

1171

Cox-prosessi pistekuvion hierarkkisena
tilastollisena mallina

Anne Riiali

5.10.1998

Pro gradu -tutkielma
Jyväskylän yliopisto
Tilastotieteen laitos

Tiivistelmä

Anne Riiali

Cox-prosessi pistekuvion hierarkkisena tilastollisena mallina.

Tilastotieteen pro gradu -tutkielma, Jyväskylän yliopisto, 1998.

72 sivua + 9 sivua liitteitä.

Tämän tutkimuksen sisällöllinen ongelma on metsäalueen puutiheyden kuvaaminen spatiaalisen pisteprosessin intensiteetin avulla. Työssä käsitellään sekä kenttähavainnointiin että täydellisiin tilajärjestyskarttoihin perustuvia tilastomenetelmiä. Sovellusaineistoina on kaksi metsän rakennetta kuvaavaa aineistoa, joista toinen on luonnonmetsän ja toinen talousmetsän puiden sijaintikartta. Molemmat metsät ovat Helsingin yliopiston metsäntutkimusaseman koemetsiä Etelä-Suomessa.

Metsien spatiaalista rakennetta tutkittiin erilaisten tilajärjestystä karakterisoivien tunnuksien avulla. Intensiteetin estimoinnissa käytettiin sekä klassista kernel-menetelmää että bayesiläistä lähestymistapaa. Bayesiläisessä mallinnuksessa lähtökohtana oli yksinkertainen gamma-Poisson-prosessi. Spatiaalisen riippuvuusrakenteen aikaansaamiseksi mallia parannettiin tekemällä priorina olevaan satunnaiskenttään kernel-tasoitus. Posteriorijakauma laskettiin käyttämällä Markovin ketjun simulointiin perustuvaa Monte Carlo -menetelmää (MCMC). Laskenta-algoritmistista tehtiin c-kielinen ohjelma. Myös puutiheydessä esiintyvää trendiä tarkasteltiin.

Metsien spatiaaliset rakenteet olivat odotetusti erilaiset. Puut sijaitsivat talousmetsässä melko säännöllisesti pientä aluetta lukuun ottamatta ja luonnonmetsässä likimain satunnaisesti. Puutiheydessä ei ollut havaittavissa selvää trendiä kummassakaan metsässä. Estimoituja intensiteettejä vertailtiin pääasiassa visuaalisesti. Molempien metsien kohdalla riippumattomuusmalli johti sekä liian suuriin intensiteetteihin että intensiteetin vaihteluihin. Spatiaalisen rakenteen mukaan ottaminen otti huomioon tiheydelle tyypillisen kontekstuaalisuuden, jolloin myös estimaatit paranivat.

Tutkimus on toteutettu osana Suomen Akatemian projekteja ”Spatiaalisen ympäristöaineiston analyysi” ja ”Paikkatiedon mallintaminen ja GIS” ja se voidaan luokitella paikkatietojärjestelmäperusteiseksi tilastoanalyysimenetelmien kehitystyöksi.

Avainsanoja: spatiaalinen pisteprosessi, intensiteetin estimointi, Bayesmenetelmä, MCMC, metsäntutkimus.

Sisältö

1	Johdanto	1
2	Tutkimusongelman ja -aineiston esittely	5
2.1	Ongelmanasettelu	5
2.2	Tutkimusaineisto	6
3	Spatiaalinen pisteprosessi	10
3.1	Käsitteitä	10
3.2	Etäisyystarkasteluja	14
3.2.1	Kahden pisteen välisen etäisyyden jakauma	14
3.2.2	Lähinaapurietäisyys	15
3.2.3	Etäisyys otantapistteestä lähimpään prosessin pisteeseen	15
3.3	Homogeeninen Poisson-prosessi	16
3.4	Epähomogeeninen Poisson-prosessi	20
3.5	Cox-prosessi	20
3.5.1	Matérnin klusteriprosessi	22
3.5.2	Log-gaussinen Cox-prosessi	23
3.5.3	Kernel-tasoitettu gamma-Poisson-prosessi	23
3.5.4	Muita Cox-prosesseja	26
3.6	Pisteprosessien simulointialgoritmeja	27
3.6.1	Homogeenisen Poisson-prosessin simulointi	27
3.6.2	Epähomogeenisen Poisson-prosessin simulointi	28
3.6.3	Cox-prosessi	29
3.7	Bayes-menetelmä	30
4	Kenttähavainnoinnin menetelmät intensiteetin estimoinnissa	33
4.1	Yleistä	33
4.2	Klassinen intensiteetin estimointi	36
4.2.1	Lukumäärämenetelmä	36
4.2.2	Etäisyysmenetelmä	37
4.3	Riippuvuuden huomioon ottaminen -satunnaiskenttämenetelmä	38
4.4	Hierarkkinen mallintaminen	39
4.4.1	Hierarkkinen mallinnus lukumäärähavainnoille	41
4.4.2	Hierarkkinen mallinnus etäisyshavainnoille	45

5	Karttamuotoisen aineiston intensiteetin estimointi	49
5.1	Intensiteetin klassinen estimointi	49
5.2	Intensiteetin bayesiläinen estimointi	50
5.2.1	Yksinkertainen Bayes-malli	51
5.2.2	Yksinkertainen Bayes-malli kontekstuaalisuudelle	53
6	Tutkimusaineistoon liittyviä empiirisiä tarkasteluja	57
6.1	Tilajärjestyksen tarkastelu	57
6.2	Trendin tarkastelu	58
6.3	Intensiteetin estimointi	61
6.3.1	Klassinen estimointi	61
6.3.2	Bayesiläinen estimointi	62
6.4	Johtopäätökset	65
	Kiitosmaininnat	69
	Kirjallisuutta	70
	Liite 1. Simulointialgoritmeja	73
	Liite 2. Trendin variogram-tarkasteluja	75
	Liite 3. Metropolissh–Hastings-algoritmi	77

1 Johdanto

Tutkimuskohteena on tason pistekuviot, joiden tilastollisena mallina käytetään spatiaalisia pisteprosesseja. Tavoitteena on kuvata pistetiheyden vaihtelua intensiteetin avulla. Tutkimuksessa käsitellään sekä kenttähavainnointiin että täydellisiin tilajärjestyskarttoihin perustuvia tilastomenetelmiä. Empiirisenä aineistona on kaksi metsän rakennetta kuvaavaa aineistoa, joista toinen on luonnonmetsän ja toinen talousmetsän puiden sijaintikartta.

Intensiteetti on pisteprosessin ensimmäisen kertaluvun ominaisuus, ja sillä kuvataan runsautta. Esimerkiksi metsätieteessä intensiteettiä käytetään puiden tiheyden ja sen vaihtelun kuvaamiseen. Intensiteetti ilmoittaa prosessin pisteiden keskimääräisen lukumäärän pinta-alayksikköä kohti. Matemaattisesti intensiteetti määritellään intensiteettifunktion avulla siten, että tapausten odotettu lukumäärä alueella B voidaan esittää integraalina $\int_B \lambda(x)dx$, jossa $\lambda(\cdot)$ on intensiteettifunktio. Stationaarisille prosesseille intensiteettifunktio on vakio, jolloin voidaan puhua pistetiheydestä. Silloin tiheys kerrottuna alueen pinta-alalla antaa tapausten odotetun lukumäärän. Epästationaarisille prosesseille tiheyttä vastaa integraali $\int_B \lambda(x)dx$ jaettuna alueen pinta-alalla.

Pisteprosessin tilajärjestystä, jolla tarkoitetaan pisteiden suhteellista sijoittumista alueella, voidaan tutkia pisteprosessin toisen kertaluvun ominaisuuksien avulla. Niitä ovat mm. toisen kertaluvun intensiteetti ja K -funktio, ja ne kuvaavat pisteparien välistä spatiaalista riippuvuutta. Muita tilajärjestyksen karakterisointeja ovat mm. lähinaapurietäisyys ja tyhjän tilan tunnusluku.

Intensiteetin mallintaminen ja perinteinen tilajärjestyksen tilastollinen analyysi eroavat toisistaan sekä oletuksiltaan että tarkoituksiltaan. Tilajärjestyksen analyysi perustuu stationaarisuusoletukseen, kun taas intensiteetin mallintamisessa lähtökohtana on oletus intensiteetin vaihtelusta tarkasteltavassa alueessa ja sen selittäminen. Tilajärjestysanalyysillä pyritään selvittämään, esiintyykö pistekuviossa tilastollista säännönmukaisuutta kuten säännöllisyyttä tai klusteroitumista ja missä mittakaavassa sitä on havaittavissa. Intensiteetin mallintamisessa tavoitteena ei ole riippuvuusrakenteen analyysi vaan tiheyden vaihtelun tutkiminen. Mikäli prosessi on epästationaarinen, perinteiset tilajärjestystarkastelut eivät ole mielekkäitä. Epästationaarista prosessia voidaan usein kuitenkin pitää lokaalisti stationaarisena, jolloin $\lambda(x)$ tulkitaan lokaaliksi intensiteetiksi pisteessä x .

Tutkittavat objektit voidaan havaita joko täydellisellä havainnoinnilla tai kenttähavainnoinnilla. Täydellisessä havainnoinnissa tutkittavan alueen kaikista objekteista tehdään tilajärjestyskartta. Jos tarkasteltava alue on laaja, tehdään analysointi alueelta poimitun otoksen avulla, jolloin puhutaan kenttähavainnoinnista. Kenttähavainnoinnissa tarkastellaan useita tutkimusalaan suhteutettuna pieniä osa-alueita, joiden valinta voi perustua otantaan. Klassisesti pyritään harvaan otantaan, jolloin koealoissa tehdyt havainnot voidaan olettaa riippumattomiksi. Harvan otannan tapauksessa pisteprosessin analysoinnissa käytetään perinteisiä lukumäärä- ja etäisyysmenetelmiä.

Jos koealat on poimittu esimerkiksi ryväotannalla, eivät harvan otannan oletukset ole voimassa. Tällöin koealoissa tehtyjä havaintoja ei voida pitää riippumattomina. Koska havaintojen välinen riippuvuus vaikuttaa mm. intensiteetin estimaattorin tilastollisiin ominaisuuksiin, on se otettava huomioon. Tämä voidaan tehdä kahdella tavalla. Satunnaiskenttämenetelmässä riippuvuus mallinnetaan kovarianssin avulla. Toinen, sovelluksissa käytökelpoinen tapa on käyttää hierarkkista mallinnusta. Se voidaan rinnastaa varianssikomponenttimalliin, jossa havaintojen välistä sisäkorrelaatiota approksimoidaan hierarkkisen rakenteen avulla.

Intensiteetin estimointi klassisen tilastotieteen periaattein johtaa satunnaiskentän parametrien estimointiin (Ripley, 1981) tai epäparametrinen menetelmä käyttäen intensiteetin kernel-estimointiin (Silverman, 1986). Viimeksi mainitussa lähestymistavassa intensiteettiä estimoidaan funktiolla

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K \left\{ \frac{1}{h}(x - X_i) \right\},$$

jossa K on kernel-funktio. Vaikka kernel-estimoinnin ongelmaksi muodostuu

yleensä ytimen ikkunaleveyden h valinta sekä tilanne, jossa pisteet ovat harvassa, menetelmä on käyttökelpoinen ja yleisesti hyväksytty eksploratiivinen työkalu. Parametrisessa lähestymistavassa pisteaineistoon sovitetaan pinta, joka on usein polynomi

$$\lambda(x, y) = a_0 + a_1x + a_2y + a_{11}x^2 + a_{22}y^2 + a_{12}xy + \dots$$

Vaihtoehtoisesti voidaan käyttää semiparametrisia menetelmiä esimerkiksi spline-funktioita soveltaen (Green & Silverman, 1994).

Bayes-menetelmän käyttö intensiteetin estimoinnissa on houkutteleva vaihtoehto, sillä se mahdollistaa paitsi intensiteetin estimoinnin niin myös intensiteetin estimaattorin satunnaisvaihtelun kuvaamisen sekä vaihtelua kuvaavien todennäköisyyksien laskemisen. Lisäksi pisteprosessin realisaatioiden simulointi voidaan tehdä ennustusjakaumaa käyttäen. Bayes-lähestymistavassa intensiteetin määräämä satunnaiskenttä ajatellaan intensiteetin priorijakumaksi ja huomio kohdistuu posteriorijakauman estimointiin.

Tutkimusaineistona on kaksi rakenteeltaan erilaista metsäaluetta Etelä-Suomesta. Toinen metsä edustaa hoidettua talousmetsää, jonka käsittelyn päämääränä on ollut maksimaalinen puun nettotuotanto. Toinen metsä on puolestaan luonnonmetsää, jossa harvennushakkuita ei ole tehty. Tutkittavana on ollut molemmissa metsissä 0.25 hehtaarin alue (50m \times 50m), ja kaikki vähintään 1.3 metriä korkeat puut on rekisteröity. Aineisto on muotoa $\{[x_n; m_n]\}$, jossa x kuvaa puun n maantieteellistä sijaintia ja m sisältää puun ominaisuuksia kuvaavat muuttujat, merkit. Näitä ovat puun pituus, halkaisija, puulaji ja puun tila (kuollut/elävä). Molemmissa metsissä valtaosa puista on kuusia. Aineistot eivät sisällä sijainnista riippuvaa kovariaattitietoa.

Spatiaalisten pisteprosessien tilastollinen analysointi ja mallintaminen homogeenisen aineiston tapauksessa tunnetaan hyvin. Heterogeenisten aineistojen yleistyessä on löydettävä keinoja heterogeenisuuden mallintamiseen. Heterogeenisuudella ymmärretään tässä yhteydessä epähomogeenisuutta sekä tiheydessä että riippuvuusrakenteessa. Niinpä keskimääräisen intensiteetin sijasta onkin mielekkäämpää tarkastella, kuinka intensiteetti alueella vaihtelee. Aluekohtaisen tiedon saamiseksi käytetään riippuvuusrakenteen kuvaamisessa hierarkkista mallia. Lähestymistapaa ovat esittäneet erilaisten Cox-prosessien yhteydessä Møller, Syversveen & Waagepetersen (1998), Ickstadt & Wolpert (1996) sekä Heikkinen & Arjas (1998). Sen sijaan kenttähavainnointiin perustuvia hierarkkisia malleja ei kirjallisuudesta löytyne.

Työn tarkoitus on ensin esitellä spatiaalisten pisteprosessien yleistä teoriaa sekä siihen perustuvia tilastomenetelmiä ja lopuksi soveltaa näitä käytäntöön. Lukija motivoidaan ensin luvussa 2 tutkimusongelman esittelyllä ja alustavalla aineiston tarkastelulla. Tutkimukseen liittyvä pisteprosessiteoria on esitelty luvussa 3. Ensimmäin on esitelty pisteprosesseihin liittyviä yleisiä käsitteitä ja analysointimenetelmiä, jotka perustuvat sekä lukumääriin että etäisyystarkasteluihin. Luvussa on määritelty yleisimpiä pisteprosesseja ja tarkasteltu niiden ominaisuuksia. Esimerkkeinä Cox-prosessin konstruktioista ovat Matérnin klusteriprosessi, log-gaussinen Cox-prosessi ja gamma-Poisson-rasteriprosessit. Simulointialgoritmeja on tarkasteltu lähinnä Poisson-prosessin tapauksessa. Koska jatkossa pisteprosessin analysoinnissa frekventistisen tilastotieteen rinnalla tullaan käyttämään bayesiläistä ajattelutapaa, on Bayes-menetelmä esitelty myös luvussa 3.

Työn painopiste on intensiteetin estimoinnissa ja sen mallintamisessa: Luvussa 4 käsitellään intensiteetin estimointia kenttähavainnoinnin yhteydessä ja luvussa 5 karttamuotoiselle aineistolle. Luku 4 alkaa kenttähavainnointiin liittyvien käsitteiden ja menetelmien esittelyllä. Intensiteetin estimointia on tarkasteltu ensin harvan otannan tapauksessa, jossa esitellään perinteiset menetelmät, lukumäärä- ja etäisyysmenetelmä. Riippuvien havaintojen tapauksessa intensiteetin estimointia on tarkasteltu satunnaiskenttämenetelmällä ja lopuksi hierarkkisella mallintamisella, joka mahdollistaa myös aineistossa esiintyvän heterogeenisuuden kuvaamisen. Myös bayesiläistä hierarkkista mallinnusta on tarkasteltu.

Karttamuotoisen aineiston intensiteetin estimoinnissa luvussa 5 on käytetty ensin klassisen tilastotieteen periaatteita noudattavaa kernel-menetelmää, mutta pääpaino on kuitenkin bayesiläisessä mallinnuksessa. Mallin rakentamisessa on lähdetty liikkelle yksinkertaisimmasta mallista, joka perustuu riippumattomuusoletukseen. Tämän jälkeen sitä on parannettu ottamalla huomioon ilmiölle tyypillinen kontekstuaalisuus.

Luvussa 2 esitettyihin metsiin liittyviä empiirisiä tarkasteluja on esitelty luvussa 6. Metsien spatiaalista rakennetta on analysoitu erilaisten tilajärjestystä karakterisoivien tunnuksien avulla. Metsien satunnaisuutta on myös testattu. Lisäksi on tarkasteltu esiintyykö puutiheydessä mahdollisesti trendiä. Intensiteetin estimoinnissa on käytetty luvussa 5 esiteltyjä menetelmiä.

2 Tutkimusongelman ja -aineiston esittely

2.1 Ongelmanasettelu

Tutkimuksen sisällöllinen ongelma on metsäalueen puiden tiheyden vaihtelun kuvaaminen intensiteetin avulla. Intensiteetin estimaattia voidaan käyttää eksploratiivisessa analyysissä vertailukohtana haettaessa visuaalisesti assosiaatiota spatiaalisiin ympäristömuuttujiin, joilla tarkoitetaan paikkaan liittyvää mitattavissa olevaa ympäristön ominaisuutta kuten maaperän laatua tai korkeutta merenpinnasta. Mikäli ympäristömuuttujia on käytettävissä esim. paikkatietojärjestelmässä, voidaan niitä soveltaa intensiteetin vaihtelun selittämiseen. On kuitenkin realistista olettaa, että selittämätöntä vaihtelua esiintyy ja se pyritään mallintamaan stokastisen prosessin avulla. Tutkimus voidaan luokitella paikkatietojärjestelmäperusteiseksi (GIS) tilastoanalyysimenetelmien kehitystyöksi.

Empiirisenä aineistona on kaksi dosentti Timo Kuuluvaisen (Helsingin yliopisto, metsäekologian laitos) tutkimuksessaan (Kuuluvainen ym., 1996) käyttämää aineistoa, joista toinen kuvaa luonnonmetsää ja toinen talousmetsää. Näitä kahta metsäaluetta tullaan vertailemaan keskenään tiheyden vaihtelun ja tilajärjestyksen avulla. Kiinnostus on erityisesti estimoitujen intensiteettien vertailussa eli kuinka spatiaalisesti homogeenisiä tai heterogeenisiä metsät ovat. Kiinnostus voisi olla edelleen puiden kokojakauman ja spatiaalisen heterogeenisuuden välisessä yhteydessä, mutta tutkimus on rajattu koskemaan vain tilajärjестystä. Puumuuttujien mukaanotto on kiinnostava jatkotutkimuksen aihe.

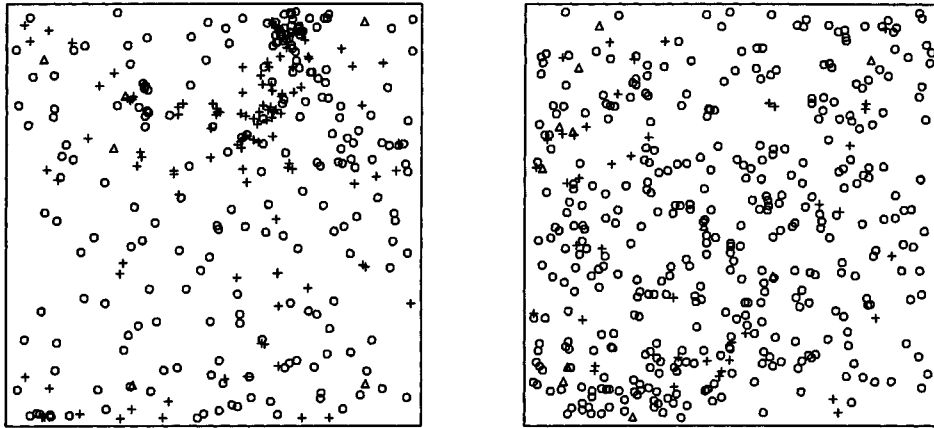
2.2 Tutkimusaineisto

Käsiteltävä aineisto koostuu kahdesta erillisestä puiden sijaintikartasta. Toinen niistä on kerätty Vuorijärven ja toinen Susimäen metsästä. Molemmat sijaitsevat Etelä-Suomessa lähellä Helsingin yliopiston metsäntutkimusasemaa Hyytiälää. Vuorijärvi on hoidettua talousmetsää, jonka käsittelyn päämääränä on ollut maksimaalinen puun nettotuotanto. Metsä on noin 105 vuotta vanha ja viimeisin hakkuu on suoritettu noin 35 vuotta sitten. Tarkkoja tietoja metsän hoitohistoriasta ei ole. Susimäki on puolestaan luonnon itsensä hoitamaa metsää, jossa harvennushakkuuta ei ole tehty. Metsä on noin 150 vuotta vanha ja se rauhoitettiin harvinaisen kasvillisuuden vuoksi vuonna 1955.

Molemmista metsistä on ollut tutkittavana 0.25 hehtaarin alue (50m × 50m). Kaikki vähintään 1.3 metriä korkeat puut on rekisteröity. Aineisto koostuu puiden sijaintikoordinaateista, pituuksista ja halkaisijoista sekä puulajista. Lisäksi on todettu, onko puu kuollut vai elävä. Puut on eroteltu ylemmän ja alemman kerrokseen sen mukaan, ovatko ne yli vai alle 15 metriä korkeita. Molemmissa metsissä valtapuulajina on kuusi, mutta joukossa on myös mäntyjä ja koivuja. Lisäksi Vuorijärven metsän alemmasta kerroksesta löytyi pihlajia ja Susimäen ylemmästä kerroksesta yksi haapa. Taulukossa 2.1 on esitetty puiden lukumäärät ja suhteelliset osuudet sekä metsien runkotilavuudet. Kummastakaan tutkitusta alueesta ei löytynyt hiljattain kaatuneita puita eikä muita luonnon häiriöitä.

