). \end{aligned} \quad (2.11)$$

Vi härleder fördelningsfunktionen för avståndet mellan två oberoende och normalfördelade variabler, $F(r)$. Låt $X = (x_1, x_2)$ och $Y = (y_1, y_2)$ vara punkter i \mathbb{R}^2 där varje x_i och y_i är oberoende och $\mathcal{N}(a, \omega^2)$. Då är, med $\|\cdot\|$ som det euklidiska avståndet,

$$\begin{aligned} F(r) &= P(\|X - Y\| \leq r) \\ &= P(\|X - Y\|^2 \leq r^2) \\ &= P((x_1 - y_1)^2 + (x_2 - y_2)^2 \leq r^2). \end{aligned} \quad (2.12)$$

Differensen $x_i - y_i$ har väntevärde 0 och varians $2\omega^2$. Vi multiplicerar båda sidor med inversen av variansen $\frac{1}{2\omega^2}$, vilket ger $\frac{x_i - y_i}{\sqrt{2\omega}} \sim \mathcal{N}(0, 1)$ innanför kvadraterna i (2.12).

$$P((x_1 - y_1)^2 + (x_2 - y_2)^2 \leq r^2) = P\left(\left(\frac{x_1 - y_1}{\sqrt{2\omega}}\right)^2 + \left(\frac{x_2 - y_2}{\sqrt{2\omega}}\right)^2 \leq \frac{r^2}{2\omega^2}\right).$$

Alltså har vi summan av kvadrater från två oberoende, standardiserade normalfördelade variabler, vilket är definitionen av en χ^2 -fördelad variabel med två frihetsgrader.

$$\left(\frac{x_1 - y_1}{\sqrt{2\omega}}\right)^2 + \left(\frac{x_2 - y_2}{\sqrt{2\omega}}\right)^2 = Q \sim \chi_2^2.$$

Fördelningsfunktionen för en χ^2 -fördelad variabel med två frihetsgrader är, se appendix A.3,

$$P(X \leq x) = F(x) = 1 - e^{-x/2}.$$

Med $x = r^2/2\omega^2$,

$$P\left(Q \leq \frac{r^2}{2\omega^2}\right) = 1 - e^{-r^2/4\omega^2}. \quad (2.13)$$

Alltså får vi utav $\mathbb{E}[N_{X-c}(B_0)] = \pi r^2 \kappa \mu$, (2.11), och (2.13),

$$\begin{aligned} \kappa \mu K(r) &= \mathbb{E}_0[N_{Xc}(B_0)] + \mathbb{E}[N_{X-c}(B_0)]. \\ &= \pi r^2 \kappa \mu + \mu F(r) \\ &= \pi r^2 \kappa \mu + \mu(1 - e^{-r^2/4\omega^2}). \end{aligned}$$

Thomasprocessens K-funktion ges således av, med σ^2 istället för ω^2 ,

$$K(r) = \pi r^2 + \kappa^{-1}(1 - e^{-r^2/4\sigma^2}). \quad (2.14)$$

Anmärkning 2.1. K-funktionen för Thomasprocessen är oberoende av parametern för intensiteten av dotterpunkterna, μ , vilket ger ett exempel på att K-funktionen inte är unik. Det vill säga om två Thomasprocesser $T_1(\kappa, \mu_1, \sigma)$ och $T_2(\kappa, \mu_2, \sigma)$ skiljs åt enbart i parametern μ har de samma K-funktion.

2.4 Uppskattning av K-funktionen

För att härleda en uppskattning av Ripleys K-funktion, betecknad $\hat{K}(r)$, utgår vi ifrån samma uttryck som för den teoretiska, (2.6), och använder liknande notation som i Diggle [2]. Då vi vill uppskatta hur många punkter i genomsnitt finns i en boll kring en godtycklig punkt av processen antas isotropi.

Anmärkning 2.2. Den uppskattade K-funktionen vi härleder här är inte specifik till Thomasprocessen utan är en allmän funktion beroende av underliggande process enbart i uppskattningen av λ .

Väntevärdet $\mathbb{E}[N_X(B_0)]$ uppskattas genom att summera antalet punkter inom radie r från varje $x_i \in A$ och dela med det totala antalet punkter n i A . Uppskattningen av väntevärdet blir

$$\hat{E}[N_X(B_0)] = n^{-1} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbf{1}\{d_{ij} \leq r\}. \quad (2.15)$$

Där $d_{ij} = \|x_i - x_j\|$ är det euklidiska avståndet mellan x_i och x_j . I uppskattningen av väntevärdet (2.15) tas ingen hänsyn till punkter utanför området A inom ett avstånd r från en godtycklig punkt $x_i \in A$ av processen. Detta leder till att antalet punkter inom en radie r då är färre än det förväntade. Detta fenomen kallas kanteffekter, vilket illustreras i Figur 7. Uppskattning blir då ej väntevärdesriktig, i att den underskattar antalet punkter inom given radie r . Vi inför en viktfunction w_{ij} med syfte att minska denna effekt.

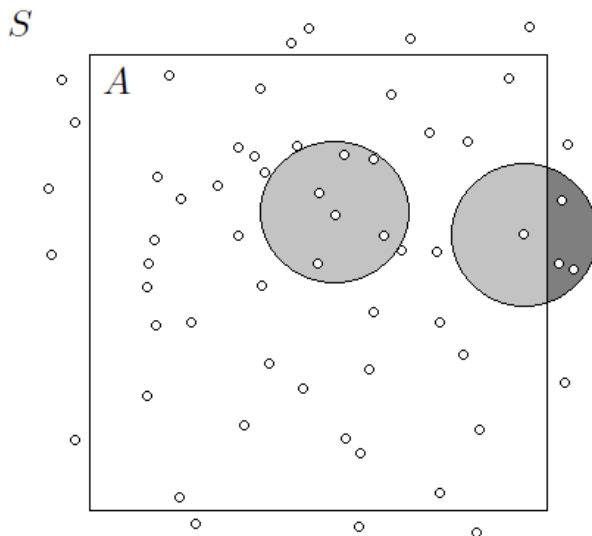
Vi kommer att använda Ripleys viktfunction [6], med illustration i Figur 8. Låt $w(x_i, r)$ vara andelen av omkretsen av cirkeln med center i $x_i \in A$ och radie r som ligger i området A . Sätt $w_{ij} = w(x_i, d_{ij})$. Givet en stationär och isotrop process är w_{ij} den betingade sannolikheten att observera en ytterligare punkt inom ett avstånd d_{ij} från en punkt x_i av processen. För punkter x_i med andel av cirkeln utanför A och x_j med hela cirkeln innanför, som i Figur 8, är $w_{ij} \neq w_{ji}$. Då vi tar inversen av vikterna får punkter med andel av cirkeln utanför A större vikt än punkter med hela cirkeln innanför A . För de punkter med hela cirkeln innanför området A är $w(x_j, d_{ji}) = 1$. En väntevärdesriktigt uppskattning av $\mathbb{E}[N_X(B_0)]$ är således

$$\hat{E}[N_X(B_0)] = n^{-1} \sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}^{-1} \mathbf{1}\{d_{ij} \leq r\}. \quad (2.16)$$

Genom att uppskatta λ med $(n-1)/|A|$, i enighet med funktioner ur `spatstat` paketet [1], där n är totala antalet punkter av processen i A får vi Ripleys uppskattning av K -funktionen [6]

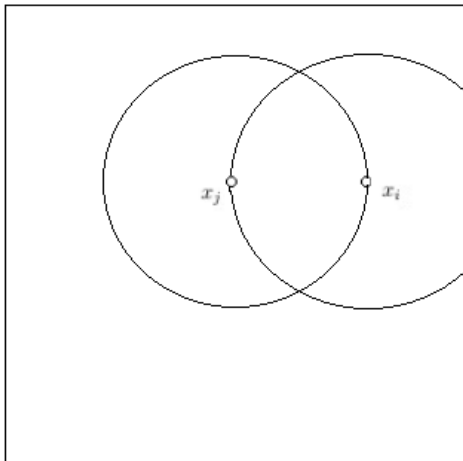
$$\hat{K}(r) = (n(n-1))^{-1} |A| \sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}^{-1} \mathbf{1}\{d_{ij} \leq r\}. \quad (2.17)$$

Att använda $n-1$ istället för n gör ingen skillnad för stora n . Denna uppskattning är en approximativt väntevärdesriktig¹ skattning av $K(r)$ för tillräckligt små r . Restriktionen på r beror på att viktfunctionerna kan bli obegränsade då r blir för stort.