Kuvassa 2.1 on esitetty tutkittujen metsäalueiden puiden sijainnit. Hoide-tussa Vuorijärven metsässä puut ovat sijoittuneet melko tasaisesti ja väljästi lukuunottamatta oikeassa yläkulmassa olevaa tiheämpää aluetta. Kyseisellä paikalla on ollut jokin aika sitten hakkuuaukko, minkä vuoksi siinä nyt kasvaa paljon nuoria koivuja. Susimäen luonnonmetsässä puut näyttävät sijaitsevan melko satunnaisesti ja tiheämmässä kuin Vuorijärven metsässä.

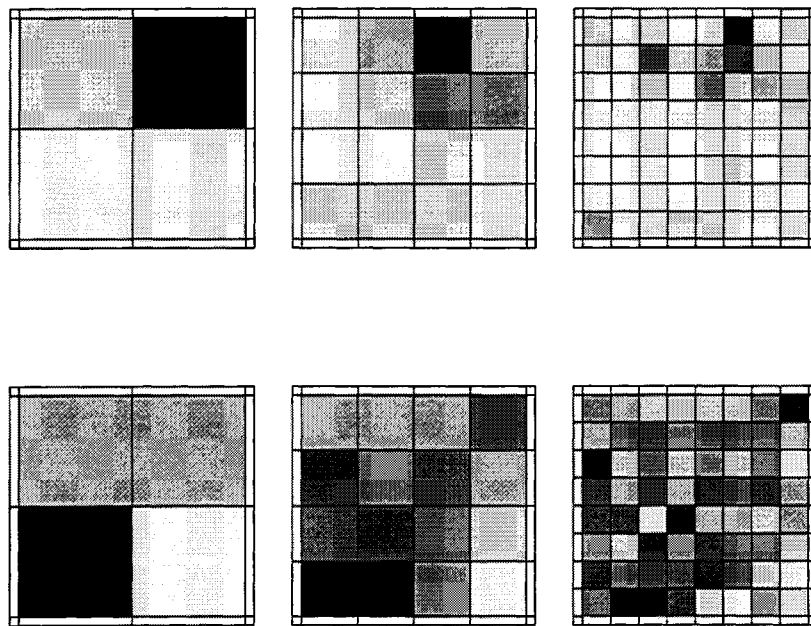
Puiden tiheyden vaihtelua tutkittavalla alueella voidaan alustavasti tarkastella aggregoitujen empiiristen intensiteettien avulla. Ne saadaan, kun tutkittava alue diskretisoidaan $k \times k$ -ruudukoksi ja lasketaan kuhunkin ruutuun kuuluvien puiden lukumäärät. Nämä ns. intensiteetin raakaestimaatit voidaan esittää harmaasävykuvana, josta saa karkean yleiskuvan tiheyden vaihtelusta. Ruutujen määrään vaikuttaa se, miten tarkka kuva tiheyden vaihtelusta halutaan. Kasvattamalla k :ta saadaan tarkempi kuva siitä, missä tiheyden vaihtelua esiintyy, mutta samalla ruutukohtaisten estimaattien luotettavuus heikkenee. Kuvassa 2.2 on metsien empiiriset intensiteetit, kun $k = 2, 4, 8$. Tummin alue kuvaa suurinta tiheyttä.



Kuva 2.1 Vasemmalla Vuorijärven ja oikealla Susimäen metsien puiden sijainnit:
o = kuusi, + = koivu ja Δ = mänty.

Taulukko 2.1 Vuorijärven ja Susimäen metsistä mitatut ominaisuudet ja puulajit (Kuuluvainen ym., 1996).

Ominaisuus	Vuorijärvi	Susimäki
Puiden lukumäärä/ha		
yhteensä	1360	1756
elävät puut	1352	1644
korkeus \geq 15m	400	664
korkeus < 15 m	952	980
Puulajit (%)		
Korkeus \geq 15m		
kuusi	92.0	76.3
mänty	4.0	4.7
koivu	4.0	18.4
haapa	-	0.6
Korkeus < 15 m		
kuusi	52.7	94.8
mänty	0.8	1.1
koivu	32.7	4.1
pihlaja	13.8	-
Metsän runkotilavuus m ³ /ha		
yhteensä	335	427
elävät puut	329	422
kuolleet puut	6	5



Kuva 2.2 Empiiriset intensiteetit 2×2 , 4×4 ja 8×8 -diskretisoinneilla. Ylhäällä Vuorijärvi ja alhaalla Susimäki. Tumma sävy kuvaa suurta intensiteettiä ja vaalea pientä.

3 Spatiaalinen pisteprosessi

Spatiaalinen pisteprosessi on sellainen stokastinen mekanismi, jonka reaalisatitot ovat R^d :n pistejoukkoja. Tässä työssä tarkastellaan vain tason R^2 pisteprosesseja ja niiden soveltamista puiden sijaintien stokastisena mallina.

Tässä luvussa on esitelty pisteprosessien teoriaa ja eräitä yleisimpiä pisteprosesseja. Luvun lopussa on tarkasteltu pisteprosessien simulointia. Tässä esitellyn pisteprosessien teorian lähteenä on käytetty kirjaa Diggle (1983).

3.1 Käsitteitä

Pisteprosessi avaruudessa R^2 on satunnaismuuttuja, joka saa arvoja mitallisessa avaruudessa $[E, \mathcal{E}]$. E on niiden R^2 :n aliavaruuksien φ perhe, jotka toteuttavat seuraavat ehdot (Stoyan, Kendall & Mecke, 1995):

1. joukko φ on *lokaalisti äärellinen*, ts. rajoitetussa joukossa on äärellinen määrä pisteitä
2. joukko φ on *yksinkertainen*, ts. $x_i \neq x_j$, jos $i \neq j$.

Jos molemmat ehdot ovat voimassa, on $\varphi \in E$ R^2 :n suljettu alijoukko. Joukon E alkia φ voidaan pitää myös mittana R^2 :ssa siten, että $\varphi(B)$ on φ :n pisteiden lukumäärä R^2 :n rajoitetussa Borel-joukossa B . E :n σ -algebra \mathcal{E} on suppein σ -algebra, jolle kaikki kuvaukset $\varphi \rightarrow \varphi(B)$ ovat mitallisia. Pisteprosessi N määritellään mitallisena kuvauksena todennäköisyysavaruudelta

$[\Omega, A, P]$ avaruudelle $[E, \mathcal{E}]$. Pisteprosessin N jakauma P avaruudessa $[E, \mathcal{E}]$ määritellään todennäköisyyksien

$$P(Y) = P(N \in Y) = P(\{\omega \in \Omega : N(\omega) \in Y\}), \quad Y \in \mathcal{E},$$

avulla. Pisteprosessi N voidaan myös tulkita satunnaismitaksi.

Pisteprosessi N on *stationaarinen*, jos kaikki prosessia koskevat todennäköisyysväittämät ovat invariantteja siirron suhteen. Ts. kun pisteprosessi kirjoitetaan muodossa $N = \{x_1, x_2, \dots\} = \{x_n\}$, prosesseilla $N = \{x_n\}$ ja $N_x = \{x_n + x\}$ on sama jakauma kaikilla $x \in R^2$.

Pisteprosessi on *isotrooppinen*, jos rotaatio eli akseleiden kierto ei muuta todennäköisyyslakia, ts. prosesseilla N ja rN on sama jakauma kaikilla rotaatioilla r .

Jos prosessi on sekä stationaarinen että isotrooppinen, niin se on *liikeinvariantti*.

Pisteprosessia sanotaan *täysin satunnaiseksi* prosessiksi, jos prosessin pisteet sijaitsevat tasajakautuneesti toisistaan riippumatta annetussa rajoitetussa alueessa. Täysin satunnainen prosessi on identtinen homogeenisen Poisson-prosessin kanssa, jonka tarkempi määrittely on kappaleessa 3.3. Pisteprosessin satunnaisuuden testaamiseen on kehitetty useita testejä (Diggle, 1983).

Pisteprosessin *ensimmäisen kertaluvun ominaisuus* eli *intensiteetti* kuvaa pistetiheyttä. Intensiteetti ilmoittaa prosessin pisteiden keskimääräisen lukumäärän pinta-alayksikköä kohti ja se määritellään *intensiteettifunktiolla*

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\}, \quad (3.1)$$

jossa dx on infinitesimaalisen pieni alue koordinaattipisteen x ympärillä, $|dx|$ on sen pinta-ala ja $N(dx)$ pisteiden lukumäärä alueella dx . Stationaariselle prosessille $\lambda(x) \equiv \lambda$.

Pisteprosessin *toisen kertaluvun ominaisuudet* kuvaavat spatiaalista riippuvuutta pisteiden välillä. Spatiaalisen riippuvuuden vaihtelua kuvataan *toisen kertaluvun intensiteettifunktiolla*

$$\lambda_2(x, y) = \lim_{|dx|, |dy| \rightarrow 0} \left\{ \frac{E[N(dx)N(dy)]}{|dx||dy|} \right\}, \quad (3.2)$$

jossa x ja y ovat koordinaattipisteitä. Stationaariselle prosessille $\lambda_2(x, y) \equiv \lambda_2(x - y)$ ja edelleen stationaariselle ja isotrooppiselle prosessille $\lambda_2(x, y) \equiv \lambda_2(t)$, missä t on pisteiden x ja y välinen euklidinen etäisyys.

Pisteprosessin *kovarianssitiheys* stationaarille ja isotrooppiselle prosessille ensimmäisen ja toisen kertaluvun intensiteetin avulla määriteltynä on

$$\gamma(t) = \lambda_2(t) - \lambda^2. \quad (3.3)$$

Toinen tapa skaalata $\lambda_2(t)$ on jakaa se λ^2 :lla, jolloin tuloksena on fyysikkojen käyttämä *parikorrelaatiofunktio*

$$g(t) = \frac{\lambda_2(t)}{\lambda^2} \quad (3.4)$$

(Stoyan ym., 1995, s. 129).

Stationaarisen ja isotrooppisen prosessin toisen kertaluvun ominaisuuksia voidaan vaihtoehtoisesti kuvata *K-funktiolla*. Funktio määritellään

$$K(t) = \lambda^{-1} E[\text{satunnaisesti valitusta prosessin pisteestä enintään etäisyydellä } t \text{ olevien muiden pisteiden lukumäärä}].$$

Prosessin tunnus $\lambda K(t)$ kertoo siten t -säteisessä ympyrässä olevien muiden pisteiden lukumäärän odotusarvon, kun ympyrän keskipiste on satunnaisesti valitussa prosessin pisteessä.

Tarkastellaan K -funktion ja toisen kertaluvun intensiteettifunktion välistä yhteyttä. Koska prosessi oletettiin yksinkertaiseksi, ts.

$$P\{N(dx) > 1\} \rightarrow 0,$$

kun $|dx| \rightarrow 0$, niin tästä seuraa, että

$$E[N(dx)] \approx P\{N(dx) = 1\}$$

ja että

$$E[N(dx)N(dy)] \approx P\{N(dx) = N(dy) = 1\}.$$

Tapahtuman x ehdollinen intensiteetti ehdolla, että origossa on piste, on $\lambda_2(0, x)/\lambda$. Siten korkeintaan etäisyyden t päässä olevien pisteiden lukumäärän odotusarvo ehdolla, että origossa on piste, on

$$\lambda K(t) = \int_0^{2\pi} \int_0^t \{\lambda_2(x)/\lambda\} x dx d\theta = 2\pi\lambda^{-1} \int_0^t \lambda_2(x)x dx$$

tai yhtäpitävästi

$$\lambda_2(t) = \lambda^2(2\pi t)^{-1} K'(t).$$

Tässä ehdollistaminen on ajateltava ns. Palmiin todennäköisyyksien mielessä.

Pisteprosessin analysointiin on kehitetty useita lukumääriin ja etäisyys-tarkasteluihin perustuvia menetelmiä. Ne sopivat sekä pistekuvion satun-naisuuden testaamiseen että prosessin tiheyden estimointiin. Lukumääräme-netelmässä ei tarvitse tietää pisteiden tarkkaa sijaintia, sillä aineisto koos-tuu osajoukoissa lasketuista lukumääristä. Osajoukot voivat olla esimerkiksi vierekkäisiä neliöitä, ns. kvadraatteja. Lukumäärien mallintaminen perustuu pisteiden lukumääräjakaumaan

$$p_n(A) = P\{N(A) = n\}, \quad n = 0, 1, \dots,$$

jossa A on tutkittavan joukon osajoukko. Pisteiden lukumäärän odotusarvo A :ssa on yleisesti

$$E[N(A)] = \int_A \lambda(x) dx \quad (3.5)$$

ja stationaariselle prosessille

$$E[N(A)] = \lambda|A|. \quad (3.6)$$

Prosessin yksinkertaisuudesta seuraa, että

$$\begin{aligned} E[N(A)^2] &= E\left[\left\{\int_A N(dx)\right\}^2\right] \\ &= E\left[\int_A N(dx) + \int_A \int_A N(dx)N(dy)\right] \\ &= \int_A \lambda(x) dx + \int_A \int_A \lambda_2(x, y) dx dy, \end{aligned} \quad (3.7)$$

ja stationaariselle prosessille

$$E[N(A)^2] = \lambda|A| + \int_A \int_A \lambda_2(x - y) dx dy. \quad (3.8)$$

Täten

$$\begin{aligned} \text{Var}[N(A)] &= E[N(A)^2] - [EN(A)]^2 \\ &= \int_A \int_A \lambda_2(x - y) dx dy + \lambda|A|(1 - \lambda|A|). \end{aligned} \quad (3.9)$$

Vastaavasti

$$\text{Cov}[N(A), N(B)] = \int_A \int_B \lambda_2(x - y) dx dy + \lambda|A \cap B| - \lambda^2|A||B|. \quad (3.10)$$

Edellä olevasta nähdään, että stationaarisen prosessin pisteiden lukumää-rän odotusarvo riippuu vain intensiteetistä ja alueen pinta-alasta, mutta va-rianssi ja kovarianssi riippuvat myös toisen kertaluvun intensiteetistä.

3.2 Etäisyystarkasteluja

Olkoon A tason rajoitettu osajoukko, jossa on n prosessin pistettä. Piste-prosessi oletetaan stationaariseksi ja isotrooppiseksi. Tarkastellaan erilaisia pisteiden välisiä etäisyyksiä ja niiden jakaumia, kun prosessi oletetaan täysin satunnaiseksi prosessiksi.

3.2.1 Kahden pisteen välisen etäisyyden jakauma

Olkoon t kahden riippumattoman ja joukossa A tasaisesti jakautuneen pisteen välinen etäisyys. Sen teoreettinen jakauma riippuu alueen koosta ja muodosta ja se voidaan joissain erikoistapauksissa (A on neliö tai ympyrä) esittää suljetussa muodossa (Bartlett, 1964). Yksikköneliölle t :n kertymäfunktio on

$$H(t) = \begin{cases} \pi t^2 - 8t^3/3 + t^4/2, & 0 \leq t \leq 1 \\ 1/3 - 2t^2 - t^4/2 + 4(t^2 - 1)^{1/2}(2t^2 + 1)/3 \\ \quad + 2t^2 \sin^{-1}(2t^{-2} - 1), & 1 < t \leq \sqrt{2} \end{cases}$$

ja yksikköympyrälle

$$H(t) = 1 + \pi^{-1} \left\{ 2(t^2 - 1) \cos^{-1}(t/2) - t(1 + t^2/2) \sqrt{1 - t^2/4} \right\},$$

kun ($0 \leq t \leq 2$). Yksikköympyrässä lyhyille etäisyyksille t asymptoottinen kertymäfunktio on

$$H(t) = \frac{\pi t^2}{a} - \frac{2t^3 b}{3a^2} + o(t^3),$$

jossa a on A :n pinta-ala, b on A :n kehän pituus ja $\lim_{t \rightarrow 0} t^3 o(t^3) = 0$ (ks. esim. Matérn, 1960, s. 24). Ylläolevilla integraaligeometrisilla kaavoilla on käyttöä mm. otantatarkasteluihin liittyvissä varianssiarvioissa (ks. luku 4.2.1).

Kun A :ssa on n pistettä, on kahden pisteen välisiä etäisyyksiä kaikkiaan $\frac{1}{2}n(n-1)$. Merkitään t_{ij} :llä i :nnen ja j :nnen prosessin pisteen välistä etäisyyttä. Etäisyyden t empiirinen kertymäfunktio määritellään enintään etäisyydellä t olevien pisteparien x_i ja x_j suhteellisenä osuutena, ts.

$$\hat{H}_1(t) = \frac{\#(t_{ij} \leq t)}{\frac{1}{2}n(n-1)}, \quad (3.11)$$

jossa $\#$ tarkoittaa lukumäärää.

3.2.2 Lähinaapurietäisyys

Tarkastellaan mielivaltaisen prosessin pisteen ja sen lähimmän naapurin välistä etäisyyttä. Olkoon y_i etäisyys i :nnestä prosessin pisteestä lähimpään toiseen prosessin pisteeseen. Etäisyyksiä y_i kutsutaan *lähinaapurietäisyyksiksi*. Merkitään etäisyyden y kertymäfunktioita G :llä ja määritellään

$$G(y) = P\{\text{etäisyys mielivaltaisesta prosessin pisteestä lähimpään prosessin pisteeseen on enintään } y\}.$$

Lähinaapurietäisyys ei ole symmetrinen, mutta pienillä etäisyyksillä esiintyysuurella todennäköisyydellä kaksinkertaisia mittauksia keskinäisten lähimpien naapureiden osalta.

Etäisyyksien $\{y_i\}$ empiirinen kertymäfunktio saadaan niiden havaittujen lähinaapurietäisyyksien y_i , jotka ovat enintään y , ja kaikkien lähinaapurietäisyyksien lukumäärien suhteena

$$\hat{G}_1(y) = \frac{\#(y_i \leq y)}{n}. \quad (3.12)$$

Lähinaapurietäisyyksien y teoreettinen jakauma riippuu prosessista, havaintojen lukumäärästä n ja alueesta A eikä sitä monimutkaisten reunavaikutusten vuoksi voi esittää suljetussa muodossa. Silloin kun pisteet ovat sijoittuneet tasajakauman mukaan toisistaan riippumatta, on approksimatiivinen kertymäfunktio y :lle, kun reunavaikutuksia ei oteta huomioon,

$$G(y) = 1 - (1 - \pi y^2 |A|^{-1})^{n-1},$$

jossa $\pi y^2 |A|^{-1}$ on todennäköisyys sille, että mielivaltainen prosessin piste on enintään etäisyydellä y valitusta pisteestä. Suurille n voidaan käyttää approksimaatiota

$$G(y) = 1 - \exp(-\lambda \pi y^2), \quad y \geq 0,$$

jossa $\lambda = n |A|^{-1}$.

3.2.3 Etäisyys otantapistestä lähimpään prosessin pisteeseen

Kiinnostus voi olla myös tason mielivaltaisen pisteen (otantapisteen) ja lähimmän prosessin pisteen välisessä etäisyydessä. Olkoon m otantapisteen lukumäärä ja d_i etäisyys i :nnestä otospisteestä lähimpään prosessin pisteeseen. Merkitään etäisyyden d kertymäfunktioita F :llä ja määritellään

$$F(d) = P\{\text{etäisyys mielivaltaisesta otantapistestä lähimpään prosessin pisteeseen on enintään } d\}.$$

F -funktioita voidaan pitää *tyhjän tilan tunnuslukuna*, sillä

$$1 - F(d) = P\{\text{origokeskinen } d\text{-säteinen ympyrä on tyhjä}\}.$$

Empiirinen kertymäfunktio, joka määritellään kaavalla

$$\hat{F}_1(d) = \frac{\#(d_i \leq d)}{m}, \quad (3.13)$$

mittaa myös tyhjää tilaa A :ssa siinä mielessä, että $1 - \hat{F}_1(d)$ estimoi pinta-alaa $|B_d|$, missä B_d koostuu kaikista niistä A :n pisteistä, jotka ovat vähintään etäisyydellä d jokaisesta A :han kuuluvasta prosessin pisteestä.

Koska täysin satunnaisessa prosessissa pisteet ovat jakautuneet tasajauman mukaisesti toisistaan riippumatta, voidaan prosessin piste ja otantapiste samaistaa. Täten on samantekevää tarkastellaanko lähinaapurietäisyyttä vai tyhjää tilaa otantapisteen ympärillä, jolloin täysin satunnaiselle prosessille tyhjän tilan tunnusluku ja lähinaapurietäisyys ovat identtiset. Siten myös etäisyyden d teoreettiselle jakaumalle voidaan suurilla n käyttää approksimaatiota

$$F(d) = 1 - \exp(-\lambda\pi d^2), \quad d \geq 0.$$

3.3 Homogeeninen Poisson-prosessi

Homogeeninen Poisson-prosessi on yksinkertaisin spatiaalisia pistekuvioita generoiva stationaarinen ja isotrooppinen pisteprosessi. Prosessille on karakterista riippumattomuusominaisuus, minkä vuoksi sitä käytetään nollahypoteesina pisteprosessin satunnaisuuden testaamisessa. Poisson-prosessia käytetään myös usein lähtökohtana muita prosesseja konstruoidessa ja simuloitaessa.

Olkkoon $\lambda > 0$ ja $N(A) = n$ pisteiden lukumäärä rajoitetussa tason joukossa A . Homogeeninen Poisson-prosessi on seuraavat ehdot täyttävä pisteprosessi:

1. $N(A)$ noudattaa Poisson-jakaumaa parametrilla $\lambda|A|$.
2. Ehdolla, että $N(A) = n$, $\{x_1, \dots, x_n\}$ on riippumaton otos tasajaumasta A :ssa.
3. Erillisille joukoille A ja B pisteiden lukumäärät $N(A)$ ja $N(B)$ ovat stokastisesti riippumattomat.

Poisson-prosessin parametri λ on prosessin intensiteetti. Riippumattomuudesta (ehto 3) seuraa, että toisen kertaluvun intensiteetti on

$$\lambda_2(t) = \lambda^2, \quad t > 0.$$

Odotusarvo ja varianssi prosessin pisteiden lukumäärälle A :ssa saadaan suoraan jakaumaoletuksesta:

$$E[N(A)] = \text{Var}[N(A)] = \lambda|A|.$$

Homogeenisen Poisson-prosessin K -funktio on

$$K(t) = \frac{E[N(b(0,t))]}{\lambda} = \frac{\lambda|b(0,t)|}{\lambda} = \pi t^2, \quad t > 0,$$

jossa $b(0,t)$ on origokeskinen t -säteinen ympyrä. Tyhjän tilan tunnusluku on

$$\begin{aligned} F(x) &= P\{N(b(0,x)) > 0\} = 1 - P\{N(b(0,x)) = 0\} \\ &= 1 - \frac{(\lambda\pi x^2)^0}{0!} \exp(-\lambda\pi x^2) = 1 - \exp(-\lambda\pi x^2) \end{aligned}$$

Koska homogeeninen Poisson-prosessi vastaa täydellisen satunnaisuuden tilannetta, lähinaapurietäisyys ja tyhjän tilan tunnusluku ovat identtiset, kuten kappaleessa 3.2.3 todettiin. Täten

$$G(x) = F(x) = 1 - \exp(-\lambda\pi x^2) \quad x > 0.$$

Homogeeninen Poisson-prosessi voidaan määritellä myös raja-arvon avulla (Hogg & Craig, 1978). Tämä lähestymistapa antaa Poisson-prosessille toisen karakterisoinnin ”harvinaisten” tapahtumien prosessina:

$$P(N(\Delta A) = 1) \approx \lambda|\Delta A| \quad \text{ja} \quad P(N(\Delta A) > 1) \approx 0,$$

kun ΔA on pieni. Tehdään seuraavassa tarkastelu 1-ulotteisessa tapauksessa. Tarkastelu voidaan yleistää R^d :hen käsittelemällä joukkofunktioiden derivaattoja.

Olkoon $|\Delta x| > 0$ ja

$$o(|\Delta x|) = \lim_{|\Delta x| \rightarrow 0} \frac{o(|\Delta x|)}{|\Delta x|} = 0.$$

Pisteprosessi on homogeeninen Poisson-prosessi, jos ja vain jos seuraavat ehdot ovat voimassa:

$$1.' \quad P(N(\Delta x) = 1) = \lambda|\Delta x| + o(|\Delta x|),$$

2.' $P(N(\Delta x) \geq 2) = o(|\Delta x|)$ ja

3.' tapahtumien lukumäärät ovat riippumattomia erillisillä väleillä.

Ehdosta 2' seuraa, että Poisson-prosessi on yksinkertainen prosessi.

Todistetaan ekvivalenssi molempiin suuntiin. Oletetaan ensin, että prosessi on homogeeninen Poisson-prosessi. Silloin

1.'

$$\begin{aligned} P(N(\Delta x) = 1) &= \lambda|\Delta x|e^{-\lambda|\Delta x|} \\ &= \lambda|\Delta x| \left(1 - \lambda|\Delta x| + \frac{\lambda^2|\Delta x|^2}{2!} - \frac{\lambda^3|\Delta x|^3}{3!} + \dots \right) \\ &= \lambda|\Delta x| + o(|\Delta x|), \end{aligned}$$

2'

$$\begin{aligned} P(N(\Delta x) \geq 2) &= 1 - P(N(\Delta x) < 2) \\ &= 1 - e^{-\lambda|\Delta x|} - \lambda|\Delta x|e^{-\lambda|\Delta x|} \\ &= o(|\Delta x|) \end{aligned}$$

3.' ehto on identtinen määritelmän ehdon 3 kanssa.

Ts. jos prosessi on homogeeninen Poisson-prosessi, ovat edellä esitetyt kolme ehtoa voimassa.

Todistus toiseen suuntaan voidaan tehdä induktioperiaatetta hyväksi käyttäen. Oletetaan siis, että edellä esitetyt kolme ehtoa ovat voimassa. Todetaan ensin, että väite pitää paikkansa, kun tapahtumien lukumäärä $k = 0$. Ehdon 3' avulla saadaan

$$\begin{aligned} P(N(x + \Delta x) = 0) &= P(N(x) = 0)P(N(\Delta x) = 0) \\ &= P(N(x) = 0)[1 - \lambda|\Delta x| - o(|\Delta x|)] \\ &= P(N(x) = 0) - \lambda|\Delta x|P(N(x) = 0) + o(|\Delta x|), \end{aligned}$$

josta edelleen

$$\frac{P(N(x + \Delta x) = 0) - P(N(x) = 0)}{|\Delta x|} = -\lambda P(N(x) = 0) + \frac{o(|\Delta x|)}{|\Delta x|}.$$

Jos nyt $|\Delta x| \rightarrow 0$, saadaan differentiaaliyhtälö

$$D_x P(N(x) = 0) = -\lambda P(N(x) = 0),$$

jonka ratkaisuksi reunaehdon $P(N(0) = 0) = 1$ kanssa saadaan

$$P(N(x) = 0) = e^{-\lambda x}.$$

Täten tapahtumien lukumäärä x :n pituisella välillä on Poisson-jakautunut parametrilla λx .

Oletetaan nyt, että väite pitää paikkansa arvolla $k - 1$ ja osoitetaan, että se pätee myös arvolla k . Nyt saadaan vastaavasti

$$\begin{aligned} P(N(x + \Delta x) = k) &= \\ &= P(N(x) = k)P(N(\Delta x) = 0) + P(N(x) = k - 1)P(N(\Delta x) = 1) + \dots \\ &= P(N(x) = k)[1 - \lambda|\Delta x| - o(|\Delta x|)] \\ &\quad + P(N(x) = k - 1)[\lambda|\Delta x| + o(|\Delta x|)] + o(|\Delta x|), \end{aligned}$$

josta

$$\begin{aligned} \frac{P(N(x + \Delta x) = k) - P(N(x) = k)}{|\Delta x|} &= \\ &= -\lambda P(N(x) = k) + \lambda P(N(x) = k - 1) + \frac{o(|\Delta x|)}{|\Delta x|} \end{aligned}$$

ja rajallemenon ($|\Delta x| \rightarrow 0$) jälkeen

$$D_x P(N(x) = k) = -\lambda P(N(x) = k) + \lambda P(N(x) = k - 1).$$

Nyt käytetään hyväksi induktio-oletusta, jolloin saadaan

$$D_x P(N(x) = k) + \lambda P(N(x) = k) = \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}.$$

Tämän differentiaaliyhtälön yleinen ratkaisu on

$$\begin{aligned} P(N(x) = k) &= e^{-\lambda x} \int_0^x e^{-\lambda t} \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} dt + ce^{-\lambda x} \\ &= \frac{\lambda^k}{(k-1)!} e^{-\lambda x} \int_0^x t^{k-1} dt + ce^{-\lambda x} \\ &= \frac{(\lambda x)^k}{k!} e^{-\lambda x} + ce^{-\lambda x} \end{aligned}$$

ja reunaehdon $P(N(0) = k) = 0$ kanssa

$$P(N(x) = k) = \frac{(\lambda x)^k}{k!} e^{-\lambda x}.$$

Täten arvolla k tapahtumien lukumäärä x :n pituisella välillä on Poisson-jakautunut parametrilla λx . Induktioperiaatteesta seuraa, että tämä on voimassa kaikilla arvoilla $k = 1, 2, 3, \dots$. Ts. jos edellä esitetyt kolme ehtoa ovat voimassa, on prosessi homogeeninen Poisson-prosessi.

3.4 Epähomogeeninen Poisson-prosessi

Homogeenisesta Poisson-prosessista saadaan epästationaarinen prosessi, kun intensiteetti λ korvataan intensiteettifunktiolla $\lambda(x)$. Pisteprosessi on epähomogeeninen Poisson-prosessi, jos seuraavat ehdot ovat voimassa:

1. $N(A)$ noudattaa Poisson jakaumaa parametrilla $\int_A \lambda(x) dx$.
2. Ehdolla $N(A) = n$ A :ssa olevat pisteet on riippumaton otos jakaumasta, joka on suoraan verrannollinen $\lambda(x)$:aan.
3. Erillisille joukoille A ja B pisteiden lukumäärät $N(A)$ ja $N(B)$ ovat stokastisesti riippumattomat.

Epähomogeeninen Poisson-prosessi tekee mahdolliseksi kovariaattien mukaan ottamisen analyysiin intensiteettifunktion

$$\lambda(x) = h\{z_1(x), z_2(x), \dots, z_p(x)\}$$

kautta, jossa $z_i(x)$ sisältää tietoa esimerkiksi maaperän laadusta yms. ja h on linkkifunktio. Erikoistapauksena on tiheydessä esiintyvä trendi, jolloin kovariaattina on koordinaatit $x = (u, v)$. Tyypillinen esimerkki on polynomipinta

$$\lambda(x) = h\left\{\sum_{r+s \leq p} a_{rs} u^r v^s\right\},$$

jossa p on trendipinnan aste. Trendipinta voi olla vakio, lineaarinen, toisen tai kolmannen asteen pinta jne.

Yksinkertaisimmillaan epähomogeeninen Poisson-prosessi voidaan ajatella seuraavasti: Diskretisoidaan alue A siten, että $A = \cup A_i$ ja $A_i \cap A_j = \phi$. Jokaisessa joukossa A_i intensiteetti on vakio λ_i , jolloin pisteiden lukumäärä joukossa A_i on $\lambda_i = \int_{A_i} \lambda(x) dx$. Muodostetaan joukkoihin A_i toisistaan riippumatta homogeeninen Poisson-prosessi intensiteettinä λ_i . Täten osittamalla prosessi ehdollisiin homogeenisiin Poisson-prosesseihin saadaan lopputuloksena epähomogeeninen Poisson-prosessi. Tällaisella "rasteriprosessilla" voidaan epähomogeenista Poisson-prosessia approksimoida mielivaltaisen tarkasti.

3.5 Cox-prosessi

Cox-prosessi on *kahdesti stokastinen prosessi*. Se muodostuu satunnaisesta intensiteettifunktiosta $\Lambda(x)$ ja epähomogeenisesta Poisson-prosessista seuraavasti:

1. $\{\Lambda(x) : x \in R^2\}$ on ei-negatiivisia arvoja saava satunnaiskenttä (stokastinen prosessi R^2 :ssa).
2. Ehdolla $\{\Lambda(x) = \lambda(x) : x \in R^2\}$ tapahtumat muodostavat epähomogeenisen Poisson-prosessin intensiteettifunktiona $\lambda(x)$.

Cox-prosessi on stationaarinen, jos ja vain jos intensiteettiprosessi on stationaarinen. Sama pätee isotrooppisuudelle. Samoin Cox-prosessin korrelaatio-ominaisuudet periytyvät satunnaiskentän vastaavista ominaisuuksista.

Tarkastellaan stationaarisen Cox-prosessin ominaisuuksia. Olkoon satunnaiskentän intensiteetin odotusarvo $E[\Lambda(x)] = \mu$ ja kovarianssifunktio $c(x, y) = E[\Lambda(x)\Lambda(y)] - \mu^2$. Silloin

$$\begin{aligned} E[N(A)] &= E_{\Lambda} E[N(A)|\Lambda] = E_{\Lambda} \int_A \Lambda(x) dx = \int_A E_{\Lambda} \Lambda(x) dx \\ &= \int_A \mu dx = \mu|A|. \end{aligned}$$

Koska stationaariselle prosessille $E[N(A)] = \lambda|A|$, niin voidaan päätellä, että Cox-prosessin intensiteetti on

$$\lambda = E[\Lambda(x)] = \mu.$$

Edelleen

$$\begin{aligned} E[N(A)N(B)] &= E_{\Lambda} E[N(A)N(B) | \Lambda] \\ &= E_{\Lambda} \left[\int_{A \cap B} \Lambda(x) dx + \int_A \int_B \Lambda(x)\Lambda(y) dx dy \right] \\ &= \int_{A \cap B} E_{\Lambda} \Lambda(x) dx + \int_A \int_B E_{\Lambda} \Lambda(x)\Lambda(y) dx dy \\ &= \mu|A \cap B| + \int_A \int_B E_{\Lambda} \Lambda(x)\Lambda(y) dx dy, \end{aligned}$$

jolloin kaavan 3.8 mukaan toisen kertaluvun intensiteetti on

$$\lambda_2(x, y) = E[\Lambda(x)\Lambda(y)] = c(x, y) + \lambda^2.$$

Stationaariselle ja isotrooppiselle Cox-prosessille toisen kertaluvun intensiteetti saadaan myös intensiteetin ja kovarianssin avulla, ts.

$$\lambda_2(t) = \lambda^2 + \gamma(t),$$

jossa

$$\gamma(t) = \text{Cov}[\Lambda(x), \Lambda(y)]$$

on intensiteettifunktioiden välinen autokorrelaatio. Huomattakoon, että intensiteettiprosessin kovarianssifunktio γ on myös pisteprosessin kovarianssitiheys (ks. kaava 3.3).

Seuraavassa tarkastellaan esimerkkejä Cox-prosessien konstruktioista. Nämä eroavat toisistaan ainoastaan erilaisen satunnaiskenttäosan vuoksi, sillä intensiteettiprosessilla ehdollistettu Cox-prosessi on aina epähomogeeninen Poisson-prosessi.

3.5.1 Matérnin klusteriprosessi

Matérnin klusteriprosessissa (Matérn, 1960) satunnaiskenttä $\Lambda(x)$ muodostetaan generoimalla ensin alueelle homogeeninen Poisson-prosessi N_0 , jonka intensiteetti on ρ . Tämän jälkeen jokaiseen generoituun pisteeseen $\{\xi_n\}$ liitetään R -säteinen ympyrä ja jokaiseen ympyrään intensiteetti η . Satunnaiskentäksi saadaan siis

$$\Lambda(x) = \frac{\eta}{\pi R^2} \sum_{\xi_n \in N_0} 1_{b(\xi_n, R)}(x), \quad (3.14)$$

jossa $b(\xi_n, R)$ on ξ_n -keskinen R -säteinen ympyrä. Nyt alueella, jota peittää kaksi ympyrää, on tiheys kaksi kertaa niin suuri kuin alueella, jota peittää vain yksi ympyrä. Sen sijaan alueella, jota ei peitä yksikään ympyrä, tiheys on nolla.

Matérnin klusteriprosessin intensiteetti saadaan Campbellin lauseen (Stoyan ym., 1995, s. 104) avulla seuraavasti:

$$\begin{aligned} \lambda &= E[\Lambda(x)] = \frac{\eta}{\pi R^2} E \left[\sum_{\xi_n \in N_0} 1_{b(\xi_n, R)}(x) \right] \\ &= \frac{\eta \rho}{\pi R^2} \int 1_{b(\xi, R)}(x) d\xi = \frac{\eta \rho}{\pi R^2} \pi R^2 = \eta \rho. \end{aligned}$$

Toisen kertaluvun intensiteetti klusteriprosessille yleisesti on (Stoyan ym., 1995, s. 160)

$$\lambda_2(r) = \lambda^2 + \frac{\lambda}{\bar{c}} Ec(c-1) \frac{f(r)}{2\pi r},$$

jossa c on klusterin koko. Nyt $c \sim \text{Pois}(\eta\pi R^2)$, jolloin $\bar{c} = \eta\pi R^2$, toinen kumulantti $Ec(c-1) = \eta^2 \pi^2 R^4$ ja

$$f(r) = \begin{cases} \frac{4r}{\pi R^2} \left(\arccos\left(\frac{r}{2R}\right) - \frac{r}{2R} \sqrt{1 - \frac{r^2}{4R^2}} \right), & \text{kun } 0 < r < 2R \\ 0 & \text{muualla.} \end{cases}$$

Täten Matérnin klusteriprosessin toisen kertaluvun intensiteetiksi saadaan

$$\lambda_2(r) = \begin{cases} \lambda^2 + \frac{2\lambda\eta}{\pi} \left(\arccos\left(\frac{r}{2R}\right) - \frac{r}{2R}\sqrt{1 - \frac{r^2}{4R^2}} \right), & \text{kun } 0 < r < 2R \\ \lambda^2, & \text{kun } r \geq 2R. \end{cases}$$

3.5.2 Log-gaussinen Cox-prosessi

Log-gaussisessa Cox-prosessissa (Møller ym., 1998) intensiteettiprosessi $\{\Lambda(s) : s \in \mathbb{R}^2\}$ on log-gaussinen satunnaiskenttä

$$\Lambda(s) = e^{Y(s)},$$

jossa $Y = \{Y(s) : s \in \mathbb{R}^2\}$ on gaussinen satunnaiskenttä keskiarvona $\mu = E[Y(s)]$ ja kovarianssifunktiona

$$k(s, t) = \text{Cov}[Y(s), Y(t)] = E[Y(s) - \mu][Y(t) - \mu].$$

Tällöin $Y(s) \sim N(\mu, \sigma^2)$, jossa $\sigma^2 = k(s, s)$. Prosessin intensiteetti on

$$\lambda = E[\Lambda(s)] = E[e^{Y(s)}] = e^{\mu + \frac{1}{2}\sigma^2}.$$

Toisen kertaluvun intensiteetti on

$$\lambda_2 = E[\Lambda(s)\Lambda(t)] = E[e^{Y(s)+Y(t)}] = e^{2\mu + \frac{1}{2}(2\sigma^2 + 2k(s,t))} = e^{2\mu + \sigma^2 + k(s,t)},$$

sillä $Y(s) + Y(t) \sim N(2\mu, 2\sigma^2 + 2k(s, t))$. Prosessin kovarianssitiheys on

$$\begin{aligned} c(s, t) &= \text{Cov}[\Lambda(s)\Lambda(t)] = E[\Lambda(s)\Lambda(t)] - E[\Lambda(s)]E[\Lambda(t)] \\ &= e^{2\mu + \sigma^2 + k(s,t)} - e^{2\mu + \sigma^2} = e^{2\mu + \sigma^2} (e^{k(s,t)} - 1). \end{aligned}$$

3.5.3 Kernel-tasoitettu gamma-Poisson-prosessi

Olkoon $\{A_i\}$ joukon A ositus siten, että $A = \cup_{i \in I} A_i$ ja $A_i \cap A_j = \emptyset$ kaikille $i \neq j$, jossa $I = \{1, \dots, p\}$. Esimerkiksi, jos suorakaiteen muotoista joukkoa A käsitellään rasterina, ositteet $\{A_i\}_{i \in I}$ ovat keskenään saman kokoisia vierekkäisiä suorakaiteita. Monissa sovelluksissa ositteet ovat kuitenkin vaihtelevankokoisia ja -muotoisia.

Yksinkertaisimmillaan gamma-Poisson-prosessi (Ickstadt & Wolpert, 1996), jota kutsutaan myös rasteriprosessiksi, saadaan muodostetuksi kahdessa vaiheessa. Ensimmäisessä vaiheessa generoidaan kuhunkin ositteeseen A_i intensiteetti λ_i toisistaan riippumattomasti Gamma(α, β^{-1})-jakaumasta.

Toisessa vaiheessa generoidaan kuhunkin ositteeseen A_i lukumäärät n_i toisistaan riippumatta Poisson-jakaumasta parametria $\lambda_i|A_i|$ käyttäen.

Näin muodostettu prosessi hallitaan kahdella parametrilla, sillä gamma-jakauman parametrit α ja β määräävät prosessin ominaisuudet. Prosessin haittapuolena on kuitenkin riippumattomuudesta aiheutuva kontekstuaalisuuden puute. Tällöin vierekkäisissä osajoukoissa A_i ja A_j pisteiden lukumäärät n_i ja n_j voivat vaihdella suuresti.

Eräs tapa saada kohinaisesta satunnaiskentästä korreloitunut on kernel-tasoittaa se, jonka jälkeen ositteiden intensiteetit ovat edelleen gammajakautuneita, mutta nyt riippuvia. Gammajakautuneita on ns. äärellisesti jakautunut eli jokaiselle positiiviselle kokonaisluvulle m gammajakautuneita voidaan esittää $m:n$ samoin jakautuneen riippumattoman satunnaismuuttujan summan jakaumana (Feller, 1971, s. 176). Tämän vuoksi kernel-tasointu on luonnollinen operaatio gammajakautuneelle muuttujalle. Diskretisoidulle aineistolle (rasterille) kernel-tasointu vastaa liukuvan keskiarvon tasointu. Tasointuun jälkeen intensiteetti muuttuu satunnaiskentässä tasaisesti eikä epätodennäköisiä suuria hyppäyksiä esiinny.

Kuten tasoittamattomassakin gamma-Poisson-prosessissa kuhunkin ositteeseen liitetään ensin toisistaan riippumattomasti gammajakautunut impulssi $\{\Gamma_j\}_{j \in J}$, $J = \{1, \dots, r\}$:

$$\Gamma_j \sim \text{Gamma}(\alpha_j, \beta_j^{-1}),$$

jolloin $E\Gamma_j = \alpha_j\beta_j$ ja $\text{Var}(\Gamma_j) = \alpha_j\beta_j^2$.

Olkoon $I = J$ ja käsitellään joukkoa A rasterina. Satunnaiskenttä saadaan muodostetuksi laskemalla jokaiseen ositteeseen viereisten ositteiden kanssa impulssien painotettu keskiarvo

$$\Lambda_i = \sum_{j \in J} k_{ij} \Gamma_j, \quad (3.15)$$

missä

$$k_{ij} = \begin{cases} k, & \text{kun } \|i - j\| = 0 \\ k', & \text{kun } 1 \leq \|i - j\| \leq \sqrt{2} \\ 0, & \text{muualla.} \end{cases}$$

Ei-negatiiviset painokertoimet k_{ij} muodostavat $p \times p$ -kerroinmatriisin K . Huomattakoon, että 3.15 johtaa riippumattomiin intensiteetteihin silloin, kun K on yksikkömatriisi.