Figur 7: Illustration av kanteffekter där vi visar att efter val av A har en punkt inga grannar inom radie r inuti A och därmed färre än förväntat.

¹Det vill säga då $n \rightarrow \infty$, går förväntningskevheten asymptotiskt mot 0.



Figur 8: Illustration till Ripleys viktfunktion för korrigerig av kanteffekter, där en punkt som ligger närmare kanten av A tilldelas en större vikt för att kompensera för kanteffekter.

2.5 'Minimum contrast'-metoden

För punktprocesser blir det traditionella valet av ML-metoden vid parameterskattningar ofta ett problem eftersom likelihoodfunktionen kan vara svårhanterlig då den inte har en sluten form för den valda modellen. Ett alternativ kan då vara MCM som kan liknas vid en minsta-kvadratmetod där vi istället för en summa använder en integral och minimerar avvikelsen mellan $S(t; \theta)$ och $\hat{S}(t)$, den teoretiska respektive uppskattade sammanfattande statistikan. Därmed uppskattas parametervektorn θ med det $\hat{\theta}$ så att integralen (2.18) blir så liten som möjligt.

$$D(\theta) = \int_{\nu}^{t_0} w(t) \left(\hat{S}(t)^c - S(t; \theta)^c \right)^p dt \quad (2.18)$$

I detta projekt användes K-funktionen som sammanfattande statistika. Då återstår valet av integrationsgränserna ν och t_0 , viktfunktionen $w(t)$ samt konstanterna p och c . Enligt Diggle [2] antyder tidigare simuleringar och observationer att $c = 0.25$ och $w(t) = 1$ är rimliga val för klustrade punktmönster, såsom t.ex. realisationer av Thomasprocessen, och valdes därför att användas i projektet. Därtill är $p = 2$, likt minsta kvadratmetoden, vanligt förekommande.

För vårt val av sammanfattande statistika kan vi av definitionen (2.6) konstatera att $\nu = 0$ är ett naturligt val av undre integrationsgräns eftersom t betecknar ett avstånd.

Valet av t_0 är centralt i detta projekt och diskuteras mer utförligt längre fram i avsnitt 3.1. Således reduceras (2.18) till

$$D(\theta) = \int_0^{t_0} \left(\hat{K}(t)^{0.25} - K(t; \theta)^{0.25} \right)^2 dt. \quad (2.19)$$

För att åter referera till Diggle rekommenderas att t_0 inte väljs större än $0.25 \min(a, b)$ för rektangulära områden med sidlängder a och b , vilket utnyttjas i avsnitt 3.1. Vi kommer att använda (2.19) genom R-funktionen `mincontrast` för uppskattning av θ . I enighet med diskussionen ovan om val av konstanter är dessa standard i `mincontrast`-funktionen.

2.6 Anpassningstest

Så småningom kommer vi vara intresserade att se om valet av Thomasprocess är ett lämpligt val för att beskriva nervdatan. Därför behövs ett sätt att mäta om en vald modell passar

tillhörande data.

Vi börjar med att konstruera ett Monte Carlo-test med teststatistika baserad på G-funktionen. Den består av $\hat{G}_1(t)$ uppskattad från observerad data och $\hat{G}_i(t)$, $i = 2, \dots, s$ från $s - 1$ stycken simuleringar från den anpassade Thomasprocessen, uppskattade med (2.1). Dessa jämförs mot $\bar{G}_i(t) = \frac{1}{s-1} \sum_{j \neq i} \hat{G}_j(t)$, som är ett medelvärde för alla $\hat{G}_j(t)$ utom för index i . G-funktionen uppskattas för varje $t \in [0, \min(a, b)]$, vilket ger

$$g_i = \sum_t \left(\hat{G}_i(t) - \bar{G}_i(t) \right)^2.$$

Varje g_i är en summa av den kvadratiske skillnaden mellan G-funktionen uppskattad från datan respektive simulerad data mot ett medelvärde av de $s - 1$ andra G-funktionerna. Då vi vill testa om den observerade datan är en realisation av den anpassade Thomasprocessen, vill vi jämföra g_1 mot simuleringarna av modellen. Testet går därför ut på att vi räknar hur många g_i som är större än g_1 , vilket ger oss ett uttryck för p-värdet,

$$\text{p-värde} = \frac{1}{s} \sum_{i=1}^s I(g_i \geq g_1)$$

För Thomasprocessen finns en sluten form på K -funktionen och vi använder den för att skapa ett liknande Monte Carlo-test. Här summerar vi den kvadratiske skillnaden mellan den uppskattade K -funktionen (2.17) med den teoretiska, vilket i vårt fall är (2.14). Detta leder till teststatistikan med uppskattade κ och σ ,

$$k_i = \sum_t \left[\hat{K}(t) - \pi t^2 - \hat{\kappa}^{-1} (1 - e^{-t^2/4\hat{\sigma}^2}) \right]^2.$$

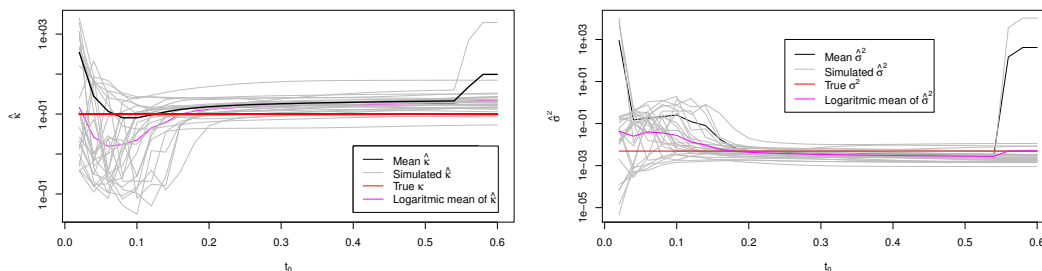
Med motsvarande uttryck för p-värdet

$$\text{p-värde} = \frac{1}{s} \sum_{i=1}^s I(k_i \geq k_1).$$

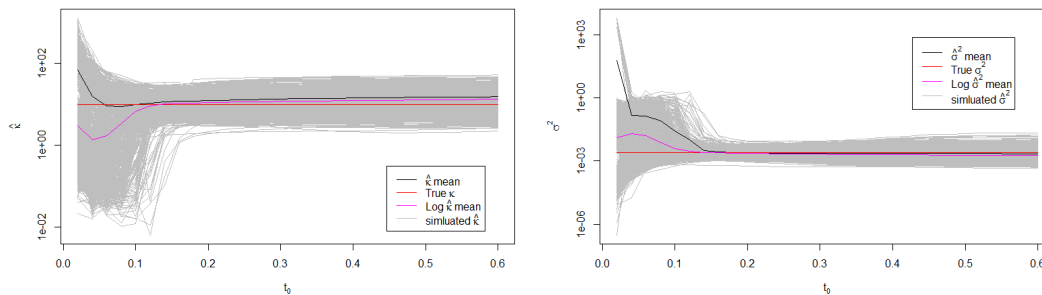
3 Utförande

3.1 Simuleringsarbete

För att undersöka thomasprocessens känslighet för t_0 genomfördes en simulationsstudie med R-paketet `spatstat`. Specifikt användes `rThomas` för att generera realisationer av Thomasprocesser, samt `thomas.estK` för parameteruppskattning. `thomas.estK` utför MCM med kantkorrigering, K -funktionen samt standardvärden $q = 0.25$ och $p = 2$. För att undersöka integrationsgränsen simulerades ett flertal realisationer av Thomasprocessen med $\kappa = \mu = 10$ och $\sigma = 0.05$ i en enhetskvadrat. Från dessa uppskattades parametrar med MCM för flera integrationsgränser t_0 i intervallet $[0, 0, 6]$. Resultatet illustreras i figur 9.