Kernel-tasoitetussa gamma-Poisson-prosessissa havaintojen lukumäärät $\{n_i\}_{i \in I}$ ovat ehdollisesti riippumattomia ehdolla, että impulssit Γ_j ovat tunnetut, ja niiden ehdollinen jakauma on

$$\{n_i\}_{i \in I} | \{\Gamma_j\}_{j \in J} \sim \text{Poisson}(\Lambda_i | A_i),$$

missä $\Lambda_i = \sum_{j \in J} k_{ij} \Gamma_j$. Odotusarvo pisteiden lukumäärälle ositteessa i on

$$\begin{aligned} En_i &= E_{\Gamma} E[n_i | \Gamma] = E_{\Gamma} \left[\int_{A_i} \Lambda_i(x) dx | \Gamma \right] = E_{\Gamma} [\Lambda_i | A_i | \Gamma] \\ &= E \sum_{j \in J} |A_i| k_{ij} \Gamma_j = \sum_{j \in J} |A_i| k_{ij} E \Gamma_j = |A_i| \sum_{j \in J} k_{ij} \alpha_j \beta_j. \end{aligned}$$

Jos $i \neq i'$, niin

$$\begin{aligned} E[n_i n_{i'}] &= E_{\Gamma} E[n_i n_{i'} | \Gamma] = E_{\Gamma} [E(n_i | \Gamma) E(n_{i'} | \Gamma)] \\ &= E_{\Gamma} \left[\int_{A_i} \Lambda_i(x) dx \cdot \int_{A_{i'}} \Lambda_{i'}(y) dy | \Gamma \right] \\ &= E_{\Gamma} [(\Lambda_i | A_i)(\Lambda_{i'} | A_{i'}) | \Gamma] \\ &= E_{\Gamma} \left[|A_i| \sum_{l \in J} k_{il} \Gamma_l |A_{i'}| \sum_{m \in J} k_{i'm} \Gamma_m \right] \\ &= |A_i| |A_{i'}| \sum_{l \in J} \sum_{m \in J} k_{il} k_{i'm} E(\Gamma_l \Gamma_m). \end{aligned}$$

Jos δ_l^m on yksikkömatriisin (l, m) :s elementti, ylläoleva voidaan kirjoittaa muodossa

$$E[n_i n_{i'}] = |A_i| |A_{i'}| \sum_{l \in J} \sum_{m \in J} k_{il} k_{i'm} [\delta_l^m E \Gamma_l^2 + (1 - \delta_l^m) E \Gamma_l E \Gamma_m]$$

ja kovarianssiksi saadaan ($i \neq i'$)

$$\begin{aligned} \text{Cov}[n_i, n_{i'}] &= E[n_i, n_{i'}] - En_i En_{i'} \\ &= |A_i| |A_{i'}| \sum_{l \in J} \sum_{m \in J} k_{il} k_{i'm} [\delta_l^m E \Gamma_l^2 + (1 - \delta_l^m) E \Gamma_l E \Gamma_m - E \Gamma_l E \Gamma_m] \\ &= |A_i| |A_{i'}| \sum_{l \in J} \sum_{m \in J} k_{il} k_{i'm} \delta_l^m [E \Gamma_l^2 - E \Gamma_l E \Gamma_m] \\ &= |A_i| |A_{i'}| \sum_{l \in J} k_{il} k_{i'l} \text{Var}(\Gamma_l) \\ &= |A_i| |A_{i'}| \sum_{l \in J} k_{il} k_{i'l} \alpha_l \beta_l^2. \end{aligned}$$

Jos $i = i'$, niin tällöin

$$\begin{aligned}
E[n_i^2] &= E_\Gamma E[n_i^2 | \Gamma] = E_\Gamma[\Lambda_i | A_i] + (\Lambda_i | A_i)^2 | \Gamma] \\
&= |A_i| E \sum_{l \in J} k_{il} \Gamma_l + |A_i|^2 E \left(\sum_{l \in J} k_{il} \Gamma_l \right)^2 \\
&= |A_i| \sum_{l \in J} k_{il} \alpha_l \beta_l + |A_i|^2 \sum_{l \in J} \sum_{m \in J} k_{il} k_{im} [\delta_l^m E \Gamma_l^2 + (1 - \delta_l^m) E \Gamma_l E \Gamma_m],
\end{aligned}$$

jolloin varianssiksi saadaan vastaavasti

$$\begin{aligned}
\text{Var}(n_i) &= \text{Cov}[n_i, n_i] = E[n_i^2] - (E n_i)^2 \\
&= |A_i| \sum_{l \in J} k_{il} \alpha_l \beta_l + |A_i|^2 \sum_{l \in J} k_{il}^2 [E \Gamma_l^2 - (E \Gamma_l)^2] \\
&= |A_i| \sum_{l \in J} k_{il} \alpha_l \beta_l + |A_i|^2 \sum_{l \in J} k_{il}^2 \alpha_l \beta_l^2.
\end{aligned}$$

Kovarianssi voidaan nyt kirjoittaa yleisesti muodossa

$$\text{Cov}[n_i, n_{i'}] = \sum_{l \in J} [\delta_i^{i'} |A_i| k_{il} \alpha_l \beta_l + |A_i| |A_{i'}| k_{il} k_{i'l} \alpha_l \beta_l^2],$$

jossa $\delta_i^{i'}$ on yksikkömatriisin (i, i') :s elementti. Jos kerroinmatriisi K on yksikkömatriisi, niin silloin

$$\text{Var}(n_i) = |A_i| \alpha_i \beta_i \quad \text{ja} \quad \text{Cov}[n_i, n_{i'}] = 0.$$

3.5.4 Muita Cox-prosesseja

Muita Cox-prosesseja ovat esimerkiksi Matérnin klusteriprosessia muistuttava Thomas-prosessi (Stoyan ym., 1995, s. 160) ja satunnaisiin Poisson-tesselaatioihin perustuva Heikkinen–Arjas-prosessi (Heikkinen & Arjas, 1998). Thomas-prosessissa satunnaiskenttänä on

$$\Lambda(x) = \eta \sum_i h(x - \xi_i),$$

jossa $\{\xi_1, \dots, \xi_K\}$ Poisson-prosessi ja h on 2-ulotteinen normaalijakauma. Heikkinen–Arjas -prosessissa satunnaiskenttänä on

$$\Lambda(x) = \sum_{k=1}^K \eta_k 1_{E_k(\xi_k)}(x),$$

jossa $\{E_k\}$ on pisteiden ξ_1, \dots, ξ_K Voronoi-tesselaatio ja η_k on porraskäyrän arvo tesselaation solussa k .

3.6 Pisteprosessien simulointialgoritmeja

Tarkastellaan Poisson-prosessin ja Cox-prosessin simulointia suorakulmiossa $A = (0, a) \times (0, b)$, jonka pinta-ala on $|A|$. Olkoon Poisson-prosessin intensiteettifunktio $\lambda(x)$, joka homogeenisessa tapauksessa on vakio λ .

3.6.1 Homogeenisen Poisson-prosessin simulointi

Koska homogeenisen Poisson-prosessin pisteet sijaitsevat tasajakautuneesti toisistaan riippumatta, voidaan pisteet generoida tasajakaumasta seuraavalla prosessin määritelmään perustuvalla algoritmilla:

1. Generoidaan pisteiden lukumäärä Poisson-jakaumasta parametrina $\lambda|A|$.
2. Generoidaan pisteen x ; x -koordinaatti tasajakaumasta väliltä $(0, a)$ ja y -koordinaatti tasajakaumasta väliltä $(0, b)$.
3. Toistetaan kohtaa 2 toisistaan riippumatta niin kauan, kunnes kohdan 1 osoittama määrä pisteitä on generoitu.

Jos simuloitava alue on muu kuin suorakulmio, voidaan pisteet generoida edellisen algoritmin mukaan suorakulmiolle, joka sisältää kyseisen alueen. Jos generoitu piste ei kuulu kyseiselle alueelle, se hylätään ja generoidaan uusi piste-ehdotus.

Toinen vaihtoehto simuloida homogeeninen Poisson-prosessi on kertymäfunktio menetelmä (Lewis & Shedler, 1979), jossa pisteet generoidaan sekventiaalisesti seuraavin periaattein: pisteiden x -koordinaatit järjestettynä, $x_{(1)}, \dots, x_{(n)}$, muodostavat yksiulotteisen homogeenisen Poisson-prosessin intensiteettinä λb , mikä on seurausta siitä, että x -koordinaattien välit ovat riippumattomia realisaatioita eksponentiaalisesti jakautuneesta satunnaismuuttujasta. Sen kertymäfunktio on

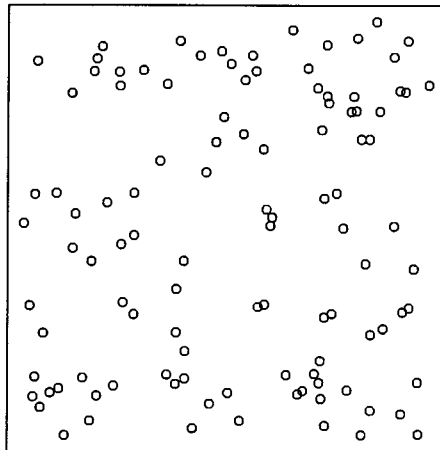
$$F(v) = 1 - e^{-\lambda b v},$$

jossa v on siis järjestettyjen ja perättäisten x -koordinaattien väli. Tällöin x -koordinaatit saadaan määrätyn tämän kertymäfunktion avulla siten, että jos $U \sim \text{tas}(0, 1)$, niin $V = -\frac{1}{\lambda b} \log U \sim \exp(\lambda b)$. Jokaista x -koordinaattia vastaava y -koordinaatti saadaan tasajakaumasta. Tällöin algoritmi on seuraava:

1. Generoidaan luku u_i tasajakaumasta väliltä $(0, 1)$.

2. Lasketaan $v_i = -\frac{1}{\lambda b} \log u_i$.
3. Pisteen x_i x -koordinaatti on $\sum_{j=1}^i v_j$.
4. Generoidaan pisteen x_i y -koordinaatti tasajakaumasta väliltä $(0, b)$.
5. Toistetaan kohtia 1–4 ja lopetetaan, kun $\sum_j v_j > a$.

Huomattakoon, että menetelmä generoi automaattisesti pisteen (x_i, y_i) x -koordinaatin kasvaessa 0:sta a :han ja pisteiden lukumäärä simuloinnin jälkeen on Poisson-jakautunut parametrilla $\lambda|A|$. Kuvassa 3.1 on simuloitu esimerkki homogeenisesta Poisson-prosessista.



Kuva 3.1 Realisaatio homogeenisesta Poisson-prosessista. Prosessi on täysin satunnainen prosessi, sillä prosessin pisteet sijaitsevat tasajakautuneesti toisistaan riippumatta.

3.6.2 Epähomogeenisen Poisson-prosessin simulointi

Epähomogeenisen Poisson-prosessin simuloinnissa voidaan käyttää joko harvennusmenetelmää tai hylkäysmenetelmää.

Harvennusmenetelmän algoritmi on seuraava:

1. Simuloidaan alueelle homogeeninen Poisson-prosessi, jonka intensiteetti on $\lambda_0 \geq \max_{x \in A} \lambda(x)$.

2. Lasketaan jokaiselle simuloidulle pisteelle x_i arvo $\frac{\lambda(x_i)}{\lambda_0}$, joka on siis aina välillä $(0, 1)$.
3. Generoidaan väliltä $(0, 1)$ tasajakautunut satunnaisluku u_i jokaista pistettä x_i kohti.
4. Piste x_i hyväksytään, jos $\frac{\lambda(x_i)}{\lambda_0} \geq u_i$. Piste x_i poistetaan, jos $\frac{\lambda(x_i)}{\lambda_0} < u_i$.

Huomattakoon, että harvennuksen jälkeen pisteiden lopullinen lukumäärä on Poisson-jakautunut parametrina $\int_A \lambda(x) dx$.

Hylkäysmenetelmää käyttäen prosessi simuloidaan seuraavasti:

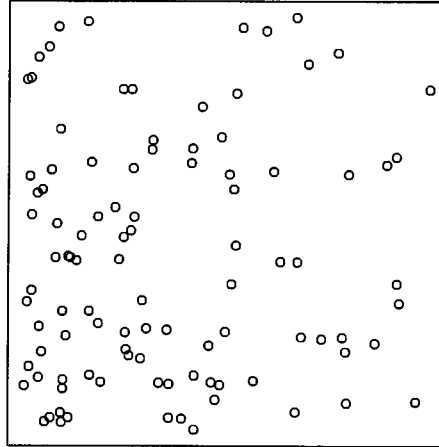
1. Generoidaan pisteiden lukumäärä Poisson-jakaumasta parametrina $\int_A \lambda(x) dx$.
2. Generoidaan alueelle piste x_i tasajakaumasta.
3. Lasketaan intensiteetti $\lambda(x)$ pisteessä x_i .
4. Arvotaan luku v_i tasajakaumasta väliltä $(0, \lambda_0)$, jossa $\lambda_0 \geq \max_{x \in A} \lambda(x)$.
5. Piste x_i otetaan mukaan, jos $\lambda(x_i) \geq v_i$. Muussa tapauksessa generoidaan edellisestä riippumaton uusi piste x_i ja toistetaan kohdat 4 ja 5.
6. Kun kohdan 1 osoittama määrä pisteitä on generoitu, lopetetaan.

Kuvassa 3.2 on simuloitu esimerkki epähomogeenisen Poisson-prosessin realisaatiosta, jossa $\lambda(x, y) = e^{-2x-y}$.

3.6.3 Cox-prosessi

Cox-prosessin simuloinnissa käytetään hyväksi epähomogeenisen Poisson-prosessin simulointialgoritmeja, sillä intensiteetillä ehdollistettu Cox-prosessi on epähomogeeninen Poisson-prosessi. Olennaista uutta edelliseen on satunnaiskentän $\{\Lambda(x)\}$ simulointi. Cox-prosessin simulointi on kaksivaiheinen:

1. Generoidaan satunnaiskenttä $\{\Lambda(x)\}$.
2. Ehdolla $\{\Lambda(x) = \lambda(x)\}$ generoidaan epähomogeeninen Poisson-prosessi, jonka intensiteettifunktio on $\lambda(x)$.



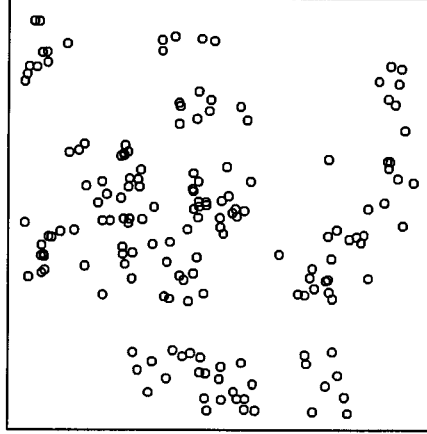
Kuva 3.2 Realisaatio epähomogeenisesta Poisson-prosessista, jonka intensiteettifunktio on $\lambda(x, y) = e^{-2x-y}$. Realisaatiosta on havaittavissa selvä trendi: pisteet harvenevat oikealle ja ylöspäin siirryttäessä.

Matérnin klusteriprosessin ja gamma-Poisson-rasteriprosessin simulointiohjelmat on esitetty liitteessä 1. Gaussisen satunnaiskentän simulointialgoritmejä tarkastelee mm. Møller ym. (1998). Kuvassa 3.3 on realisaatio Matérnin klusteriprosessista ja kuvassa 3.4 realisaatiot sekä tasoittamattomasta että tasoitetusta gamma-Poisson-prosessista.

3.7 Bayes-menetelmä

Olkoon satunnaismuuttuja X , jonka jakauma on $p(x|y)$ ($= p(x; y)$) ja jossa y on tuntematon parametri. Frekventistisen ajattelutavan mukaan y :tä estimoitaessa muodostetaan uskottavuusfunktio ja maksimoidaan se y :n suhteen. Parametri y ajatellaan siis kiinteäksi vakioksi.

Bayesiläisessä päättelyssä tuntematon y ajatellaan satunnaismuuttujaksi. Parametriin y liittyy siis epävarmuutta, jota voidaan kuvata asettamalla sille *priorijakauma* $p(y)$. Se kuvaa tutkijan käsityksiä y :stä ennen havainnointia, ts. se kertoo kuinka todennäköisiä y :n eri arvot ovat. Bayes-menetelmässä muodostetaan y :lle ”päivitetty” jakauma, *posteriorijakauma*, $p(y|x)$, ottamalla aineiston $x = \{x_1, \dots, x_n\}$ sisältämä informaatio huomioon. Tässä käyte-



Kuva 3.3 Realisaatio Matérnin klusteriprosessista, jossa $R = 0.1$, $\rho = 15$ ja $\eta = 250$.

tään hyväksi Bayesin kaavaa:

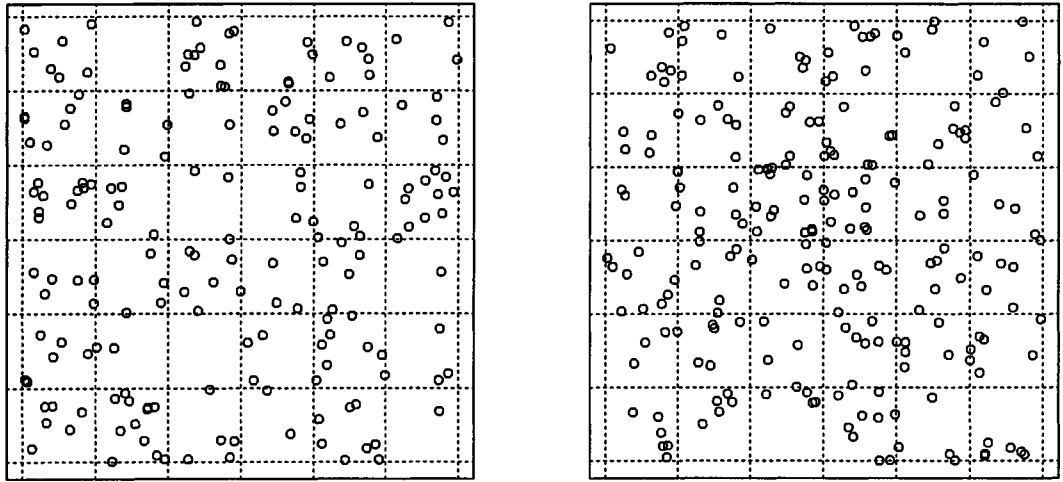
$$p(y|x) = \frac{p(y)p(x|y)}{\int p(y)p(x|y) dy} = \frac{p(y)p(x|y)}{p(x)}, \quad (3.16)$$

jossa $p(x|y)$ on aineistosta laskettu uskottavuus ja $p(x)$ skaalaustekijä (=vakio) eikä siten vaikuta jakauman muotoon. Täten voidaan kirjoittaa

$$p(y|x) \propto p(y)p(x|y) \quad (3.17)$$

eli posteriorijakauma on suoraan verrannollinen priorijakauman ja uskottavuuden tuloon. Posteriori $p(y|x)$ voidaan tulkita priorin ja datan ”kompromissiksi” eli se kuvaa tutkijan uskomuksia y :stä, kun data x on otettu huomioon. Päätelmät y :stä tehdään tarkastelemalla posteriorijakaumaa, josta voidaan laskea y :lle tunnuslukuja.

Bayes-teorian periaatteita ja tekniikoita käsittelee yleisellä tasolla mm. Gelman ym. (1995). Ne ovat sellaisenaan sovellettavissa pisteprosessien bayesiläiseen analyysiin (ks. esim. Møller ym., 1998; Heikkinen & Arjas, 1998).



Kuva 3.4 Realisaatioita gamma-Poisson-rasteriprosesseista, joissa $\alpha = 5$ ja $\beta = 1$. Vasemmalla on tasoittamaton ja oikealla kernel-tasoitettu gamma-Poisson-prosessi.

4 Kenttähavainnoinnin menetelmät intensiteetin estimoinnissa

Suuren alueen tilajärjestyksen ja intensiteetin täydelliseen tilajärjestyskarttaan pohjautuva analysointi on lähes mahdotonta, koska se on erittäin työlästä. Niinpä suuren alueen analysoinnissa käytetään kenttähavainnointia, joka on rinnastettavissa otantatutkimukseen. Tällöin tutkittavasta alueesta tarkastellaan vain useita tutkimusalueeseen suhteutettuna pieniä koealoja, joiden valinta voi perustua otantaan. Tiheyden kuvaaminen alueesta poimitun otoksen avulla on lisäksi varsin luonnollista, sillä tiheys on usein tasaisesti muuttuva ominaisuus.

Tässä luvussa on esitelty kenttähavainnointiin liittyviä yleisiä käsitteitä ja menetelmiä intensiteetin estimointiin. Ensin on tarkasteltu klassisia menetelmiä, lukumäärä- ja etäisyysmenetelmää, ja lopuksi satunnaiskenttämenetelmää ja hierarkkista mallinnusta riippuvien havaintojen kuten ryväsotannan tapauksessa.

4.1 Yleistä

Kenttähavainnoinnin tavoitteena on alueellista otantaa hyödyntäen hankkia tietoa tutkittavasta alueesta ilman, että alueen jokainen objekti täytyy tutkia erikseen. Niinpä tutkittavalta alueelta valitaan m pistettä, s_1, \dots, s_m , joita kutsutaan *otantapisteksiksi*.

Otantapisteissä voidaan tehdä pistekuviosta joko lukumäärähavainnointia tai etäisyyshavainnointia. Lukumäärähavainnot $N(B_{s_1}), \dots, N(B_{s_m})$ tehdään

otantakoealoilta B_{s_1}, \dots, B_{s_m} , jotka ovat $|B|$:n kokoisia alueita otantapisteen ympärillä. Koealat ovat tyypillisesti ympyröitä tai neliöitä ja jokaisen koealan B_{s_i} ”karakterisoivana” pisteenä (esim. keskipisteenä) on otantapiste s_i .

Etäisyshavainnoinnissa voidaan tarkastella etäisyyttä otantapistestä lähimpään naapuriin, toiseksi lähimpään tai k :nneksi lähimpään, etäisyyttä satunnaisesta objektista lähimpään objektiin tai näiden modifikaatioita kuten T-otanta (Besag & Gleaves, 1973).

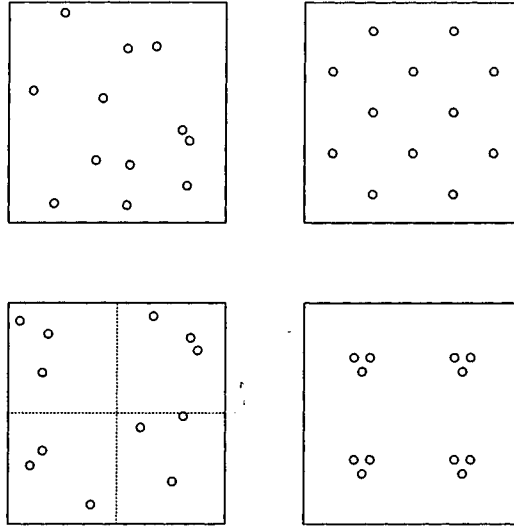
Otantapistheet voidaan valita esimerkiksi tasaisella satunnaisotannalla, systemaattisella otannalla, ositetulla otannalla tai ryväotannalla. Jos otantapistheet ovat suurella todennäköisyydellä erillään toisistaan, havaintoja voidaan pitää lähes stokastisesti riippumattomina. Tällaisesta otannasta käytetään nimitystä *harva otanta*. Ryväotantaa lukuunottamatta edellä luetellut otantatavat täyttävät yleensä harvan otannan oletukset.

Pistekuviota voidaan havainnoida myös poimimalla otantapisteen sijaan satunnaisia objekteja ja mitata niihin liittyviä suureita kuten lähinaapurietäisyyttä. Satunnaisen objektin poiminta on kenttähavainnoinnin yhteydessä kuitenkin ongelmallisempaa. Esimerkiksi satunnaisen otantapisteen lähin objekti ei täytä satunnaisen objektin vaatimusta. Byth & Ripley (1980) ehdottavat menetelmää, jossa valitaan pieni alue otantapisteen ympäriltä. Alueen objektit numeroidaan ja niiden joukosta arvotaan yksi. Tämä objekti toteuttaa likimain satunnaisen objektin vaatimuksen.

Kenttähavainnointiin perustuvia tilastomenetelmiä kutsutaan *kenttämenetelmiksi*. Perinteisesti niitä ovat *lukumäärämenetelmä* ja *etäisyysmenetelmä*, ja niihin liittyvä klassinen tilastoanalyysi toimii harvan otannan tapauksessa. Molempia menetelmiä voidaan käyttää sekä tilajärjestyksen analyysiin että intensiteetin estimointiin. Lukumäärämenetelmä sopii kuitenkin paremmin intensiteetin estimointiin, sillä etäisyysmenetelmä on herkkä poikkeamalle Poisson-oletuksesta. Etäisyysmenetelmä taas sopii paremmin tilajärjestyksen analyysiin, sillä se ei ole aivan niin tiukasti skaalaan sidottu kuin lukumäärämenetelmä. Tosin myös etäisyysmenetelmä on pienen skaalan menetelmä.

Lukumäärämenetelmä perustuu pisteiden lukumääriin $N(B_{s_1}), \dots, N(B_{s_m})$, jossa B_{s_1}, \dots, B_{s_m} ovat koealoja, joiden pinta-ala on $|B|$. Jos kyseessä on Poisson-prosessi, jokaisen koealan pisteiden lukumäärä noudattaa Poisson-jakaumaa parametrilla $\lambda|B|$, ts.

$$P(N(B) = n) = \frac{(\lambda|B|)^n}{n!} e^{-\lambda|B|}, \quad n = 0, 1, \dots \quad (4.1)$$



Kuva 4.1 Neljä spatiaalista otanta-asetelmaa 12 otantapisteeille (vasemmalta oikealle, ylhäältä alas): tasainen satunnaisotanta, systemaattinen otanta, ositettu otanta ja (systemaattinen) ryväotanta.

Lukumäärämenetelmässä aineiston muodostavat koealojen lukumäärät n_1, \dots, n_m , joten pisteiden tarkkaa sijaintia ei tarvitse tietää. Intensiteetin estimaattorin ominaisuuksiin kuten tehokkuuteen vaikuttaa luonnollisesti koealojen koko ja muoto, niiden määrä ja kuinka ne on sijoitettu tutkimusalueeseen.

Etäisyysmenetelmässä yksinkertaisinta on havaita etäisyys otantapistees-
tä lähimpään naapuriin, d_i , jolloin kyseessä on tyhjän tilan havainnointi otan-
tapisteen ympärillä. Intensiteetin estimointi perustuu etäisyyden jakauman
kertymäfunktioon

$$F(d) = 1 - e^{-\lambda\pi d^2}, \quad d > 0,$$

ja siitä saatuun ”tyhjän tilan” jakauman kertymäfunktioon

$$F(u) = 1 - e^{-\lambda u},$$

jossa $u = \pi d^2$ on tyhjä tila otantapisteen ympärillä.

Jos harvan otannan oletukset eivät jostain syystä ole voimassa tai koealojen valinnassa ei ole käytetty ollenkaan otantaa, ongelmaksi muodostuu

havaintojen välinen riippuvuus. Riippuvuus on otettava huomioon, sillä se vaikuttaa intensiteetin estimaattorin varianssiin ja sen estimointiin. Eräs tapa ottaa riippuvuus huomioon on satunnaiskenttämenetelmä, joka mallintaa riippuvuuden suoraan kovarianssin avulla. Toinen tapa ottaa riippuvuus huomioon approksimatiivisesti on hierarkkinen mallinnus. Tämä sopii erityisesti heterogeenisille aineistoille, jotka aiheuttavat tyypillisesti ongelmia perinteisten menetelmien yhteydessä. Hierarkkisella mallilla on lisäksi paljon etuja: Se mahdollistaa mm. intensiteetin vaihtelun kuvaamisen sekä rypään sisäisen korrelaation approksimoiminnin.

4.2 Klassinen intensiteetin estimointi

Perinteisesti intensiteetin estimointi kenttämenetelmillä perustuu harvaan otantaan, jolloin havainnot koealoissa voidaan pitää riippumattomina. Tällöin intensiteettiä voidaan estimoida lukumäärä- ja etäisyysmenetelmällä. Etäisyysmenetelmä toimii vain Poisson-prosessin tapauksessa, mutta lukumäärämenetelmä ei vaadi Poisson-oletusta. Lukumäärämenetelmää voidaan käyttää myös heterogeenisessa tapauksessa keskimääräisen intensiteetin $\frac{1}{|B|} \int_B \lambda(x) dx$ estimointiin.

4.2.1 Lukumäärämenetelmä

Oletetaan yksinkertaisuuden vuoksi, että koealat ovat keskenään saman kokoisia ja muotoisia. Lisäksi oletetaan, että otantapisteen on valittu joko tasaiselle satunnaisotannalle tai systemaattiselle otannalle. Käytetään seuraavassa havainnoista $N(B_{s_i})$ lyhennettyä merkintää n_i . Koealojen satunnaisuuden huomioonottava estimaattori intensiteetille on pisteiden kokonaislukumäärä jaettuna koko pinta-alalla

$$\hat{\lambda} = \frac{\sum_{i=1}^m n_i}{m|B|}. \quad (4.2)$$

Estimaattori on homogeenisessä tapauksessa λ :n harhaton estimaattori, sillä

$$E\hat{\lambda} = \frac{1}{m|B|} \sum_{i=1}^m E n_i = \frac{1}{m|B|} m\lambda|B| = \lambda.$$

Vaikka saadun estimaattorin harhattomuus ei riipu jakaumasta, sen ominaisuudet riippuvat. Esimerkiksi varianssi riippuu paitsi otanta-asetelmasta, koealojen koosta ja muodosta, tarkasteltavan alueen kokonaispinta-alasta ja taustalla olevan prosessin kovarianssirakenteesta, ks. kaava 3.9. Varianssi on

harvan otannan tapauksessa

$$\begin{aligned}
\text{Var}(\hat{\lambda}) &= \text{Var}\left(\frac{\sum_{i=1}^m n_i}{m|B|}\right) \approx \frac{1}{m^2|B|^2} \sum_{i=1}^m \text{Var}(n_i) \\
&= \frac{1}{m^2|B|^2} \sum_{i=1}^m \left\{ \lambda|B|(1 - \lambda|B|) + \int_B \int_B \lambda_2(x - y) dx dy \right\} \\
&= \frac{\lambda}{m|B|} \left\{ 1 - \lambda|B| + \frac{1}{\lambda|B|} \int_B \int_B \lambda_2(x - y) dx dy \right\} \\
&= \frac{\lambda}{m|B|} \left\{ 1 - \lambda|B| + \frac{\lambda}{|B|} \int_B \int_B g(x - y) dx dy \right\} \\
&= \frac{\lambda}{m|B|} \{1 - \lambda|B| + \lambda|B|E_B g(U - V)\},
\end{aligned}$$

missä (U, V) sisältää kaksi B :ssa tasajakautunutta riippumatonta satunnaismuuttujaa. $E_B g(U - V)$:n laskemiseen voidaan käyttää luvun 3.2.1 eksakteja ja asymptoottisia tuloksia. Poisson-prosessille

$$\text{Var}(\hat{\lambda}) = \frac{\lambda}{m|B|}.$$

Estimointitilanteessa λ on tuntematon. Silloin voidaan käyttää ”lähes” harhatonta varianssiestimaattoria

$$\widehat{\text{Var}}(\bar{n}) = \frac{\sum_{i=1}^m (n_i - \bar{n})^2}{m(m-1)}.$$

Ositetun otannan tapauksessa intensiteetin estimaattori saadaan painotettuna keskiarvona

$$\hat{\lambda} = \sum_{i=1}^K \frac{|A_i|}{a} \hat{\lambda}_i, \quad (4.3)$$

jossa A_1, \dots, A_K ovat ositteet, $a = \sum_{i=1}^K |A_i|$ ja $\hat{\lambda}_i$ on kaavan 4.2 mukaan laskettu intensiteetin estimaattori ositteessa i . Varianssiksi saadaan

$$\text{Var}(\hat{\lambda}) = \sum_{i=1}^K \left(\frac{|A_i|}{a}\right)^2 \text{Var}(\hat{\lambda}_i).$$

4.2.2 Etäisyysmenetelmä

Tarkastellaan tyhjän tilan tunnuslukuun perustuvaa etäisyysmenetelmää, joka toimii vain Poisson-prosessille. Oletetaan, että jokaisesta otantapistestä

m on mitattu etäisyys d_i lähimpään prosessin pisteeseen. Prosessin tiheyttä voidaan estimoida etäisyyksien d_i jakauman avulla. Etäisyyksiin perustuvan menetelmän yhteydessä on kuitenkin luonnollista estimoida tiheyden λ sijasta parametria $\gamma = \lambda^{-1}$, joka kuvaa keskimääräistä alaa pistettä kohti. Poisson-hypoteesin ollessa voimassa γ :n suurimman uskottavuuden estimaattori on

$$\hat{\gamma} = \frac{\pi}{m} \left(\sum_{i=1}^m d_i^2 \right). \quad (4.4)$$

Estimaattori on γ :n harhaton estimaattori ja sen asymptoottinen varianssi on

$$\text{Var}(\hat{\gamma}) = \frac{\gamma^2}{m}.$$

Tyhjän tilan tunnuslukuun perustuva estimaattori on yksinkertaisin intensiteetin estimaattori, mutta se on hyvin herkkä poikkeamalle Poisson-oletuksesta. Muita etäisyyksimenetelmään liittyviä estimaattoreita löytyy kirjasta Diggle (1983).

4.3 Riippuvuuden huomioon ottaminen -satunnaiskenttämenetelmä

Oletetaan, että koealoissa tehtyjä havaintoja ei voida pitää stokastisesti riippumattomina, ts. harvan otannan oletukset eivät ole voimassa. Tarkastellaan pisteiden lukumääriin liittyvää menetelmää, joka mallintaa havaintojen välisen riippuvuuden jatkuvaparametrisen satunnaiskentän kovarianssin avulla. Menetelmä on identtinen edellisessä kappaleessa esitetyn lukumäärämenetelmän kanssa, jos havainnot ovat riippumattomat.

Olkoon N pisteiden sijaintia mallintava pisteprosessi, jonka intensiteetti on $\lambda(x)$. Oletetaan lisäksi, että koealat ovat identtiset. Määritellään

$$Y(s) = \frac{1}{|B|} N(B_s), \quad (4.5)$$

jossa $B_s = B_0 + s = \{x + s | x \in B_0\}$ ja s on koalaa B_s karakterisoiva piste. Tällöin $\{Y(s), s \in R^2\}$ on jatkuvaparametrinen satunnaiskenttä. $\{Y(s), s \in R^2\}$ on stationaarinen, jos N on stationaarinen ja edelleen isotrooppinen, jos N on isotrooppinen. Tämän muunnoksen avulla estimointilanne palautuu Matérnin (1960) kehittämään teoriaan otannasta toisen kertaluvun jatkuvaparametrisesta satunnaiskentästä.

Koealojen ollessa kiinteitä satunnaiskentän odotusarvo stationaariselle prosessille on

$$E[Y(s)] = \frac{1}{|B|} E[N(B_s)] = \frac{1}{|B|} \lambda |B| = \lambda$$

ja kovarianssi

$$\begin{aligned} \text{Cov}[Y(s), Y(t)] &= E[Y(s)Y(t)] - E[Y(s)]E[Y(t)] \\ &= \frac{1}{|B|^2} E[N(B_s)N(B_t)] - \lambda^2 \\ &= \frac{1}{|B|^2} \left\{ \lambda |B_s \cap B_t| + \int_{B_s} \int_{B_t} \lambda_2(x-y) dx dy \right\} - \lambda^2. \end{aligned}$$

Varianssi saadaan tästä erikoistapauksena ($s = t$):

$$\text{Var}[Y(s)] = \frac{1}{|B|^2} \left\{ \lambda |B| + \int_B \int_B \lambda_2(x-y) dx dy \right\} - \lambda^2.$$

Stationaarisesta satunnaiskentästä poimitun otoksen $Y(s_1), \dots, Y(s_n)$ avulla saadaan intensiteetille estimaattori

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y},$$

joka on λ :n harhaton estimaattori edellä mainituille otanta-asetelmille. Intensiteetin estimointi palautuu siis keskiarvon estimointiin (vrt. kaava 4.2). Estimaattorin varianssi saadaan satunnaiskentän kovarianssin avulla.

Jos koealojen valinnassa on käytetty otantaa, tulee koealojen satunnaisuus ottaa huomioon ottamalla odotusarvot myös asetelman suhteen (Matérn, 1960). Klassisesti pyritään harvaan otantaan, jolloin $\text{Cov}[Y(s_i), Y(s_j)] \approx 0$, kun $i \neq j$. Tällöin voidaan siis käyttää kappaleessa 4.2 esiteltyjä perinteisiä kenttämenetelmiä.

4.4 Hierarkkinen mallintaminen

Kuten jo edellä todettiin, havaintojen välinen riippuvuus voi aiheuttaa intensiteetin estimoinnissa ongelmia. Lisäksi yleisenä ongelmana on jonkin, joko havaitun tai havaitsemattoman, taustamuuttujan aiheuttama heterogeisuus. Esimerkiksi metsän heterogeisuuden syynä on usein mm. maaperän laatu, korkeus merenpinnasta, maankäytön historia, taudit tai tuhot.

Mahdollinen havaintojen välinen riippuvuus ja aineiston heterogeisuus voidaan havaita aineistosta helposti. Havaintojen välinen riippuvuus näkyy

sisäkorrelaationa, jolloin esimerkiksi ryväsotannan yhteydessä saman rypään havainnot muistuttavat enemmän toisiaan kuin havainnot keskimäärin. Heterogeenisuus on usein havaittavissa klusteroitumisena, mikä puolestaan näkyy aineistossa ylihajontana. Ylihajonnalla tarkoitetaan tässä tilanteita, jossa lukumäärähavainnoille

$$\text{Var}[N(A)] > E[N(A)].$$

Aineistossa esiintyvä riippuvuus ja heterogeenisuus aiheuttavat tyypillisesti ongelmia mm. varianssin estimoinnissa lukumäärämenetelmän yhteydessä ja ylipäättään tiheyden estimoinnissa etäisyysmenetelmällä. Heterogeenisen aineiston ollessa kyseessä ongelmana on lisäksi heterogeenisuuden voimakkuuden arviointi. Kuten kappaleessa 4.2 todettiin, lukumäärämenetelmää voidaan heterogeenisessä tilanteessa käyttää keskimääräisen intensiteetin estimointiin, mutta tällöin estimaattorin varianssi voi tulla hyvinkin suureksi. Jos heterogeenisuus on ennalta odotettua, on keskimääräisen intensiteetin sijasta mielekkäämpää tarkastella, kuinka intensiteetti alueella vaihtelee. Jotta aluekohtaisen tiedon saaminen olisi mahdollista, käytetään riippuvuusrakenteen kuvaamisessa hierarkkista mallia.

Hierarkkisen mallin etu perinteiseen sekoitettuun jakaumaan verrattuna tulee esille siinä, että kun esimerkiksi ylihajontatilanteessa sovitetaan havainnoille sekoitettua jakaumaa, joka tässä tapauksessa voisi olla gamma-Poisson-jakauma

$$\begin{aligned} f(n_i|\alpha, \beta) &= \int p(n_i|\lambda)p(\lambda|\alpha, \beta) d\lambda \\ &= \frac{\Gamma(\alpha + n_i)}{\Gamma(\alpha)} \left(\frac{\beta}{|A| + \beta} \right)^\alpha \left(\frac{|A|}{|A| + \beta} \right)^{n_i}, \end{aligned}$$

jonka varianssi on suurempi kuin Poisson-jakauman varianssi. Koska intensiteetti λ on integroitu pois, mallintavat parametrit α ja β suoraan pisteiden lukumääriä koealoissa eikä havaitsematon intensiteetti ole mukana mallissa. Tällöin intensiteettiä ei voida estimoida suoraan jakauman avulla. Hierarkkisessa mallissa intensiteettitaso on mukana, jolloin saadaan helposti estimaatit rypäille.

Hierarkkinen mallinnus on luontevaa ryväsotannan yhteydessä, sillä ryväsotanta on kaksiasteinen: ensin poimitaan rypäät ja sitten koealat. Siten rypäät ovat hierarkiassa ensimmäisellä (ylimmällä) tasolla ja koealat toisella (alemmalla) tasolla. Ryväsotantaisten estimaattien avulla saadaan tietoa intensiteetin vaihtelusta tarkasteltavalla pienalueella ja koealojen (toistojen) avulla saadaan estimaatti keskimääräiselle intensiteetille kyseisessä rypäessä. Otosrypäät poimitaan tyypillisesti harvalla otannalla, jolloin ne oletetaan

keskenään riippumattomiksi. Otosrypään sisällä havainnot ovat puolestaan riippuvia, koska ne ovat suurella todennäköisyydellä lähellä toisiaan.

Havainnoille, jotka ovat joko lukumäärä- tai etäisyyshavaintoja, muodostetaan kaksitasoinen hierarkkinen malli, jossa on mukana havaitsematon intensiteetti. Hierarkkisen mallinnuksen ideana on, että mahdollisesti riippuvat koealataso havainnot oletetaan ehdollisesti riippumattomiksi ehdolla, että tiedetään, mihin rypäeseen ne kuuluvat.

Hierarkkinen mallinnus voidaan rinnastaa kiinteiden vaikutusten malliin (varianssianalyysimalliin). Esimerkiksi lukumäärähavainnoissa esiintyvä kokonaisvaihtelu voidaan osittaa rypäiden sisäiseen ja rypäiden väliseen vaihteluun

$$\sum_{i=1}^m \sum_{j=1}^k (n_{ij} - \bar{n})^2 = \sum_{i=1}^m \sum_{j=1}^k (n_{ij} - \bar{n}_i)^2 + k \sum_{i=1}^m (\bar{n}_i - \bar{n})^2, \quad (4.6)$$

jossa m on rypäiden lukumäärä, k koealojen lukumäärä rypäessä, \bar{n} on kaikkien koealojen keskiarvo ja \bar{n}_i on rypään i koealojen keskiarvo. Rypäiden välinen vaihtelu kuvaa heterogeenisuutta ja sisäinen vaihtelu lokaalia rypään sisäistä vaihtelua mitatulla alueella. Hierarkkisella mallilla voidaan siis arvioida otosrypään sisäistä ja otosrypäiden välistä vaihtelua. Lisäksi sen avulla voidaan analysoida rypään sisäistä korrelaatiota.

Vastaava dekompositio voitaisiin tehdä myös kiinteiden vaikutusten malleja käyttäen, mutta silloin kutakin ryvästä vastaisi oma intensiteettiparametri. Hierarkkista mallia käyttäen selvittää huomattavasti vähemmällä parametreilla.

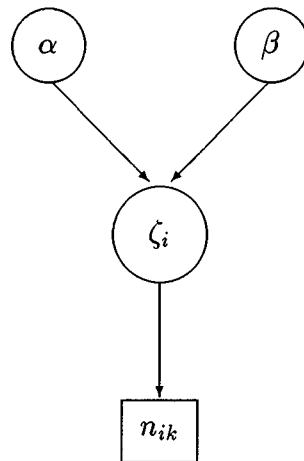
Seuraavassa on sovellettu hierarkkista mallia sekä lukumäärämenetelmään että etäisyysmenetelmään. Molemmissa tapauksissa on tarkasteltu myös bayesiläistä lähestymistapaa. Bayes-menetelmä on esitetty yleisesti luvussa 3.7.

4.4.1 Hierarkkinen mallinnus lukumäärähavainnoille

Oletetaan, että koealat on poimittu ryväsoitannalla. Oletetaan edelleen, että intensiteetti muuttuu alueella hitaasti ja prosessi on lokaalisti Poisson-prosessi. Tällöin saman rypään koealoilla on sama intensiteetti λ_i , jolloin rypäessä i havaintojen lukumäärä koealassa noudattaa Poisson-jakaumaa parametrina $\lambda_i |B_{ik}|$. Merkitään $\zeta_i = \lambda_i |B_{ik}|$ ja oletetaan ζ_i :n jakaumaksi Gamma(α, β^{-1}). Mallina on siis itse asiassa gamma-Poisson -prosessi (vrt. luku 3.5.3). Huomattakoon, että jatkossa käytetään myös parametrissa ζ nimitystä intensiteetti. Parametri ζ on pinta-alalla kerrottu intensiteetti,

jolloin se on vain eri skaalassa (ilmoittaa lukumäärää) kuin parametri λ . Täten, kun ζ :n jakaumana on $\text{Gamma}(\alpha, \beta^{-1})$, on parametrin λ jakauma $\text{Gamma}(\alpha, |B|\beta^{-1})$. Koska myöhemmin Bayes-lähestymistavassa ei tehdä eroa havainnon ja parametrin välillä, vaan molempia käsitellään satunnaismuuttujina, niin yhdenmukaisuuden vuoksi merkitään λ :aa ja ζ :aa pienillä kirjaimilla.