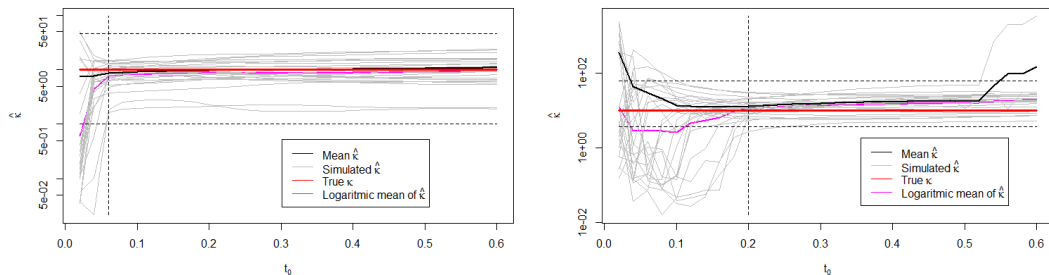


Figur 9: Figurerna ovan visar parametrar κ (vänster) och σ^2 (höger) uppskattade genom MCM från 100 simulerade Thomasprocesser, som funktion av integrationsgränsen t_0 . De gråa linjerna representerar parametervärdet för varje enskild realisation av Thomasprocessen. Den svarta linjen representerar medelvärdet av alla realisationer, den magenta det logaritmiska medelvärdet och den röda det faktiska värdet på parametern.

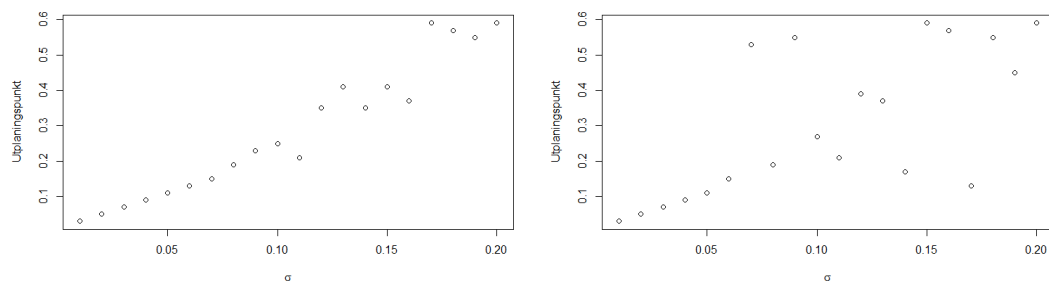


Figur 10: 1000 realisationer κ och σ^2 .

Detta upprepades sedan för ett flertal olika σ mellan 0 och 0,2. För alla dessa tenderade de uppskattade parametrarna att fluktuera kraftigt fram till ett visst t_0 , där de planade ut, se figur 11. Dessa utplaningspunkter hittades genom att först kasta bort högsta och minsta värdena av $\hat{\kappa}$. Därefter valdes det högsta t_0 där något $\hat{\kappa}$ låg utanför ett visst intervall, här valt till inom hälften av näst minsta värdet och dubbla näst största värdet. Utplaningspunkter som funktion av σ kan ses i figur 12.



Figur 11: Uppskattat κ som funktion av integrationsgränsen t_0 för $\sigma = 0.02$ (vänster) respektive $\sigma = 0.08$. Notera hur båda varierar kraftigt fram till en utplaningspunkt markerat med en lodrät streckad linje

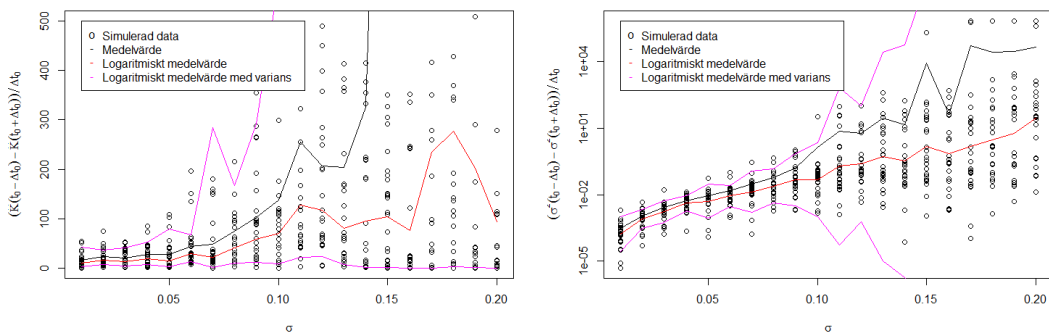


Figur 12: Utplaningspunkt t_0 av $\hat{\kappa}$ till vänster och $\hat{\sigma}^2$ till höger som funktion av σ . Notera att kring σ större än 0.1 verkar ett linjärt samband stämma lite sämre. För $\hat{\sigma}^2$ fungerade även två värden illa innan $\sigma = 0.1$

För att noggrannare undersöka känsligheten studerades beloppet av differenskvoten

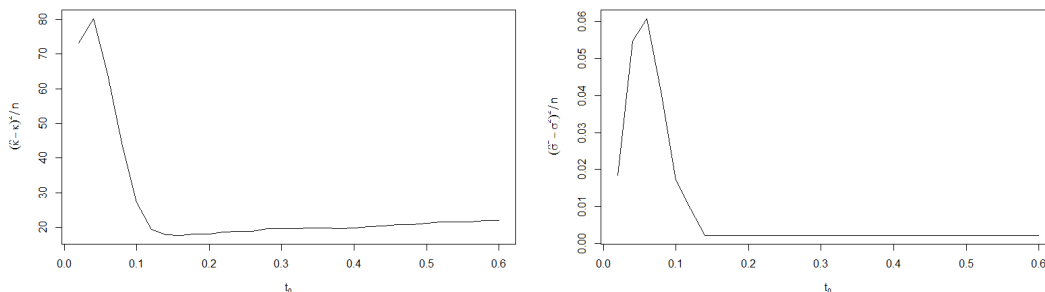
$$\frac{\hat{\kappa}(t_0 + \Delta t) - \hat{\kappa}(t_0 - \Delta t)}{\Delta t},$$

där $\hat{\kappa}$ är den uppskattade parametern κ , Δt är litet, här valt till 0.001 och t_0 är värdet på integrationsgränsen, här valt nära standardvärdet 0.25. Värdet på differenskvoten för olika σ med $\kappa = \mu = 10$ kan ses i figur 13. Notera hur värdet blir mycket stort vid $\sigma = 0.13$, då utplaningspunkten hamnar efter $t_0 = 0.25$.



Figur 13: Graferna ovan visar differenskvoten av de MCM-skattade parametrarna kring $t_0 = 0.24$ som funktion av σ . Notera hur medelvärdet (vänster figur) sticker iväg ungefär vid $\sigma = 0.13$, vilket är var utplaningspunkten hamnar bakom $t_0 = 0.24$, vilket drastiskt ökar differenskvoten.