Lukumäärähavaintojen hierarkkinen rakenne on esitetty kuvassa 4.2. Malli on kaksitasoinen: ylimmällä tasolla ovat gammajakauman parametrit α ja β ja näiden alapuolella tuntematon intensiteetti ζ_i . Havainnot $N(B_{ik})$ riippuvat intensiteeteistä ζ_i ja intensiteetit riippuvat edelleen gammajakauman parametreista. Havainnot $N(B_{ik})$ ovat keskenään riippuvia. Hierarkkisesta rakenteesta ja Poisson-oletuksesta seuraa kuitenkin, että havainnot $N(B_{ik})$ ovat ehdollisesti riippumattomia ehdolla ζ_i , ts. havainnot ovat ehdollisesti riippumattomia ehdolla, että ne kuuluvat samaan rypäeseen.



Kuva 4.2 Hierarkkinen rakenne lukumäärähavainnoille. Havaintojen lukumäärä n_{ik} riippuu tuntemattomasta ryväsintensiteetistä ζ_i , joka puolestaan riippuu gammajakauman parametreista α ja β .

Intensiteetin ja sen vaihtelun kuvaamisessa huomio kohdistuu gammajakauman parametreihin α ja β . Tarkasteltavan alueen A keskimääräisen intensiteetin $\frac{1}{|A|} \int_A \lambda(x) dx$ estimaatti on $\frac{\alpha\beta}{|B|}$ ja odotettu objektien lukumäärä $\frac{\alpha\beta|A|}{|B|}$. Kun estimoidaan keskimääräinen intensiteetti jokaiselle otosrypälle erikseen, voidaan alueesta muodostaa interpoloinnilla (Ripley, 1981, s. 28)

intensiteettipinta.

Kuten aikaisemmin todettiin, voidaan havainnoissa esiintyvä kokonaisvaihtelu osittaa rypäiden väliseen ja rypäiden sisäiseen vaihteluun kaavan 4.6 mukaan. Kokonaisvaihtelulle saadaan nyt approksimaatio

$$\begin{aligned}\text{Var}(n_{ij}) &= E[\text{Var}(n_{ij}|\zeta)] + \text{Var}[E(n_{ij}|\zeta)] \\ &= E(\zeta_i) + \text{Var}(\zeta_i) = \alpha\beta + \alpha\beta^2 = \alpha\beta(1 + \beta).\end{aligned}\quad (4.7)$$

Rypäiden sisäistä vaihtelua approksimoidaan varianssilla $\text{Var}(\zeta_i) = \alpha\beta^2$, jolloin rypäiden väliseksi vaihteluksi saadaan $\alpha\beta$. Edelleen havaintojen väliselle sisäkorrelaatiolle saadaan approksimaatio

$$\rho = \frac{\text{Var}(\zeta_i)}{\text{Var}(n_{ij})} = \frac{\alpha\beta^2}{\alpha\beta + \alpha\beta^2} = \frac{\beta}{1 + \beta}.$$

Edellisten lisäksi kiinnostavaa on varianssi-keskiarvo-suhteen

$$\frac{\text{Var}(\zeta_i)}{E(\zeta_i)} = \frac{\alpha\beta^2}{\alpha\beta} = \beta$$

estimointi. On huomattava, että gamma-Poisson-mallille varianssi-keskiarvosuhde ja sisäkorrelaatio ovat 1-1-suhteessa keskenään. Varianssi-keskiarvosuhde voidaan samaistaa *hajontaindeksiin* (Fisher, Thornton & Mackenzie, 1922)

$$I = \frac{\sum_{i=1}^m (n_i - \bar{n})^2}{(m-1)\bar{n}}, \quad (4.8)$$

jossa n_i on havaintojen lukumäärä koealassa i ja $\bar{n} = \frac{1}{m} \sum_{i=1}^m n_i$. Hajontaindeksin tulkinta perustuu Poisson-jakauman odotusarvon ja varianssin yhtäsuuruuteen, jolloin tunnuslukua voidaan pitää varianssisuhdetunnuslukuna, jossa osoittaja

$$s^2 = \frac{1}{m-1} \sum_{i=1}^m (n_i - \bar{n})^2$$

estimoi $N(B_{ik})$:n varianssia, kun jakaumaoletuksia ei tehdä, ja nimittäjä \bar{n} estimoi varianssia, kun Poisson-hypoteesi on voimassa. Suuret I :n arvot kertovat aggregoituneesta datasta ja pienet arvot puolestaan säännöllisestä. Homogeeniselle Poisson-prosessille $I = 1$.

Huomattakoon, että nyt estimoitu tunnusluku käyttäytyy samalla tavalla kuin hajontaindeksikin. Mitä suurempi parametrin β arvo on, sitä heterogeenisemmasta aineistosta on kyse. Parametri β kuvaa siis tässä tapauksessa aineiston heterogeenisuutta.

Tarkastellaan seuraavaksi bayesiläistä hierarkkista mallinnusta tilanteessa, jossa on m ryvästä ja jokaisessa ryvässä on r koealaa. Olkoon λ_i intensiteetti ryvässä i ja $N(B_{ik}) = n_{ik}$ pisteiden lukumäärä koealassa B_{ik} , $i = 1, \dots, m$ ja $k = 1, \dots, r$. Yksinkertaisuuden vuoksi oletetaan, että koealat ovat keskenään saman kokoisia. Tässä yhteydessä λ tarkoittaa itse asiassa intensiteettiä kerrottuna pinta-alalla, jolloin se ilmoittaa siis lukumäärää. Merkitään $\lambda = (\lambda_1, \dots, \lambda_m)$, $n = (n_{11}, \dots, n_{mr})$, $n_i = (n_{i1}, \dots, n_{ir})$ ja koealan pinta-alaa $|B|$:llä.

Nyt λ_i :n jakauma, joka tässä on siis $\text{Gamma}(\alpha, \beta^{-1})$, ajatellaan intensiteetin priorijakaumaksi $p(\lambda|\alpha, \beta)$. Koska havaintojen lukumäärät ovat Poisson-jakautuneita, on kyseessä konjugaattipriori. Gammajakauman parametrien α ja β epävarmuus otetaan huomioon asettamalla niille hyperpriori $p(\alpha, \beta)$. Hyperparametreilla α ja β ja niiden priorilla $p(\alpha, \beta)$ mallinnetaan siis intensiteettiä. Intensiteettiä koskevat päätelmät tehdään intensiteetin posteriorijakaumasta $p(\lambda, \alpha, \beta | n)$, joka saadaan hyperpriorin, priorin ja aineistosta laskettavan uskottavuuden tulona

$$\begin{aligned} p(\lambda, \alpha, \beta | n) &\propto p(\alpha, \beta) p(\lambda | \alpha, \beta) p(n | \lambda) \\ &= p(\alpha, \beta) \prod_{i=1}^m \frac{(\frac{1}{\beta})^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\frac{\lambda_i}{\beta}} \prod_{i=1}^m \prod_{k=1}^r \frac{(\lambda_i | B|)^{n_{ik}}}{n_{ik}!} e^{-\lambda_i | B|}. \end{aligned}$$

Koska päähuomio on siinä, miten intensiteetti vaihtelee alueella, meitä kiinnostavat intensiteetit ryväissä ehdolla havaitut lukumäärät $\lambda_i | n$. Estimaatit saadaan λ :n marginaaliposteriorista, λ :n prediktiivisestä jakaumasta

$$p(\lambda | n) = \int \int p(\lambda, \alpha, \beta | n) d\alpha d\beta.$$

Jos parametreille α ja β on annettu arvot tai ne on estimoitu, jolloin kyseessä on empiirinen Bayes-menetelmä, saadaan ryväiden estimaatit λ :n ehdollisesta marginaaliposteriorista ehdolla α ja β . Se on

$$\begin{aligned} p(\lambda | \alpha, \beta, n) &= \prod_{i=1}^m p(\lambda_i | \alpha, \beta, n) \\ &= \prod_{i=1}^m \frac{(\frac{1}{\beta} + r|B|)^{\alpha + \sum_{k=1}^r n_{ik}}}{\Gamma(\alpha + \sum_{k=1}^r n_{ik})} \lambda_i^{\alpha + \sum_{k=1}^r n_{ik} - 1} e^{-\lambda_i (\frac{1}{\beta} + r|B|)}, \end{aligned}$$

sillä konjugaattisuudesta johtuen λ_i :den posteriorit ovat riippumattomia $\text{Gamma}(\alpha + \sum_{k=1}^r n_{ik}, \frac{1}{\beta} + r|B|)$ -jakaumia.

Intensiteetin odotusarvoksi saadaan

$$E(\lambda_i | n_i) = \frac{\alpha + \sum_{k=1}^r n_{ik}}{\frac{1}{\beta} + r|B|}$$

ja varianssiksi

$$\text{Var}(\lambda_i | n_i) = \frac{\alpha + \sum_{k=1}^r n_{ik}}{(\frac{1}{\beta} + r|B|)^2}.$$

Parametrien α ja β riippuvuutta toisistaan voidaan tarkastella (α, β) :n marginaaliposteriorin

$$\begin{aligned} p(\alpha, \beta | n) &= \int p(\lambda, \alpha, \beta | n) d\lambda \\ &\propto p(\alpha, \beta) \prod_{i=1}^m \frac{\Gamma(\alpha + \sum_{k=1}^r n_{ik}) (\frac{1}{\beta})^\alpha |B|^{\sum_{k=1}^r n_{ik}}}{\Gamma(\alpha) (\frac{1}{\beta} + r|B|)^{\alpha + \sum_{k=1}^r n_{ik}} n_{ik}!} \end{aligned}$$

avulla.

4.4.2 Hierarkkinen mallinnus etäisyshavainnoille

Tarkastellaan etäisyyttä otantapisteestä lähimpään naapuriin d_i . Kun oletetaan, että prosessi on lokaalisti homogeeninen Poisson-prosessi, niin etäisyyksille d_i on voimassa

$$F(d_i | \lambda_i) = 1 - e^{-\lambda_i \pi d_i^2},$$

jossa λ_i on intensiteetti otantapisteessä s_i . Kun merkitään $u_i = \pi d_i^2$, saadaan tyhjän tilan u_i jakauman kertymäfunktio

$$F(u_i | \lambda_i) = 1 - e^{-\lambda_i u_i},$$

joka on eksponenttijakauman kertymäfunktio parametrina λ_i .

Intensiteetin alkeellisena estimaattorina voidaan pitää etäisyyden käänteisarvoa eli

$$\hat{\lambda}_i = \frac{1}{u_i},$$

joka on yhteen havaintoon perustuva su-estimaattori ylläolevalle mallille. Huomattakoon, että tämä on voimassa vain otantapisteessä s_i ja hyvin pienellä alueella sen ympärillä. Lisäksi tällainen estimaattori on erityisen herkkä poikkeamalle Poisson-oletuksesta.

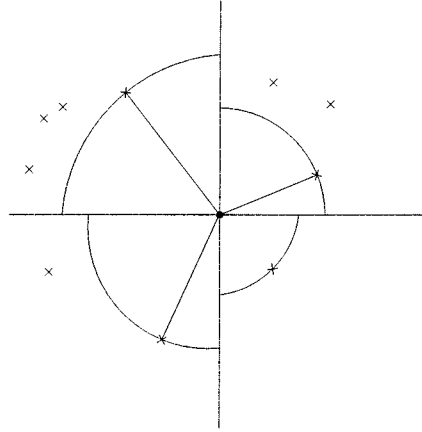
Jotta hierarkkinen mallinnus olisi luontevaa myös etäisyshavainnoinin yhteydessä, tarvitaan tutkittavista otantapisteistä toistoja. Oletetaan seuraava yksinkertainen asetelma: ryvästä vastaa otantapisteen ympäröivän alueen ositus neljään 90 asteen sektoriin. Etäisyys otantapisteestä lähimpään

naapuriin mitataan jokaisessa sektorissa erikseen, jolloin saadaan neljä mitausta. Otanta-asetelma on esitetty kuvassa 4.3. Poisson-oletuksesta seuraa, että etäisyydet d_{ik} , $k = 1, 2, 3, 4$, ovat tällöin stokastisesti riippumattomia ehdolla, että ne kuuluvat samaan otantapisteeseen i . Nyt etäisyyden d_{ik} jakauman kertymäfunktio on

$$F(d_{ik}|\lambda_i) = 1 - e^{-\frac{1}{4}\lambda_i\pi d_{ik}^2}$$

ja tyhjän tilan u_{ik} jakauman kertymäfunktio on

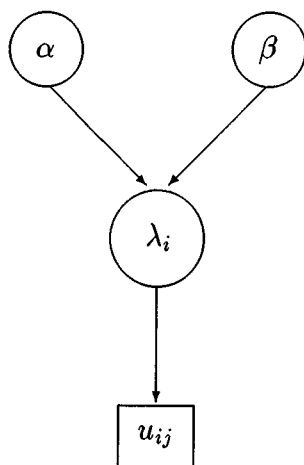
$$F(u_i|\lambda_i) = 1 - e^{-\frac{1}{4}\lambda_i u_i}.$$



Kuva 4.3 Eräs otanta-asetelma etäisyshavainnoille: ● = satunnainen otantapiste ja × = prosessin piste.

Hierarkkinen rakenne on vastaava kuin lukumäärähavainnoinninkin yhteydessä. Nyt vain lukumääriä vastaa tyhjä tila u_i , joka noudattaa eksponenttijakaumaa parametrilla λ_i . Intensiteetin λ_i jakaumaksi oletetaan $\text{Gamma}(\alpha, \beta^{-1})$. Nyt havainnot ovat siis ehdollisesti riippumattomia ehdolla, että ne on mitattu samasta otantapisteesestä. Hierarkkinen rakenne on esitetty kuvassa 4.4.

Bayesiläinen hierarkkinen mallinnus etäisyshavainnoille menee vastaavasti kuin lukumäärähavainnoillekin. Olkoon otantapisteiden määrä m ja λ_i



Kuva 4.4 Hierarkkinen rakenne etäisyshavainnoille. Tyhjä tila u_{ij} riippuu tuntemattomasta intensiteetistä λ_i , joka riippuu puolestaan gammajakauman parametreista α ja β .

pinta-alalla skaalattu intensiteetti otantapisteessä s_i . Olkoon otantapisteessä s_i tehdyt havainnot $u_i = (u_{i1}, \dots, u_{i4})$. Merkitään $u = (u_{11}, \dots, u_{m4})$ ja $\lambda = (\lambda_1, \dots, \lambda_m)$.

Yhteisposteriori intensiteetille ja hyperparametreille on

$$\begin{aligned}
 p(\lambda, \alpha, \beta | u) &\propto p(\alpha, \beta) p(\lambda | \alpha, \beta) p(u | \lambda) \\
 &= p(\alpha, \beta) \prod_{i=1}^m \frac{(\frac{1}{\beta})^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\frac{\lambda_i}{\beta}} \prod_{i=1}^m \prod_{j=1}^4 \frac{1}{4} \lambda_i e^{-\frac{1}{4} \lambda_i u_{ij}} \\
 &= p(\alpha, \beta) \prod_{i=1}^m \prod_{k=1}^4 \frac{(\frac{1}{\beta})^\alpha}{4\Gamma(\alpha)} \lambda_i^\alpha e^{-\lambda_i (\frac{1}{\beta} + \frac{1}{4} u_{ik})}.
 \end{aligned}$$

Marginaaliposteriori λ :lle ehdolla α ja β

$$\begin{aligned}
 p(\lambda | \alpha, \beta, u) &= \prod_{i=1}^m p(\lambda_i | \alpha, \beta, u) \\
 &= \prod_{i=1}^m \frac{(\frac{1}{\beta} + \sum_{k=1}^4 u_{ik})^{\alpha+1}}{\Gamma(\alpha+1)} \lambda_i^\alpha e^{-\lambda_i (\frac{1}{\beta} + \sum_{k=1}^4 u_{ik})}
 \end{aligned}$$

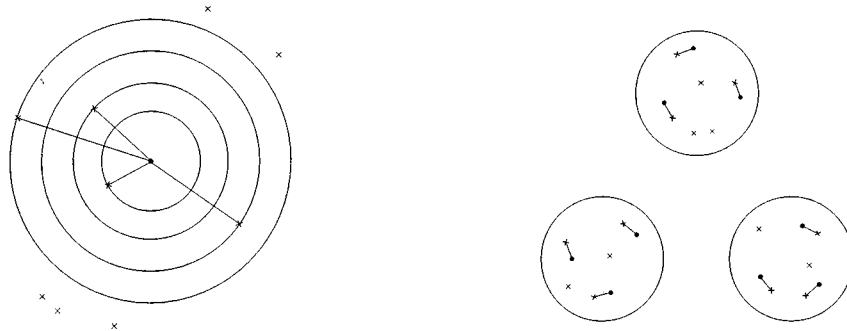
sillä λ_i :den posteriorit ovat riippumattomia $\text{Gamma}(\alpha + 1, \frac{1}{\beta} + \sum_{k=1}^4 u_{ik})$ -jakaumia (konjugaattisuus).

Parametrien α ja β marginaaliposteriori on

$$p(\alpha, \beta | u) = \frac{p(\lambda, \alpha, \beta | u)}{p(\lambda | \alpha, \beta, u)}$$

$$\propto p(\alpha, \beta) \prod_{i=1}^m \frac{\Gamma(\alpha + 1) (\frac{1}{\beta})^\alpha}{\Gamma(\alpha) (\frac{1}{\beta} + \sum_{k=1}^4 u_{ij})^{\alpha+1}}.$$

Muita vaihtoehtoisia otanta-asetelmia etäisyysmittauksille on esitetty kuvassa 4.5. Vasemmanpuoleisessa otanta-asetelmassa on mitattu etäisyydet satunnaisesta otantapisteestä ensimmäiseksi, toiseksi, kolmanneksi ja neljänneksi lähimpään prosessin pisteeseen. Kiinnostus on ympyröiden, joiden keskipisteenä on otantapiste ja säteinä mitatut etäisyydet, pinta-aloissa. Poisson-prosessin tapauksessa ympyröiden muodostamien renkaiden pinta-alat ovat toisistaan riippumattomia. Oikeanpuoleinen kuva esittää yhtä otantapisteryvystä, jossa jokaisessa koealassa on mitattu etäisyydet kolmesta satunnaisesta otantapisteestä lähimpään prosessin pisteeseen. Tällainen asetelma on käyttökelpoinen hitaasti muuttuvalle tiheydelle.



Kuva 4.5 Otanta-asetelmia etäisyyshavainnoille: • = satunnainen otantapiste ja × = prosessin piste. Vasemmalla on otantapisteympyrät ja oikealla otantapisteryväs.

5 Karttamuotoisen aineiston intensiteetin estimointi

Olkoon $\{x_1, \dots, x_n\}$ A :ssa havaittava Cox-prosessi, jonka satunnainen intensiteettifunktio on $\Lambda(s)$. Ehdolla $\Lambda(s) = \lambda(s)$ prosessi on epähomogeeninen Poisson-prosessi ja sen intensiteettifunktio on $\lambda(s)$. Ehdollinen uskottavuusfunktio on tyypillisesti muotoa

$$\exp \left\{ \int_A \lambda(s) ds \right\} \prod_{i=1}^N \lambda(x_i). \quad (5.1)$$

Intensiteetin mallinnuksessa kiinnostus on satunnaisessa intensiteettifunktiossa ehdolla havaitut pisteet $\{x_1, \dots, x_n\}$.

Tunnettuja lähestymistapoja ovat epäparametrinen (kernel-)menetelmä (Diggle, 1985), parametrinen suurimman uskottavuuden menetelmä (Ogata & Tanemura, 1986), hierarkkinen bayesiläinen malli (Møller ym., 1998), epäparametrinen bayesiläinen malli (Heikkinen & Arjas, 1998; Ickstadt & Wolpert, 1996) ja tyypin II suurimman uskottavuuden menetelmä (Ogata, 1996).

Tässä luvussa tarkastellaan kernel-menetelmää ja bayesiläistä lähestymistapaa.

5.1 Intensiteetin klassinen estimointi

Tarkastellaan aluksi intensiteetin estimointia kernel-menetelmällä. Menetelmä on epäparametrinen, ja se on rinnastettavissa tiheysfunktion estimointiin.

Itse asiassa epähomogeenisen Poisson-prosessin intensiteetin estimointi voidaan tulkita kaksiulotteisen jatkuvan jakauman tiheysfunktion estimoinniksi sillä erolla, että $\int_A \lambda(s) ds$ ei ole tunnettu. Tiheysfunktioille se olisi yksi.

Tiheyden kernel-estimaattori on

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K \left\{ \frac{1}{h}(x - X_i) \right\}, \quad (5.2)$$

jossa K on kernel-funktio ja h ikkunaleveys (tasoittava parametri). Kernel-funktio K vaikuttaa estimoidun tiheysjakauman muotoon tasoittamalla sitä, ja ikkunaleveydellä määrätään tasoituksen aste. Tämän vuoksi estimoitu intensiteettipinta on korreloitunut, jolloin intensiteetti muuttuu tasaisesti tutkittavalla alueella.

Kernel-funktio on yleensä symmetrinen yksihuippuinen todennäköisyysjakauma, jolle on voimassa

$$\int_{R^2} K(x) dx = 1.$$

Yleisesti käytettyjä kernel-funktioita ovat esimerkiksi Gaussin kernel-funktio

$$K(x) = (2\pi)^{-1} \exp\left(-\frac{1}{2}x^T x\right)$$

ja Epanechnikovin kernel-funktio

$$K_e(x) = \begin{cases} 2\pi^{-1}(1 - x^T x), & \text{kun } x^T x < 1 \\ 0 & \text{muualla.} \end{cases}$$

Kernel-estimoinnissa ongelmaksi muodostuu yleensä ikkunaleveyden h valinta. Mitä suurempi h on sitä tasaisempi tiheysfunktioista saadaan. Yleensä tiheysfunktion käyttötarkoitus määrää ikkunaleveyden ja monesti se voidaan määrätä silmämääräisesti. Objektivisia ikkunaleveyden valintaan liittyviä menetelmiä kuten cross-validation on esitetty mm. kirjassa Silverman (1986).