I figur 14 visas $(\hat{\kappa} - \kappa)^2$ och $(\hat{\sigma} - \sigma)^2$, $\hat{\kappa}$ och $\hat{\sigma}$ är de uppskattade parametrarna från en process med sanna värden κ och σ . Detta görs för flera olika t_0 . Man kan tydligt se att variationen är stor fram till utplaningspunkten, varefter $(\hat{\sigma} - \sigma)^2$ planar ut, medan $(\hat{\kappa} - \kappa)^2$ landar på en svag men tydlig ökning. Ett optimalt val av t_0 bör således ligga så nära efter utplaningspunkten som möjligt.



Figur 14: Genomsnittligt kvadratisk fel för $\hat{\kappa}$ (vänster) och $\hat{\sigma}$ (höger). Samma parametervärden som ovan, $\kappa = 10$ och $\sigma = 0.05$

3.1.1 Överskattning av $\hat{\kappa}$

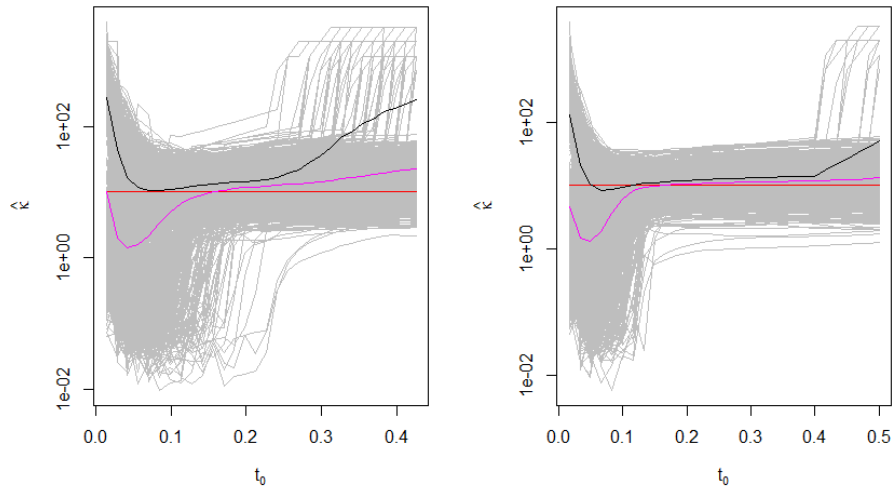
Under simuleringsstudien observerades hur medelvärdet av parameterskattningen $\hat{\kappa}$ har en svag positiv lutning, vilket kan ses i figur 10 och 14. Det innebär att om integrationsgränsen t_0 väljs för stort kommer vi i genomsnitt överskatta parametern och det räcker alltså inte enbart att ta hänsyn till utplaningspunkten vid val av t_0 .

Ytterligare simuleringar genomfördes för att se om detta beteende var unikt för Ripleys kantkorrigering som användes under simuleringsstudien. Figur 15 nedan visar att så inte är fallet.

De kantkorrigeringar som ses i figuren kallas gränskorrigering och translationskorrigering. Gränskorrigering innebär att man utesluter de punkter som ligger på ett avstånd mindre än t_0 från områdesgränsen till A. Translationskorrigering av K-funktionen ges av

$$\hat{K}_{trans}(t) = \frac{1}{\hat{\lambda}n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbf{1}\{d_{ij} \leq t\} \frac{|A|}{|A| - (x_i - x_j)} \quad (3.1)$$

där $\hat{\lambda}$ är någon uppskattning av λ .



Figur 15: $\hat{\kappa}$ med gränskorrigering (vänster) och translationskorrigering (höger).

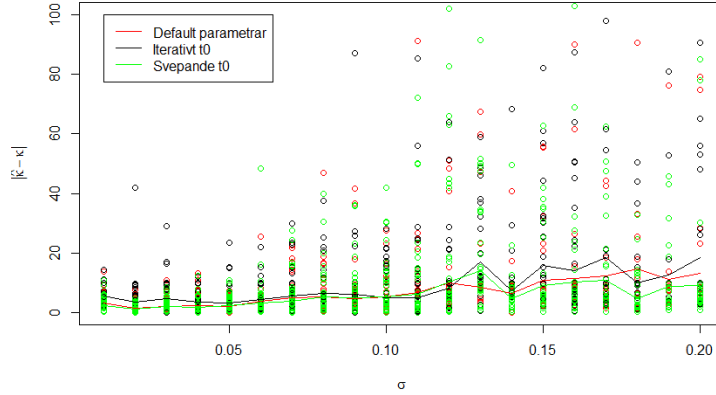
3.1.2 Val av t_0

För t_0 lägre än utplaningspunkten blir uppskattningen dålig, men ett högre t_0 tenderar att överskatta κ , således är det önskvärt att integrera upp till utplaningspunkten vid MCM. I appendix, A.1, finns en genom simulering framtagen tabell att tillgå där utplaningspunkten för olika värden av σ^2 och κ finns beräknade. Dessa är framtagna med 25 simuleringar per parameteruppsättning. Tanken är att den kan användas som vägledning vid val av t_0 . Observera att tabellen är beräknad utifrån simuleringar på en enhetskvadrat. För att översätta uppskattade parametrar beräknade i något annat område behöver vi skala $\hat{\kappa}$ med $\sqrt{|A|}$ och $\hat{\sigma}$ med $\frac{1}{\sqrt{|A|}}$.

För att använda tabellen behövs alltså κ och σ men generellt är dessa parametrar inte kända. En strategi för att lösa detta är att iterativt uppdatera t_0 efter $\hat{\sigma}$, $\hat{\kappa}$ och $\hat{\mu}$, tills dess att t_0 är tillräckligt nära det tidigare.

En annan strategi är att hitta utplaningspunkten genom att svepa över t_0 ovanifrån tills dess att $\hat{\kappa}$ lämnar ett band kring det första rimliga $\hat{\kappa}$. Därefter används ett t_0 lite över det hittade, då denna metod per konstruktion underskattar det lite.

Dessa strategier implementerades i R, kod finns i appendix. Dessa testades mot standardvärdet av t_0 genom att generera ett flertal realisationer av Thomasprocessen för ett par olika σ , därefter låta alla metoderna uppskatta κ och σ och mäta avvikelsen från sanna värdet. Resultatet av detta kan ses i figur 16.

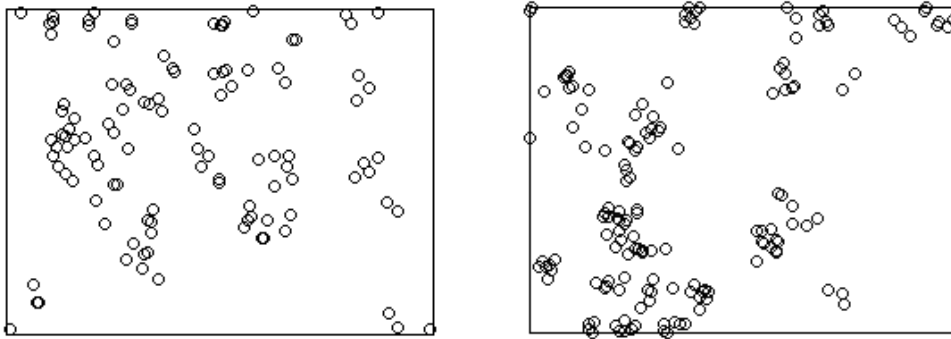


Figur 16: Fel av $\hat{\kappa}$ med olika val av t_0 , punkter representerar uppskattningar från enskilda utfall och linjerna representerar det logaritmiska medelvärdet med de värsta extremvärdena borttagna.