5.2 Intensiteetin bayesiläinen estimointi

Bayesiläisen ajattelutavan mukaan intensiteettiä kuvaava satunnaiskenttä $\{\lambda(x), x \in R^2\}$ ajatellaan intensiteetin priorijakaumaksi. Satunnaiskenttä voi olla joko jatkuvaparametrinen tai diskreetti tilanteesta riippuen. Satunnaiskentän konstruoinnissa tunnettuja lähestymistapoja ovat log-gaussinen Cox-prosessi (Møller ym., 1998), kernel-tasoitettu gamma-Poisson-prosessi

(Ickstadt & Wolpert, 1996) ja Heikkinen–Arjas-mallit (Heikkinen & Arjas, 1998). Bayes-menetelmässä intensiteettiä kuvataan satunnaiskentän posteriorijakaumalla $p(\lambda|x)$, joka yleensä lasketaan simulointia hyväksi käyttäen.

Sovelletaan Ickstadtin ja Wolpertin gamma-Poisson-mallia hieman yksinkertaistettuna. Käsitellään sekä empiiristä että täysin Bayesilaista menetelmää. Menetelmien ero näkyy niiden suhtautumisessa priorijakauman parametreihin: empiirisessä Bayes-menetelmässä gammajakauman parametrien vaihtelua ei oteta huomioon, kun taas täysin Bayes-menetelmässä se huomioidaan.

Käsitellään ensin yksinkertaisinta tapausta, jossa satunnaiskenttä $\{\lambda(x)\}$ on korreloimaton. Tarkasteltava tilanne on hyvin lähellä kenttähavainnoinnin yhteydessä luvussa 4.4 tehtyä Bayes-mallinnusta.

5.2.1 Yksinkertainen Bayes-malli

Tarkastellaan aluetta A rasterina, jolloin A :n osajoukot A_i ovat keskenään samanmuotoisia neliöitä, kvadraatteja. Kaksinkertaisten indeksien sijasta käytetään merkintöjä yksinkertaistamaan vain yhtä indeksiä i viittaamaan i :nteen ruutuun. Priorina on diskreetti satunnaiskenttä, joka muodostetaan riippumattomista prosesseista, jonka realisaatiot ovat luvussa 3.5.3 esitettyjä rasteriprosesseja. Ts. jokaiseen kvadraattiin A_i generoidaan gammajakautunut intensiteetti λ_i toisistaan riippumatta, jonka jälkeen kunkin kvadraatin objektien lukumäärä n_i saadaan Poisson-jakaumasta parametrina $\lambda_i|A_i|$. Mallina on siis yksinkertainen gamma-Poisson-prosessi:

$$\begin{aligned}\lambda_i &\sim \text{Gamma}(\alpha, \beta^{-1}) \\ n_i|\lambda_i &\sim \text{Poisson}(\lambda_i|A_i|).\end{aligned}$$

Tarkastellaan ensin empiiristä Bayes-menetelmää, jossa gammajakauman parametrit α ja β estimoidaan. Estimoinnissa käytetään momenttimenetelmää, jolloin estimaatit saadaan yhtälöparista

$$\begin{aligned}E(\lambda_i|\alpha, \beta) &= \alpha\beta \\ \text{Var}(\lambda_i|\alpha, \beta) &= \alpha\beta^2,\end{aligned}$$

jossa $E(\lambda_i|\alpha, \beta)$ ja $\text{Var}(\lambda_i|\alpha, \beta)$ ovat aineistosta laskettujen raakaestimaattien keskiarvo ja varianssi. Parametrien estimaatit ovat täten

$$\hat{\alpha} = \frac{[E(\lambda)]^2}{\text{Var}(\lambda)} \quad \text{ja} \quad \hat{\beta} = \frac{E(\lambda)}{\text{Var}(\lambda)}.$$

Intensiteetin priorina on siis yksinkertaisen gamma-Poisson-prosessin avulla muodostettu satunnaiskenttä. Satunnaiskenttä on korreloimaton, sillä intensiteetit kvadraateissa ovat stokastisesti riippumattomia. Intensiteetin priorijakauma on täten

$$p(\lambda_i) = \frac{\left(\frac{1}{\beta}\right)^{\hat{\alpha}}}{\Gamma(\hat{\alpha})} \lambda_i^{\hat{\alpha}-1} e^{-\frac{\lambda_i}{\beta}}$$

ja yhteen havaintoon perustuva uskottavuus on Poisson-mallista johtuen

$$p(n_i|\lambda_i) = \frac{(|A_i|\lambda_i)^{n_i}}{n_i!} e^{-|A_i|\lambda_i}.$$

Siten intensiteetin posteriorijakauma on

$$\begin{aligned} p(\lambda_i|n_i) &\propto p(\lambda_i)p(n_i|\lambda_i) \\ &= \frac{\left(\frac{1}{\beta}\right)^{\hat{\alpha}}}{\Gamma(\hat{\alpha})} \lambda_i^{\hat{\alpha}-1} e^{-\frac{\lambda_i}{\beta}} \frac{(|A_i|\lambda_i)^{n_i}}{n_i!} e^{-|A_i|\lambda_i} \\ &\propto \lambda_i^{(\hat{\alpha}+n_i)-1} e^{-\lambda_i\left(\frac{1}{\beta}+|A_i|\right)}. \end{aligned}$$

Koska gammajakauma on Poisson-jakauman konjugaatti, on posteriori myös gammajakautunut. Ts. $p(\lambda_i|n_i) \propto \text{Gamma}(\hat{\alpha} + n_i, \frac{1}{\beta} + |A_i|)$. Täten intensiteetin Bayes-estimaatiksi kvadraatissa i saadaan

$$\hat{\lambda}_i = E(\lambda_i|n_i) = \frac{\hat{\alpha} + n_i}{\frac{1}{\beta} + |A_i|}. \quad (5.3)$$

Empiiriset Bayes-estimaatit $\hat{\lambda}_i$ voitaisiin laskea myös iteroimalla siten, että ensin lasketaan estimaatit $\hat{\alpha}$ ja $\hat{\beta}$ kuten edellä. Tämän jälkeen lasketaan kaavan (5.3) mukaisesti estimaatit $\hat{\lambda}_i$, joiden avulla lasketaan uudet $\hat{\alpha}$ ja $\hat{\beta}$ jne. Iterointia jatketaan niin kauan kunnes $\hat{\alpha}$ ja $\hat{\beta}$ eivät enää muutu.

Täysin bayesiläisessä lähestymistavassa gammajakauman parametrien α ja β vaihtelu otetaan huomioon asettamalla niille hyperpriori $p(\alpha, \beta)$. Tällöin posteriorijakauma on suoraan verrannollinen hyperpriorin, priorin ja aineistosta lasketun uskottavuuden tuloon. Posteriorin laskeminen johtaa nyt kuitenkin Markov Chain Monte Carlo -menetelmään, jota on käsitelty ”kontekstuaalisuusmallin” yhteydessä kappaleessa 5.2.3. Täysin bayesiläisen lähestymistavan mukainen mallinnus tässä kappaleessa esitellylle mallille menee vastaavasti, sillä se on itse asiassa erikoistapaus kontekstuaalisuusmallista.

Koska triviaali Bayes-menetelmä perustuu kvadraattien riippumattomuuteen, on intensiteettien estimointi helppoa. Menetelmä on perusteltavissa kenttähavainnoinnin yhteydessä, jossa rypäät ovat (lähes) riippumattomia. Nyt menetelmä on kuitenkin puutteellinen, sillä se ei huomioi ilmiölle luonteenomaista kontekstuaalisuutta, ts. sitä, että tiheys muuttuu tasaisesti. Koska estimaatit $\hat{\lambda}_i$ ovat riippumattomia, ei niiden välillä ole spatiaalista korrelaatiota. Tämän vuoksi vierekkäisissä kvadraateissa intensiteetit voivat vaihdella suuresti, mikä voi olla ilmiölle hyvin epätodennäköistä.

5.2.2 Yksinkertainen Bayes-malli kontekstuaalisuudelle

Edellä tarkastellun menetelmän puuttena todettiin olevan riippumattomuus vierekkäisten kvadraattien intensiteettien välillä. Tämä puute voidaan poistaa usealla tavalla. Tarkastellaan seuraavaksi kernel-tasoituksen käyttöä.

Oleellinen ero nyt esiteltävän mallin ja gamma-Poisson-mallin välillä on priorina olevaan riippumattomaan gamma-Poisson-prosessiin tehtävä tasointus. Priorin konstruoinnissa lähdetään liikkeelle riippumattomista prosesseista kuten triviaalissakin tapauksessa. Ts. jokaiseen kvadraattiin generoidaan gammajakautunut impulssi γ_i toisistaan riippumatta. Tämän jälkeen tehdään kernel-tasointus, jossa impulsseille lasketaan vierekkäisten kvadraattien kanssa painotettuja keskiarvoja. Satunnaiskenttä $\{\lambda(x)\}$ muodostuu siten painotetuista keskiarvoista λ_i .

Mallina on siis kernel-tasoitettu gamma-Poisson-prosessi (ks. kappale 3.5.3):

$$\begin{aligned}\gamma_j &\sim \text{Gamma}(\alpha, \beta^{-1}) \\ \lambda_i &= \sum_{j \in J} k_{ij} \gamma_j \\ \{n_i\}_{i \in I} | \{\lambda_j\}_{j \in J} &\sim \text{Poisson}(\lambda_i | A_i|),\end{aligned}$$

jossa γ_i on impulssi ja λ_i intensiteetti kvadraatissa i . Nyt gammajakautuksen parametreihin α ja β liittyvä epävarmuus on luontevinta ottaa huomioon asettamalla niille hyperpriori $p(\alpha, \beta)$, mikä johtaa täysin bayesiläiseen menetelmään. Parametri k_{ij} oletetaan tunnetuksi. Ickstadt & Wolpert (1996) mallintavat myös k_{ij} :t ja asettavat niihin liittyville parametreille hyperpriorit.

Olkoon kvadraattien lukumäärä N ja käsitellään rasteria vektorina (luetaan riveittäin ylhäältä alas). Havaintojen lukumäärävektori on täten $n = (n_1, \dots, n_N)$, impulssivektori $\gamma = (\gamma_1, \dots, \gamma_N)$ ja intensiteettivektori $\lambda = (\lambda_1, \dots, \lambda_N)$. Posteriori on

$$p(\alpha, \beta, \gamma | n) \propto p(\alpha, \beta) p(\gamma | \alpha, \beta) p(n | \gamma), \quad (5.4)$$

jonka laskeminen on analyttisesti mahdoton tehtävä. Ongelman ratkaisemisessa joudutaan käyttämään *Markov Chain Monte Carlo* (lyh. MCMC) -menetelmää. Menetelmän ideana on simuloida toisistaan riippuvia realisatioita käyttäen sellaista ergodista Markovin ketjua, jonka tasapainojakaumana on $p(\alpha, \beta, \gamma | n)$. Tällaisen Markovin ketjun konstruointi voidaan tehdä usealla tavalla. Tässä käytetään *Metropolis–Hastings*-algoritmia.

Metropolis–Hastings-algoritmin periaate on seuraava: Olkoon satunnaismuuttuja X , jonka jakauma on $\pi(\cdot)$. Ketjun tila hetkellä t on x_t . Seuraava tila x_{t+1} saadaan siten, että ensin generoidaan arvo x^* ehdotusjakaumasta $q(\cdot | x_t)$, joka voi riippua ketjun sen hetkisestä tilasta x_t . Uusi arvo x^* hyväksytään todennäköisyydellä $\alpha(x_t, x^*)$, jossa

$$\alpha(x_t, x^*) = \min \left(1, \frac{\pi(x^*)q(x|x^*)}{\pi(x)q(x^*|x)} \right).$$

Jos ehdotettu arvo hyväksytään, uudeksi tilaksi tulee $x_{t+1} = x^*$. Jos ehdotettu arvo hylätään, ketju ei liiku, ts. $x_{t+1} = x_t$. Jos ehdotusjakauma on symmetrinen, ts. siirtymistodennäköisyyksille on $q(x|x^*) = q(x^*|x)$, hyväksymistodennäköisyys on

$$\alpha(x_t, x^*) = \min \left(1, \frac{\pi(x^*)}{\pi(x)} \right).$$

Seuraavaksi on konstruoitu Metropolis–Hastings-algoritmi posteriorin 5.4 simuloinnille. Parametreja on yhteensä $N + 2$ kappaletta, joita päivitetään vuorotellen. Ajatellaan parametrit α ja β riippumattomiksi, jolloin $p(\alpha, \beta) = p(\alpha)p(\beta)$, ts. molemmille parametreille voidaan määrätä oma hyperpriori. Algoritmi on seuraava:

1. Annetaan parametreille alkuarvot $\alpha_0, \beta_0, \gamma_{10}, \dots, \gamma_{N0}$.
2. Päivitetään α
 - generoidaan uusi arvo α^* ehdotusjakaumasta $q(\cdot | \alpha)$
 - lasketaan Metropolis–Hastings-suhde

$$\begin{aligned} P^* &= \frac{p(\alpha^* | \beta, \gamma, n)q(\alpha | \alpha^*)}{p(\alpha | \beta, \gamma, n)q(\alpha^* | \alpha)} = \frac{p(\alpha^*, \beta, \gamma | n)q(\alpha | \alpha^*)}{p(\alpha, \beta, \gamma | n)q(\alpha^* | \alpha)} \\ &= \frac{p(\alpha^*)p(\gamma | \alpha^*, \beta)q(\alpha | \alpha^*)}{p(\alpha)p(\gamma | \alpha, \beta)q(\alpha^* | \alpha)} = \frac{p(\alpha^*)p(\gamma | \alpha^*, \beta)q(\alpha | \alpha^*)}{p(\alpha)p(\gamma | \alpha, \beta)q(\alpha^* | \alpha)} \end{aligned}$$

$$\begin{aligned}
&= \frac{p(\alpha^*)q(\alpha|\alpha^*)}{p(\alpha)q(\alpha^*|\alpha)} \prod_{i=1}^N \frac{\Gamma(\alpha^*)}{\Gamma(\alpha)} \beta^{-(\alpha^*-\alpha)} \gamma_i^{\alpha^*-\alpha} \\
&= \frac{p(\alpha^*)q(\alpha|\alpha^*)}{p(\alpha)q(\alpha^*|\alpha)} \left(\frac{\Gamma(\alpha)}{\Gamma(\alpha^*)} \right)^N \beta^{-N(\alpha^*-\alpha)} e^{(\alpha^*-\alpha) \sum_{i=1}^N \log \gamma_i}
\end{aligned}$$

- $\alpha_{t+1} = \alpha^*$ todennäköisyydellä $\min(1, P^*)$
muuten $\alpha_{t+1} = \alpha_t$

3. Päivitetään β

- generoidaan uusi arvo β^* ehdotusjakaumasta $q(\cdot|\beta)$
- lasketaan Metropolis–Hastings-suhde

$$\begin{aligned}
P^* &= \frac{p(\beta^*|\alpha, \gamma, n)q(\beta|\beta^*)}{p(\beta|\alpha, \gamma, n)q(\beta^*|\beta)} = \frac{p(\beta^*)p(\gamma|\alpha, \beta^*)q(\beta|\beta^*)}{p(\beta)p(\gamma|\alpha, \beta)q(\beta^*|\beta)} \\
&= \frac{p(\beta^*)q(\beta|\beta^*)}{p(\beta)q(\beta^*|\beta)} \prod_{i=1}^N \left(\frac{\beta}{\beta^*} \right)^\alpha e^{-\gamma_i \left(\frac{1}{\beta^*} - \frac{1}{\beta} \right)} \\
&= \frac{p(\beta^*)q(\beta|\beta^*)}{p(\beta)q(\beta^*|\beta)} \left(\frac{\beta}{\beta^*} \right)^{N\alpha} e^{-\left(\frac{1}{\beta^*} - \frac{1}{\beta} \right) \sum_{i=1}^N \gamma_i}
\end{aligned}$$

- $\beta_{t+1} = \beta^*$ todennäköisyydellä $\min(1, P^*)$
muuten $\beta_{t+1} = \beta_t$

4. Päivitetään impulssi $\gamma_i, i = 1, \dots, N$

- generoidaan uusi arvo γ_i^* ehdotusjakaumasta $q(\cdot|\gamma_i)$
- lasketaan Metropolis–Hastings-suhde
merkitään $\gamma^* = (\gamma_1, \dots, \gamma_i^*, \dots, \gamma_N), \gamma_{-i} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_N)$
ja a on kvadraatin pinta-ala

$$\begin{aligned}
P^* &= \frac{p(\gamma_i^*|\gamma_{-i}, \alpha, \beta, n)q(\gamma|\gamma^*)}{p(\gamma_i|\gamma_{-i}, \alpha, \beta, n)q(\gamma^*|\gamma)} = \frac{p(\gamma^*, \alpha, \beta|n)q(\gamma|\gamma^*)}{p(\gamma, \alpha, \beta|n)q(\gamma^*|\gamma)} \\
&= \frac{p(\alpha)p(\beta)q(\gamma|\gamma^*) \prod_{k \neq i} p(\gamma_k|\alpha, \beta)p(\gamma_i^*|\alpha, \beta)p(n_i|\gamma_i^*)}{p(\alpha)p(\beta)q(\gamma^*|\gamma) \prod_{k \neq i} p(\gamma_k|\alpha, \beta)p(\gamma_i|\alpha, \beta)p(n_i|\gamma_i)} \\
&= \frac{q(\gamma|\gamma^*)p(\gamma_i^*|\alpha, \beta)}{q(\gamma^*|\gamma)p(\gamma_i|\alpha, \beta)} \prod_{j \sim i} \frac{\frac{(a\lambda_j^*)^{n_j}}{n_j!} e^{-a\lambda_j^*}}{\frac{(a\lambda_j)^{n_j}}{n_j!} e^{-a\lambda_j}} \\
&= \frac{q(\gamma|\gamma^*)}{q(\gamma^*|\gamma)} \left(\frac{\gamma_i^*}{\gamma_i} \right)^{\alpha-1} e^{-\frac{1}{\beta}(\gamma_i^*-\gamma_i)} \prod_{j \sim i} \left(\frac{\lambda_j^*}{\lambda_j} \right)^{n_j} e^{-a(\lambda_j^*-\lambda_j)}
\end{aligned}$$

- $\gamma_{i,t+1} = \gamma_i^*$ todennäköisyydellä $\min(1, P^*)$
muuten $\gamma_{i,t+1} = \gamma_{i,t}$

5. Kasvatetaan $t = t + 1$ ja palataan kohtaan 2.

6 Tutkimusaineistoon liittyviä empiirisiä tarkasteluja

Tarkastellaan luvussa 2 esiteltyjen metsien, Vuorijärven ja Susimäen, spatiaalista rakennetta sekä intensiteetin estimointia. Puiden sijainnit on nähtävissä kuvasta 2.1. Analysoinnissa on käytetty Splus-ohjelmistoa ja erityisesti sen SpatialStats-modulia.

6.1 Tilajärjestyksen tarkastelu

Metsien tilajärjestyksiä on tarkasteltu G -, F - ja K -funktioiden avulla. G -funktio tarkastelee lähinaapurietäisyyttä ja F -funktio etäisyyttä otantapistteestä lähimpään prosessin pisteeseen. K -funktio voidaan tulkita siten, että $\lambda K(t)$ kertoo t -säteisessä ympyrässä olevien muiden pisteiden lukumäärän odotusarvon, kun ympyrän keskipiste on satunnaisesti valitussa prosessin pisteessä. Tarkasteluissa on ollut mukana kaikki puulajit, sillä tilajärjestyksen tarkastelu puulajeittain ei tässä yhteydessä ole järkevää, koska muita puita kuin kuusia on vain vähän.

Kuvassa 6.1 on vertailtu metsien G -funktioita. Kuvaajista nähdään, että molemmassa metsässä on hyvin lähekkäin sijaitsevia puita. Vuorijärven G -funktioista on havaittavissa klusteroitumista etäisyyden ollessa alle metrin, mikä johtunee oikean yläkulman tiheästä koivutaimikosta. Jos Vuorijärven metsästä tarkasteltaisiin vain täysikasvuisia puita, jäisivät lyhyet (alle metri)

etäisyydet pois, jolloin G -funktioista olisi havaittavissa selvää säännöllisyyttä. Susimäen G -funktioista ei ole havaittavissa klusteroitumista eikä säännöllisyyttäkään. Samat johtopäätökset voidaan tehdä metsien F -funktioista, jotka on esitetty kuvassa 6.2.

Tilajärjestyksen satunnaisuuden testauksessa on käytetty G -funktiota (Diggle, 1983, s. 16). Tulokset on esitetty kuvassa 6.3. Jos kuvaaja kulkee simuloitujen virherajojen sisäpuolella, voidaan tulkita puiden sijoittuneen tasaisesti ja toisistaan riippumatta, jolloin Poisson-hypoteesi on voimassa. Vuorijärven puiden sijainteja ei voida pitää satunnaisena, sillä lyhyillä etäisyyksillä puita on enemmän kuin täysin satunnaisessa tapauksessa ja pidemmillä etäisyyksillä liian vähän. Susimäen metsässä puiden sijainteja voidaan pitää lähes satunnaisina.

Metsien K -funktiot on laskettu käyttäen Rippleyn estimaattoria (Ripley, 1976). Näitä on verrattu kuvassa 6.4 vastaavaan homogeenisen Poisson-prosessin K -funktioon. Jälleen huomataan, että Susimäen metsä on hyvin lähellä homogeenisen Poisson-prosessin tilannetta.

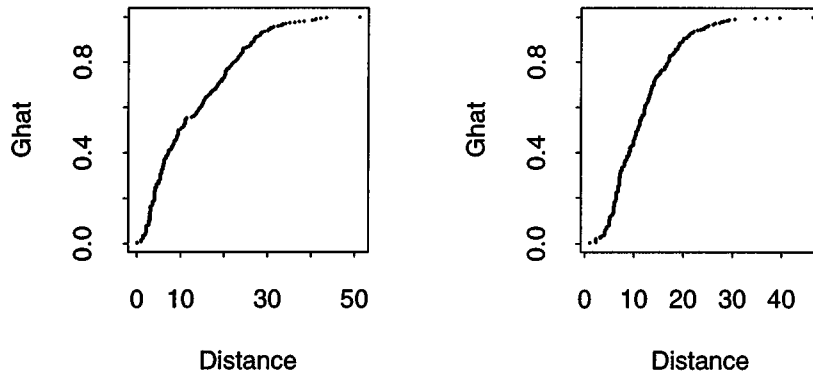
Edellä tehdyt tarkastelut tukevat sitä johtopäätöstä, mikä voidaan tehdä silmämääräisesti jo kuvasta 2.1: Vuorijärven metsässä puut sijaitsevat nuorta koivikkoa lukuun ottamatta melko säännöllisesti, kun taas Susimäen metsässä puut ovat sijoittuneet satunnaisesti.