3.2 Nervdatan

Nervdatan består av 8 nervmönster från två patienter utan diabetesneuropati samt 6 nervmönster från två patienter med milda symptom. Vi refererar till dessa som grupp 1 respektive grupp 2.

Resultat från simulationsstudien ger oss riktlinjer i arbetet med att undersöka nervmönstren. Speciellt är det av intresse att se om Thomasprocessen beskriver nervdatan väl och hur metoderna som tagits fram i avsnitt 3.1 jämför sig med MCM med standardparametrar.



Figur 17: (Vänster) nervmönster från patient i grupp 1. (Höger) Nervmönster från patient i grupp 2. Möjligen går det att se en något mer klustrad fördelning i grupp 2.

3.2.1 Parameterskattningar på nervmönster

I vanliga fall går det inte att direkt jämföra parameterskattningar eftersom de sanna parametervärdena inte är kända. I detta fall finns dock information från nervmönstren att tillgå om vilka ändpunkter som hör ihop med respektive baspunkt. Med hjälp av detta kan vi beräkna mer tillförlitliga parametervärden och jämföra dem med våra egna uppskattningar.

För att uppskatta κ från nervmönstren kan vi helt enkelt använda

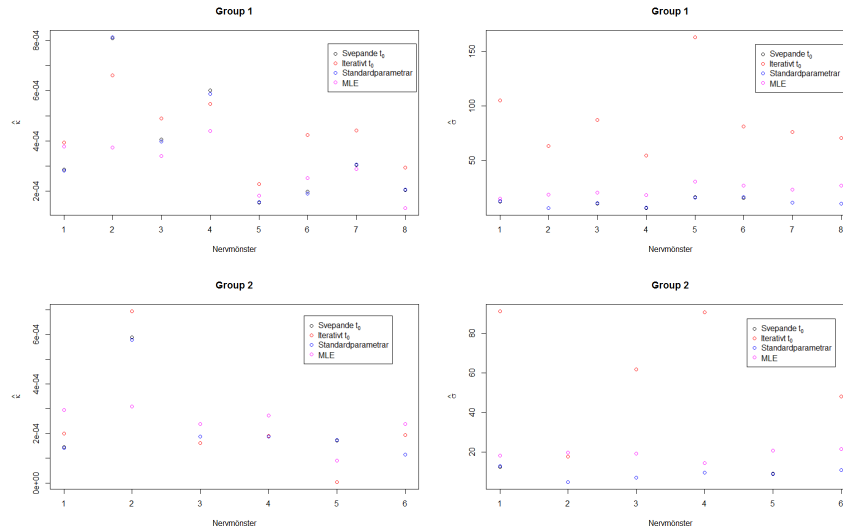
$$\hat{\kappa} = \frac{\#\text{baspunkter}}{|A|} \quad (3.2)$$

För σ utnyttjar vi ML-metoden och att ändpunkterna fördelar sig $N(0, \sigma^2 I)$ kring respektive baspunkt. Vi får då att

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n-1} \quad (3.3)$$

där x_i är avståndet mellan ändpunkt i och tillhörande baspunkt. Utförligare beräkning ges i appendix A.2.

Figur 18 visar parameterskattningarna på nervmönster från alla patienter från de två olika grupperna. Skattningarna är framställda med ML-metoderna ovan, MCM med standardparametrar och de två metoderna som presenterades i 3.1.2.



Figur 18: Parameterskattningar av nervmönster, grupp 1 och grupp 2.

För grupp 2, nervmönster 5, bör nämnas att skattningen av σ med iterativt t_0 togs bort då den var mer än tio gånger större än de andra skattningarna.

3.2.2 Anpassningstest på nervdatan

Vi applicerade ett anpassningstest från avsnitt 2.6 på de tillhandahållna nervmönstren. Detta gjordes med hjälp av simuleringar med `envelope` funktionen, där den övre integrationsgränsen t_0 i (2.19) här är vald som standardvärdet 0.25. Vi skapar en nollhypotes, H_0 , för varje observerat nervmönster och förkastar utifrån signifikansnivån $\alpha = 0.05$. Nollhypotesen är för varje enskilt nervmönster i båda grupperna: Nervmönstret är en realisation av den anpassade Thomasprocessen. Vi utför 499 simuleringar för båda teststatistikorna. Vi presenterar resultaten nedan i tabell 1 och 2. Utifrån dessa kan vi i samtliga fall inte förkasta nollhypotesen. Då flera tester utförs samtidigt, resulterar det i att nervmönster 7 ur grupp 1 i tabell 1 möjligtvis blivit signifikant enbart på chans, men då signifikansnivån behöver korrigeras för flera tester, är även detta ett ej signifikant resultat.

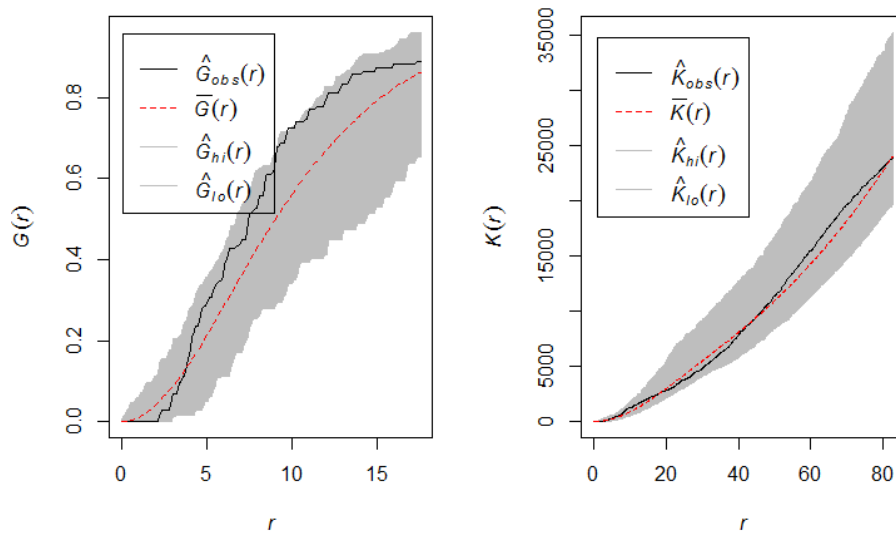
grupp 1	p-värde	grupp 2	p-värde
1	0.09	1	0.21
2	0.74	2	0.48
3	0.51	3	0.43
4	0.67	4	0.96
5	0.13	5	0.14
6	0.25	6	0.14
7	0.04		
8	0.89		

Tabell 1: P-värden för teststatistikan av G -funktionen

grupp 1	p-värde	grupp 2	p-värde
1	0.38	1	0.85
2	0.51	2	0.60
3	0.48	3	0.40
4	0.48	4	0.67
5	0.74	5	0.47
6	0.46	6	0.62
7	0.60		
8	0.59		

Tabell 2: P-värden för teststatistikan av K -funktionen

I figur 19 ses två figurer från `spatstat`-funktionen `envelope` som används i konstruktionen av p-värdena där det gråa området begränsas av det minsta respektive största uppskattade G - och K -funktionsvärdena.



Figur 19: Anpassningstest med G - och K -funktionen för nervmönster 7 i grupp 1.

4 Diskussion

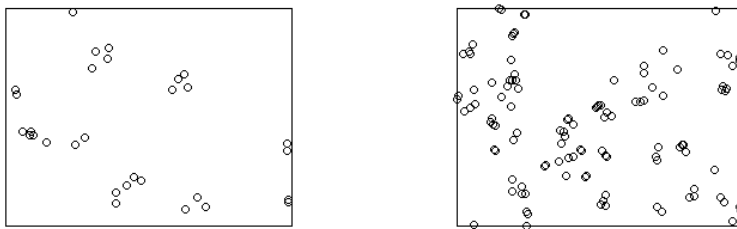
Som simuleringsstudien visat är valet av integrationsgräns i MCM viktigt för att få rimliga parameterskattningar eftersom resultaten varierar så pass kraftigt innan utplaningspunkten. Simuleringarna visade även att integrationsgränser större än utplaningspunkten tenderar att ge överskattade parametrar vilket också behöver tas i beaktning. Det är däremot viktigare att välja ett tillräckligt stort t_0 eftersom konsekvensen av att välja ett för stort t_0 inte är lika allvarligt som att välja ett för litet. Ett undantag är dock om σ^2 är stort då förvisso alla parametrar är svårskattade. Valet av kantkorrigering verkar också ha en viss inverkan vilket kan ses i figur 15 där gränskorrigering generellt ger skattningar med större varians.

Till en början verkade resultaten i simuleringsstudien tyda på att det var möjligt att hitta metoder för att härleda ett lämpligare t_0 än rekommenderade standardvärden och i förlängningen att förbättra parameterskattningarna. Det visade sig däremot att standardvärdena i MCM är väl valda eftersom resultaten inte skiljer sig nämnvärt från de metoder föreslagna i projektet. Dessutom kan uppskattningarna bli såpass dåliga att de marginella fördelarna är en onödig risk.

För att sammanfatta tillvägagångssättet för att välja t_0 rekommenderas att t_0 väljs så nära utplaningspunkten som möjligt. Det finns förvisso en del att förtydliga i detta påstående eftersom utplaningspunkten här inte har någon strikt definition. I detta projekt har den beskrivits med: det t_0 där parameterskattningarna planar ut och för figur 9 ser vi till exempel att åtagandet att bestämma exakt var utplaningspunkten ligger är subjektivt.

Avsnitt 3.1.2 beskriver två förslag på ickegrafiska metoder för att välja ett t_0 nära utplaningspunkten utifrån given data. Vi kan se i figur 18 att för $\hat{\kappa}$ håller sig alla metoder inom samma storleksordning medan $\hat{\sigma}$ kan skilja sig mer. Generellt verkar svepande t_0 och MCM med standardvärden hålla sig närmast skattningarna med ML-metoden vilket antyder att dessa metoder är mer tillförlitliga. Anledningen till detta är för att skattningarna med ML-metoden kan anses ligga nära de sanna parametervärdena. Ett problem med ML-metoden i detta fall är dock att ändpunkternas sanna föräldrapunkt snarare ligger vid första förgreningen efter baspunkten och är således lite förskjuten från baspunkten. Detta resulterar i att avstånden blir större och förklarar varför de uppskattade värdena av σ är något större än de från MCM.

Som nämnts tidigare har skattningen av σ för nervmönster 5 i grupp 2 tagits bort eftersom den var flera storleksordningar större då vi använde iterativt t_0 . Detta är ett exempel på hur illa det kan gå då vi går ifrån standardvärden i MCM. En trolig anledning till detta är att nervmönstret i detta fall har färre punkter (se figur 20) och därmed har svårskattade parametrar. Det går även att föreställa sig att om den iterativa metoden börjar med en dålig skattning leder det i sin tur till att metoden itererar i fel riktning och gör kommande skattningar dåliga. Koden i appendix B.2 innehåller försök att förhindra detta men är inte alltid tillförlitliga.



Figur 20: Nervmönster 5 (vänster) och nervmönster 2 från grupp 2.

Anpassningstest som utfördes på nervmönstren visade att nollhypotesen inte kunde förkastas i samtliga fall. Det var enbart ett test som låg under $\alpha = 0.05$, men efter korrigering för flera tester förblir även det icke signifikant. Det faktum att nollhypotesen inte kan förkastas

för någon av grupperna tyder på att båda grupperna kan beskrivas med Thomasprocessen. Däremot säger det inget om att de patienter med diabetesneuropati har mer klustrade nervmönster än friska patienter. Därför styrker inte resultaten tidigare forskningsresultat men reducerar dem inte heller. De skattade parametrarna i figur 18 skiljer sig förvisso mellan de olika grupperna men rör sig kring samma storleksordning vilket gör det svårt att säga om parametrarna generellt skiljer sig åt gruppvis. För det här fallet presenteras medelvärdet av parameterskattningarna från de två grupperna i tabellerna nedan för att få en tydligare uppfattning om hur grupperna skiljer sig åt. En viss skillnad kan ses där grupp 2 har mindre parametervärden där ett mindre σ tyder på att nervmönstren i grupp 2 är aningen mer klustrade. Däremot behövs andra statistiska metoder för att med säkerhet fastställa att det verkligen är så. Hypotestest för att undersöka om skillnaden i parametervärdena mellan de två grupperna är signifikant är ett tillvägagångssätt. Resultaten från ett sådant test bör i detta fall ifrågasättas då urvalspopulationen är så pass liten.

	$\hat{\kappa}$	$\hat{\sigma}$	$\hat{\mu}$
Svepande t_0	3.70×10^{-4}	11.6	3.21
Iterativt t_0	4.35×10^{-4}	87.7	2.26
Standardparametrar	3.67×10^{-4}	11.7	3.24
MLE	2.98×10^{-4}	22.9	3.37

Tabell 3: Medelvärdet av parameterskattningar, grupp 1

	$\hat{\kappa}$	$\hat{\sigma}$	$\hat{\mu}$
Svepande t_0	2.32×10^{-4}	9.0	3.79
Iterativt t_0	2.40×10^{-4}	-	11.2
Standardparametrar	2.30×10^{-4}	9.04	3.83
MLE	2.41×10^{-4}	19.0	3.37

Tabell 4: Medelvärdet av parameterskattningar, grupp 2

Denna studie har fokuserat på att studera den övre integrationsgränsen i MCM med realisationer av Thomasprocessen. Projektet kan lätt utökas med ytterligare undersökningar, till exempel hur olika val av viktfunktion $w(t)$ eller konstanter p och c påverkar MCM-skattningar. Att studera ett annat val av klusterprocess är också ett alternativ där Matern-processen föreslagits. Denna process påminner om Thomasprocessen men skiljer sig med hur dess dotterpunkter fördelar sig kring respektive föräldrapunkt. För att se om någon av de olika processerna är att föredra går det att utnyttja `envelope`-funktionen för att se hur de olika processerna håller sig inom de punktvisa testerna.

Då projektet har varit simuleringsinriktad är det önskvärt att, om möjligt, stärka resultaten med teoretiska studier rörande valet av integrationsgräns. Till exempel verkar det finnas en relation mellan utplaningspunkten och σ^2 som skulle kunna vara en möjlig ingångspunkt vid teoretiska studier.

Referenser

- [1] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015.
- [2] Peter J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. Chapman and Hall/CRC, 3rd edition, 2013.
- [3] RM Fewster, BC Stevenson, and DL Borchers. Trace-contrast models for capture-recapture without capture histories. *STATISTICAL SCIENCE*, 31(2):245–258, 2016.
- [4] Janine Illian, Helga Stoyan, Dietrich Stoyan, and Antti Penttinen. *Statistical analysis and modelling of spatial point patterns*. John Wiley, 2008.
- [5] Jesper Møller and Rasmus P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, 2004.
- [6] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2):255–266, 1976.
- [7] Lance A. Waller, Aila Särkkä, Viktor Olsbo, Mari Myllymäki, Ioanna G. Panoutsopoulou, William R. Kennedy, and Gwen Wendelschafer-Crabb. Second-order spatial analysis of epidermal nerve fibers. *Statistics in Medicine*, 30(23):2827–2841, 2011.

A Appendix A

A.1 Utplaningspunkt

Nedan följer empiriskt framtagna utplaningspunkter för olika parameteruppsättningar.

$\kappa \setminus \sigma$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
1	$\frac{1}{50}$	$\frac{1}{10}$	$\frac{2}{25}$	$\frac{7}{50}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{9}{50}$	$\frac{13}{50}$	$\frac{3}{10}$	$\frac{9}{25}$	$\frac{2}{5}$	$\frac{23}{50}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{27}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
3.04	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{7}{50}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{3}{10}$	$\frac{6}{25}$	$\frac{13}{50}$	$\frac{8}{25}$	$\frac{27}{50}$	$\frac{21}{50}$	$\frac{17}{50}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
5.08	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{1}{5}$	$\frac{4}{25}$	$\frac{6}{25}$	$\frac{8}{25}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{21}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{29}{50}$
7.12	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{1}{5}$	$\frac{9}{50}$	$\frac{1}{5}$	$\frac{11}{50}$	$\frac{13}{50}$	$\frac{3}{10}$	$\frac{8}{25}$	$\frac{3}{5}$	$\frac{9}{50}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
9.16	0	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{9}{50}$	$\frac{9}{50}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{13}{50}$	$\frac{7}{25}$	$\frac{11}{25}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{19}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
11.20	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{4}{25}$	$\frac{11}{50}$	$\frac{9}{50}$	$\frac{13}{50}$	$\frac{17}{50}$	$\frac{2}{5}$	$\frac{27}{50}$	$\frac{29}{50}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
13.25	0	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{7}{50}$	$\frac{7}{50}$	$\frac{7}{50}$	$\frac{9}{50}$	$\frac{6}{25}$	$\frac{6}{25}$	$\frac{3}{10}$	$\frac{14}{25}$	$\frac{13}{25}$	$\frac{17}{50}$	$\frac{29}{50}$	$\frac{27}{50}$	$\frac{3}{5}$	$\frac{3}{5}$
15.29	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{9}{50}$	$\frac{9}{50}$	$\frac{6}{25}$	$\frac{11}{25}$	$\frac{3}{10}$	$\frac{8}{25}$	$\frac{3}{5}$	$\frac{13}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
17.33	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{9}{50}$	$\frac{7}{25}$	$\frac{13}{25}$	$\frac{13}{50}$	$\frac{29}{50}$	$\frac{29}{50}$	$\frac{11}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
19.37	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{4}{25}$	$\frac{7}{50}$	$\frac{14}{25}$	$\frac{6}{25}$	$\frac{17}{25}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$
21.41	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{1}{5}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{19}{50}$	$\frac{13}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
23.45	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{9}{50}$	$\frac{1}{5}$	$\frac{9}{50}$	$\frac{13}{50}$	$\frac{3}{50}$	$\frac{23}{50}$	$\frac{17}{50}$	$\frac{3}{5}$	$\frac{27}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
25.50	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{9}{50}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
27.54	0	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{11}{50}$	$\frac{4}{25}$	$\frac{27}{50}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
29.58	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{13}{50}$	$\frac{3}{50}$	$\frac{21}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
31.62	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{12}{25}$	$\frac{7}{25}$	$\frac{13}{50}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
33.66	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{9}{50}$	$\frac{11}{50}$	$\frac{3}{5}$	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
35.70	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{14}{25}$	$\frac{6}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
37.75	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{7}{50}$	$\frac{11}{50}$	$\frac{19}{50}$	$\frac{11}{50}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
39.79	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{9}{50}$	$\frac{7}{50}$	$\frac{3}{25}$	$\frac{9}{50}$	$\frac{12}{25}$	$\frac{6}{25}$	$\frac{3}{5}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
41.83	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{7}{50}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{23}{50}$	$\frac{14}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
43.87	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{2}{25}$	$\frac{7}{50}$	$\frac{9}{50}$	$\frac{4}{25}$	$\frac{11}{50}$	$\frac{9}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
45.91	0	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{50}$	$\frac{1}{10}$	$\frac{7}{50}$	$\frac{7}{50}$	$\frac{4}{25}$	$\frac{27}{50}$	$\frac{9}{50}$	$\frac{13}{25}$	$\frac{3}{5}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
47.95	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{3}{50}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{1}{50}$	$\frac{1}{5}$	$\frac{11}{50}$	$\frac{13}{25}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
50.00	0	$\frac{1}{50}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{1}{10}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{11}{50}$	$\frac{29}{50}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$

Tabell 5: Observera att t_0 bör väljas en aning större än vad som angivits i tabellen. Beräkningarna underskattar nämligen utplaningspunkten en aning per konstruktion.

A.2 Maximum likelihood estimation av σ^2

Vi antar att varje punkt x_i ur mönstret är oberoende och normalfördelade, $\mathcal{N}(0, \sigma^2)$

$$\begin{aligned}
 L(\theta) &= f(x_1, \dots, x_n | \theta) \\
 &= \prod_{i=1}^n f(x_i | \theta) \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}} \\
 &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}}.
 \end{aligned}$$

Den naturliga logaritmen används på $L(\theta)$ för ett mer lätthanterligt uttryck.

$$\begin{aligned}
 l(\theta) &= \ln L(\theta) \\
 &= \frac{-n}{2} (\ln 2\pi + \ln \sigma^2) - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2}
 \end{aligned}$$

Vi tar den partiella derivatan på σ och sätter den lika med noll.

$$0 = \frac{\partial l(\theta)}{\partial \sigma} = \frac{\sum_{i=1}^n x_i^2}{\sigma^3} - \frac{n}{\sigma}$$

Detta ger oss uppskattningen $\hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n-1}$, där vi multiplicerat $\hat{\sigma}^2$ med $\frac{n-1}{n}$ för att korrigera för förväntnings skevheten.

A.3 χ^2 -fördelning med två frihetsgrader

Med k stycken frihetsgrader är fördelningsfunktionen för en χ^2 fördelad variabel,

$$F(x, k) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})}. \quad (\text{A.1})$$

Där $\gamma(\frac{k}{2}, \frac{x}{2})$ är den undre ofullständiga gamma funktionen,

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt. \quad (\text{A.2})$$

Med $k = 2$ blir (A.2)

$$\gamma\left(1, \frac{x}{2}\right) = \int_0^{\frac{x}{2}} e^{-t} dt.$$

Uttrycket i nämnaren är gamma funktionen med ett heltal som argument, $\Gamma(1) = (1-1)! = 1$, således blir (A.1)

$$F(x, 2) = \int_0^{\frac{x}{2}} e^{-t} dt = 1 - e^{-x/2}.$$

B Appendix B

B.1 Svepande t_0

```

sigmaavt0<-function(dat,plot=FALSE,print=FALSE,tsteps=100){
#Indata:
#dat är ett punktmönster innehållande datan vi vill studera
#plot är en boolean för huruvida vi vill ha en plot av kappa/sigmahatt
# mot integrationsgränsen
#tsteps är hur många tidssteg man vill ta,

#Först tittar vi på geometrin från datan och tar reda på skalfaktorn
#geo och maximala t0 som Khatt är definerat för.
x=dat$window$ xrange
x=abs(x[[1]]-x [[2]])
y=dat$window$ xrange
y=abs(y[[1]]-y [[2]])
geo=sqrt(x*y)
tmax=sqrt(x^2+y^2)/2.01
#2.01 för att vi vill vara lite mindre än hälften för att undvika
#numeriskt strul

#Beräkna Khatt så att det inte behöver göras om varje gång.
khatt=Kest(dat,rmax=tmax)
#Initiera lite variabler som kommer behövas
#Här sparas kappahatt
Kappas=rep(0,tsteps)
#Här sparas sigmahatt
Sigmas=rep(0,tsteps)
#Vektor med integrationsgränserna
t0s=seq(0,tmax,tmax/(tsteps-1))
#Den här kommer bli true när vi får första rimliga kappat

```



```

rimligStart=FALSE
#Index i t0s för utplaningspunkten kommer sparas här
Utplaningsindex=0

#Loopa igenom integrationsgränserna uppifrån.
for(n in tsteps:1){
  #Tills dess att vi hittat ett rimligt kappahatt körs den här delen
  if(!rimligStart){
    #Använd MCM och spara resultatet.
    res=thomas.estK(khatt,rmax=t0s[[n]])
    Kappas[[n]]=res$modelpar[[1]]
    Sigmas[[n]]=res$modelpar[[2]]
    #Vi kollar om kappahatt är rimligt.
    #Om det är det så körs andra halvan av loopen framöver
    #Vi definerar också ett intervall utanför vilket
    #det inte är utplanat längre.
    if(Kappas[[n]]<100*x*y&&Kappas[[n]]>x*y){
      rimligStart=TRUE
      lowerBound=Kappas[[n]]/2
      UpperBound=2*Kappas[[n]]
    }
    #När vi hittat ett rimligt kappahatt körs den här biten i loopen.
  } else {
    #Vi använder MCM med tidigare sigmahatt och kappahatt
    #som startvärde för optimering.
    P=c(kappa=Kappas[[n+1]],scale=sqrt(Sigmas[[n+1]]))
    res=thomas.estK(khatt,startpar = P,rmax=t0s[[n]])
    Kappas[[n]]=res$modelpar[[1]]
    Sigmas[[n]]=res$modelpar[[2]]
    #Om vi är utanför intervallet för första gången
    #har vi hittat utplaningspunkten
    urIntervall=(Kappas[[n]]>UpperBound||Kappas[[n]]<lowerBound)
    if(Utplaningsindex==0&&urIntervall){
      #Spara UP och tillhörande kappas
      Utplaningsindex=n;
      UP=t0s[[n]];
      kappaUP=Kappas[[n]];
      #Skifta tillbaka UP lite då metoden naturligt överskattar.
      UPmod=t0s[[n+tsteps/20]];
      kappaUPmod=Kappas[[n+tsteps/20]];
      #Här kan man bryta loopen om vi inte vill plotta
    }
  }
}

#Om man vill ha en plot plottas kappahatt mot t0, med utplanat
#intervall och utplaningspunkten.
if(plot){
  plot(t0s,Kappas,type="l",ylim=c(0,UpperBound*2))
  lines(t0s,rep(lowerBound,tsteps),col='red')
  lines(t0s,rep(UpperBound,tsteps),col='red')
  if(Utplaningsindex!=0){
    points(UP,kappaUP,col="magenta")
    points(UPmod,kappaUPmod,col="green")
  }
}
#Returnera kappa från t0 lite större än UP,

```

```

#då metoden naturligt underrskattar UP något.
if (Utplaningsindex!=0){
  return(c(kappa=kappaUPmod,scale=UP/2))
} else {
  #Om vi inte hitta någon UP använd default t0
  i=min(x,y)/4/tmax*tsteps
  return(c(kappa=Kappas[[i]],scale=Sigmas[[i]])
)
}
}

```

B.2 Iterativt t_0

```

thomparam <- function(dat,plot=FALSE,print=FALSE,maxit=15){
#Indata:
#dat är ett punktmönster innehållande datan vi vill studera
#plot är en boolean för huruvida vi vill plotta ut de t0 som testas
#print är huruvida vi vill printa vissa värden (för fels ökning)
#maxit är maximala antal iterationer, lågt då metoden oftast testades
#    på många punktmönster åt gången.

#Först tittar vi på geometrin från datan och tar reda på skalfaktorn
#geo och maximala t0 som Khatt är definerat för.
x=dat$window$xrange
x=abs(x[[1]]-x [[2]])
y=dat$window$yrange
y=abs(y[[1]]-y [[2]])
geo=sqrt(x*y)
tmax=sqrt(x^2+y^2)/2.01
#2.01 för att vi vill vara lite mindre än hälften för att undvika
#numeriskt strul

#Khatt sparas så det ej behöver räknas ut varje gång
khatt=Kest(dat,rmax=tmax)

#t0 från linjärisering av framtagna tabell. Stämmer bra om kappa och my inte är för
små.
teot0 <- function(param){
  return(1.936*param$modelpar[[2]]-0.0005*param$modelpar[[1]]/geo-0.0013*param$modelpar[[3]]/geo)
}

#Testa default värdet.
t0=geo/4
param=thomas.estK(khatt,lambda = dat$n,rmax=t0)
if (plot){#omplot plotta kappahatt mot t0
  points(param$ctrl$rmax,param$modelpar[[1]],col='red')
}
#Iterationsräknare
it=1;
#Om vi inte började med rimligt sigma så testas vi några t0 till :
if (2*param$modelpar[[2]]>tmax){
  sp=c(kappa=param$modelpar[[1]], scale=param$modelpar[[2]])
  t0=3*tmax/4
  param=thomas.estK(khatt,startpar=sp,rmax=t0,control=c(maxit=1000),lambda=dat$n)
  if (plot){
    points(param$ctrl$rmax,param$modelpar[[1]],col='red')
  }
}
}

```

```

}
if(2*param$modelpar[[2]]>tmax){
  sp=c(kappa=param$modelpar[[1]], scale=param$modelpar[[2]])
  t0=tmax/4
  param=thomas.estK(khatt,startpar=sp,
    rmax=t0,control=c(maxit=1000),lambda=dat$n)
}
if(plot){
  points(param$ctrl$rmax,param$modelpar[[1]],col='red')
}
#Om sigma fortfarande är dåligt ger vi upp och returnerar det vi hade.
if(2*param$modelpar[[2]]>tmax){
  if(print){
    print("Bad sigma approximation")
  }
  return(param)
}
#Sen itererar vi antingen till vi nått maximalt antal iterationer
#Eller vi får ett t0 som stämmer bra överens med tidigare parametrar
while (it < maxit && (t0 < teot0(param) || t0 > (teot0(param) + 0.05 * geo))) {
  #Om vi har rimliga kappa och my använder vi linjärisering av
  #matrisen för t0.
  if (param$modelpar[[1]]/x*y > 3 && param$modelpar[[3]]/x*y > 3) {
    t0 = teot0(param)
  } else {
    # Annars tar vi lite mer än dubbla sigma.
    t0 = 2.5 * param$modelpar[[2]]
  }

  #Vid felsökning printar vi t0
  if (print && it == 1) {
    print(c("1:t0", t0))
  }

  #Vi kollar så att t0 är inom möjliga intervallet
  if (t0 > 0 && t0 <= tmax) {
    sp = c(kappa = param$modelpar[[1]], scale = param$modelpar[[2]])
    param = thomas.estK(khatt, startpar = sp,
      rmax = t0, control = c(maxit = 1000), lambda = dat$n)
  } else {
    #Annars tar vi ett slumpat t0 och försöker igen
    t0 = min(max(geo/8, geo/4 + geo/8 * rnorm(1)), tmax)
    param = thomas.estK(khatt, rmax = t0, control = c(maxit = 1000), lambda = dat$n)
  }

  it = it + 1
  #Plotta punkterna om plot är på.
  if (plot) {
    points(param$ctrl$rmax, param$modelpar[[1]], col = 'red')
  }
  #Skriv ut parametrar vid felsökning
  if (print) {
    print(c(t0, param$modelpar[[2]], param$modelpar[[1]]))
  }
}
#Om vi felsöker vill vi veta om max iterationer nåddes
if (it >= maxit && print) {

```

```
    print("För många iterationer")
}
#Returnera senaste parametrarna
return(param)
}
```