6.2 Trendin tarkastelu

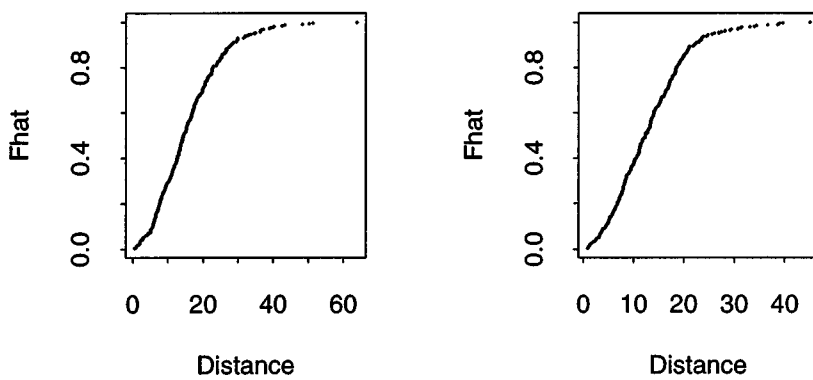
Molemmat metsät on diskretisoitu 16×16 -ruudukoksi ja jokaisen ruudun keskipisteeseen (x_i, y_i) on liitetty tiheys z_i kyseisessä ruudussa. Tarkastellaan tiheyksissä z_i esiintyviä muutoksia, ts. tutkitaan esiintyykö puutiheydessä jonkin suuntaista trendiä.

Trendin poistamiseen on käytetty iteratiivista median polish -menetelmää (Cressie, 1991, s. 46), joka perustuu sekä rivi- että sarakesuunnassa laskettuihin tiheyden mediaaneihin. Trendipinnat on esitetty kuvassa 6.5, jossa ovat myös alkuperäiset aineistot tulkintaa helpottamassa. Kummassakaan metsässä ei ole havaittavissa selvää trendiä missään suunnassa. Trendipinnat eivät ole tässä tapauksessa kovinkaan informatiivisia.

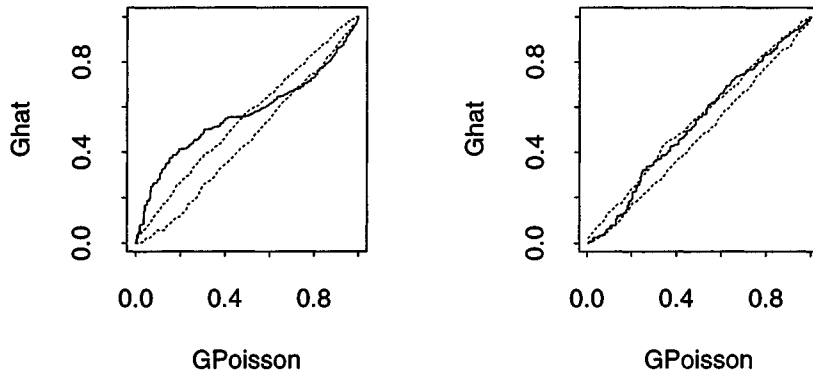
Trendin poistamisen vaikutuksia on tutkittu median polish -jäännösten variogrammeista vertaamalla niitä vastaaviin alkuperäisen aineiston variogrammeihin. Liitteessä 2 on esitetty variogrammit sekä itä-länsi-suunnassa että pohjois-etelä-suunnassa alkuperäiselle aineistolle ja median polish -jäännöksille. Ainoastaan Vuorijärven pohjois-etelä-variogrammista on havaittavissa vähäistä korrelaatiota (variogrammi kasvaa), mutta sitäkään ei voida selittää trendillä, sillä vastaava jäännösten variogrammi on myös kasvava (ei



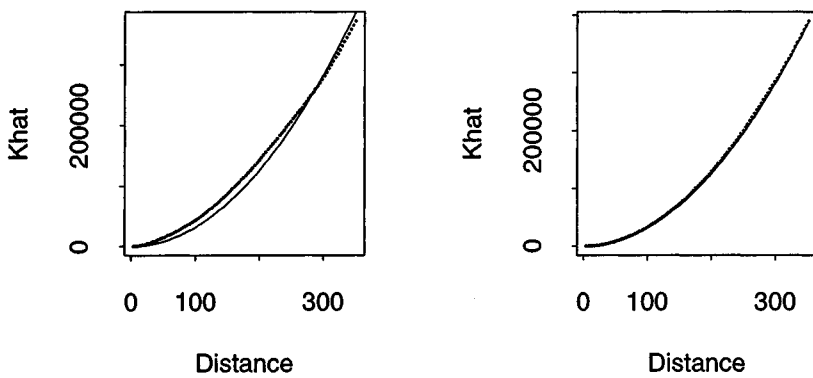
Kuva 6.1 Vasemmalla Vuorijärven ja oikealla Susimäen G -funktio. Vuorijärven G -funktioista on havaittavissa pienillä etäisyyksillä klusteroitumista ja suurilla etäisyyksillä säännöllisyyttä, kun taas Susimäen G -funktioista ei ole havaittavissa kumpaakaan.



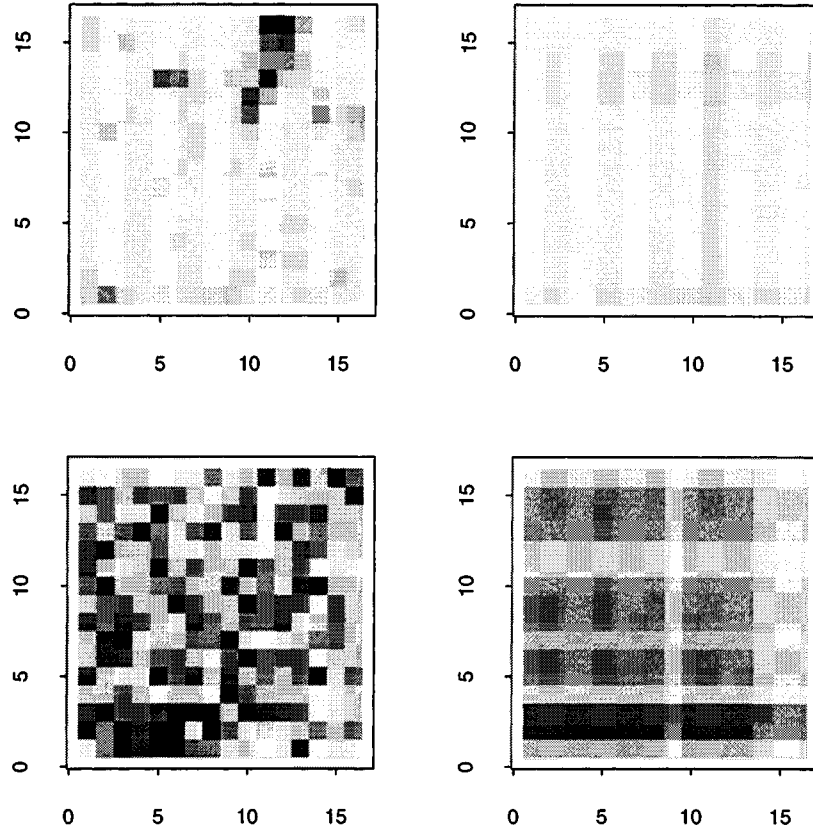
Kuva 6.2 Vasemmalla Vuorijärven ja oikealla Susimäen F -funktio.



Kuva 6.3 Vasemmalla Vuorijärven ja oikealla Susimäen satunnaisuuden testaus G -funktion avulla. Yhtenäinen viiva kuvaa \hat{G} -funktiota ja pisteiviivojen välinen alue \hat{G} -funktion pisteittäistä 90 %:n luottamusväliä.



Kuva 6.4 Vasemmalla Vuorijärven ja oikealla Susimäen estimoitu K -funktio (pisteiviiva). Yhtenäinen viiva esittää homogeenisen Poisson-prosessin K -funktiota.



Kuva 6.5 Median polish -menetelmällä saadut trendipinnat Vuorijärven (ylhäällä) ja Susimäen (alhaalla) metsille. Vasemmalla alkuperäinen aineisto ja oikealla trendi.

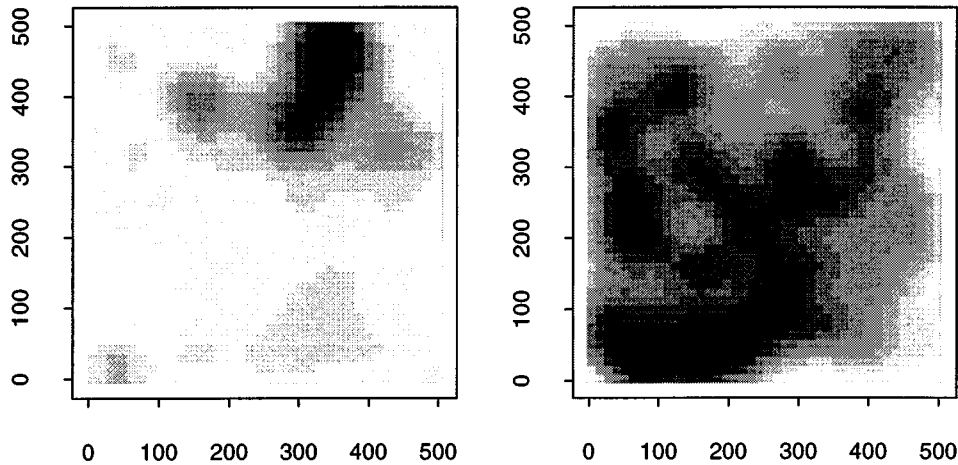
ole vakio etäisyyden suhteen).

6.3 Intensiteetin estimointi

6.3.1 Klassinen estimointi

Sovelletaan ensin luvussa 5.1 esitettyä kernel-menetelmää metsien intensiteettien estimointiin. Kernel-funktiona on käytetty Gaussin kernel-funktiota ja ikkunaleveytenä 3 metriä. Ikkunaleveys on valittu silmämääräisesti. Estimoidut intensiteettipinnat on esitetty kuvassa 6.6 harmaasävykuvina. Tum-

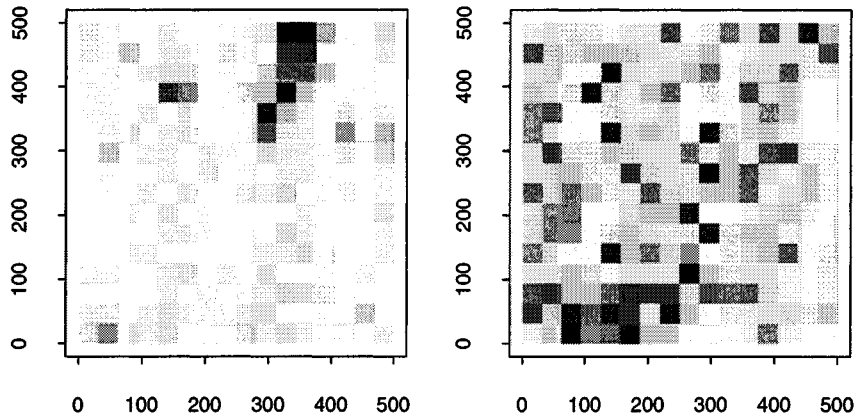
ma sävy kuvaa suurta intensiteettiä ja vaalea pientä. Pinnan korreloituneisuudesta johtuen intensiteetti muuttuu tasaisesti molemmissa kuvissa.



Kuva 6.6 Vasemmalla Vuorijärven ja oikealla Susimäen kernel-estimoidut intensiteettipinnat. Tumma sävy kuvaa suurta intensiteettiä ja vaalea pientä.

6.3.2 Bayesiläinen estimointi

Tarkastellaan ensin intensiteetin estimointia empiirisellä Bayes-menetelmällä käyttäen luvussa 5.2.1 esitettyä yksinkertaista Bayes-mallia, jossa priorina oleva satunnaiskenttä on korreloimaton. Aineistot on diskretisoitu 16×16 -ruudukoksi ja gammajakauman parametrit on estimoitu aineistosta momenttimenetelmällä. Parametrien estimaatit ovat Vuorijärvelle $\hat{\alpha}_1 = 0.5$ ja $\hat{\beta}_1 = 360$ ja Susimäelle $\hat{\alpha}_2 = 1.4$ ja $\hat{\beta}_2 = 794$. Kaavan (5.3) mukaan lasketut Bayes-estimaatit on esitetty harmaasävykuvina kuvassa 6.7. Koska intensiteettien välillä ei ole spatiaalista korrelaatiota, voivat vierekkäisten ruutujen estimaatit vaihdella suuresti kuten kuvasta on havaittavissa. Lisäksi intensiteetit ovat yliestimoituja (liian suuria). Tämä johtuu siitä, että β :n estimointi aineistosta epäonnistuu. Koska spatiaalista riippuvuutta ei ole otettu huomioon on $\text{Var}(\lambda|\alpha, \beta)$ liian pieni, jolloin β tulee liian suureksi.



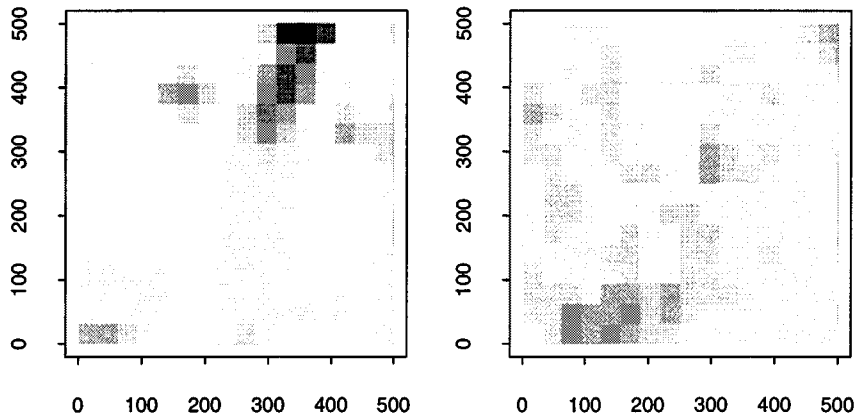
Kuva 6.7 Yksinkertaisen gamma-Poisson-mallin empiiriset Bayes-estimaatit. Oikealla Vuorijärvi ($\alpha_1 = 0.5$ ja $\beta_1 = 360$) ja vasemmalla Susimäki ($\alpha_2 = 1.4$ ja $\beta_2 = 794$). Molemmista on selvästi havaittavissa kontekstuaalisuuden puute.

Jotta ilmiölle luonteenomainen kontekstuaalisuus saadaan mukaan, tarkastellaan seuraavaksi intensiteetin estimointia käyttäen luvussa 5.2.2. esitettyä kontekstuaalisuusmallia. Tasoitusta varten painokertoimet k_{ij} on määrätty siten, että naapuriruutua painotetaan puolta vähemmän kuin itse ruutua ja $\sum_j k_{ij} = 1$. Mallinnuksessa käytetään täyttä Bayes-menetelmää, mikä johtaa Metropolis–Hastings-algoritmin konstruointiin (ks. s. 54). Algoritmin c-kielinen toteutus aliohjelmineen on esitetty liitteessä 3.

Molempien metsien tapauksessa parametrin α priorina on käytetty Gamma(1,2)-jakaumaa ja β :n priorina Gamma(100,1.5)-jakaumaa. Kaikille päivitettäville parametreille ehdotusjakauman konstruoinnissa on käytetty artikkelissa Besag ym. (1995) esitettyä menetelmää: uusi ehdotus x_{t+1} riippuu edellisestä arvosta x_t seuraavasti:

$$x_{t+1} = x_t \cdot e^{C(2U-1)},$$

jossa C on vakio ja U välillä $(0,1)$ tasajakautunut satunnaismuuttuja. Tällöin ehdotusjakaumat eivät ole symmetrisiä, mikä on otettava huomioon Metropolis–Hastings-suhteita laskettaessa. Parametrien α ja β päivityksessä vakiolle on käytetty arvoa $C = 0.5$ ja impulsseille $C = 1$. Iteraatioita on



Kuva 6.8 Kontekstuaalisuusmallin täydellä Bayes-menetelmällä estimoidut intensiteetit. Oikealla Vuorijärvi ja vasemmalla Susimäki.

tehty 20000, joista ensimmäiset 2000 on jätetty pois visuaalisen tarkastelun perusteella.

Molempien metsien tapauksessa α :n hyväksymistodennäköisyys oli noin 20 %, β :n 25 % ja impulssien kohdalla se vaihteli 40–90 %. Kuvassa 6.9 on esitetty Vuorijärven α :n, β :n ja ruudun (3, 14) impulssin simulointiketjut ja marginaaliposteriorijakaumat. Simulointiketjuista on havaittavissa selvä konvergenssi. Muiden 255 ruudun impulssien simulointiketjut ja posteriorijakaumat ovat likimain samannäköisiä. Susimäen tapauksessa tulokset ovat vastaavanlaiset.

Parametrien α ja β posteriorijakaumien keskiarvot ja 90 %:n Bayes-välit molemmista metsistä on koottu taulukkoon 6.1. Estimoidut intensiteetit, jotka on saatu kernel-tasoitettujen impulssien posteriorikeskiarvoista, on esitetty harmaasävykuvina kuvassa 6.8. Nyt intensiteettipinnoista on selvästi havaittavissa korreloituneisuus (vrt. kuva 6.7). Intensiteettien avulla on estimoitu myös puiden lukumäärät jokaisessa ruudussa. Tulokset on esitetty numeerisessa muodossa 16×16 -rasterina taulukoissa 6.2 ja 6.3. Vertailun vuoksi myös alkuperäiset aineistot on esitetty samoissa taulukoissa.

Taulukko 6.1 Parametrien α ja β posteriorikeskiarvot ja 90 %:n Bayes-välit (1= Vuorijärvi, 2= Susimäki).

Parametri	Keskiarvo	90 %:n Bayes-väli
α_1	0.18	0.14 – 0.21
β_1	77.61	66.53 – 89.40
α_2	0.25	0.21 – 0.29
β_2	82.71	71.31 – 95.56

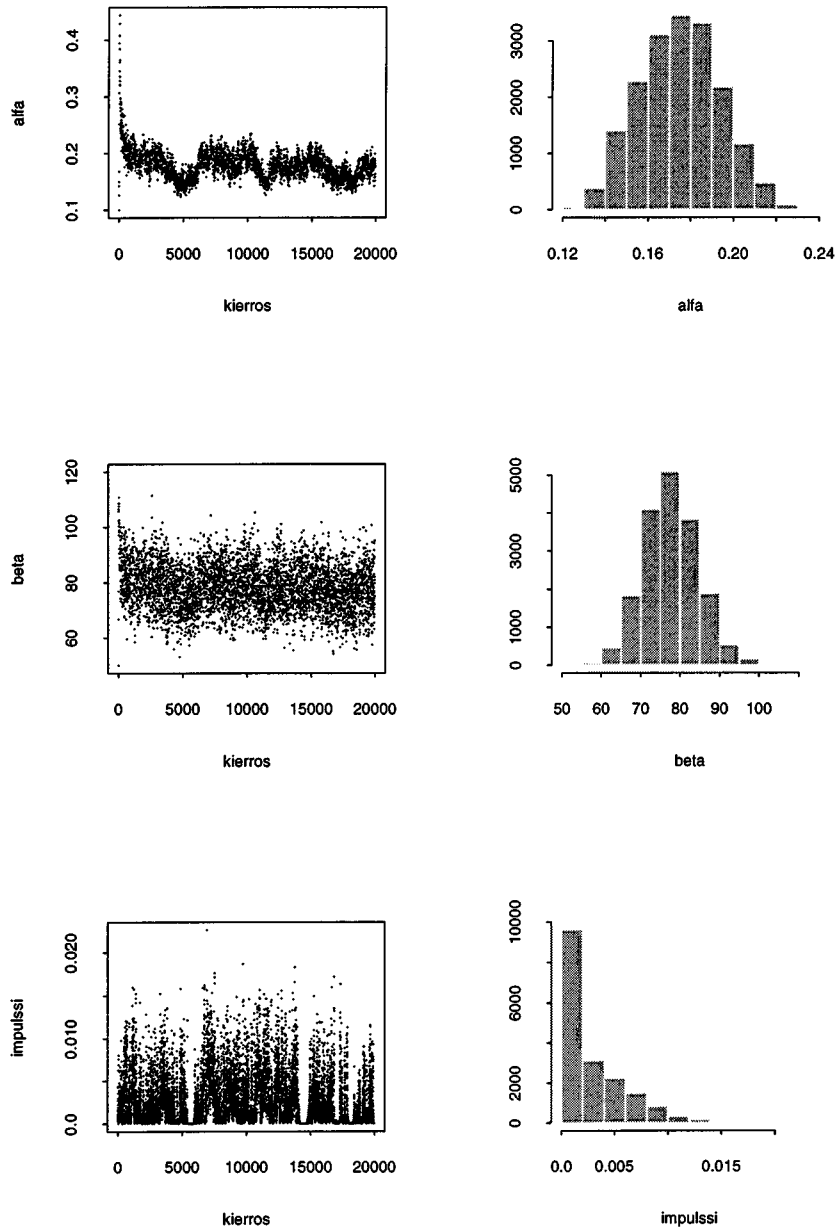
6.4 Johtopäätökset

Metsien spatiaaliset rakenteet vastasivat hyvin odotuksia. Tilastollista säännönmukaisuutta havaittiin vain Vuorijärven metsässä. Vuorijärven puut sijaitsivat pientä tiheää koivutaimikkoa lukuun ottamatta säännöllisesti. Hoidetut talousmetsät ovat tyypillisesti säännöllisiä, sillä onhan maksimaalisen puun nettotuotannon yhtenä edellytyksenä se, että puilla on tilaa kasvaa, jolloin ne eivät voi olla liian lähellä toisiaan. Sen sijaan Susimäen metsä, jossa harvennushakkuita ei ole tehty, vastasi melko hyvin täydellisen satunnaisuuden tilannetta. Puut sijaitsivat likimain satunnaisesti ja toisistaan riippumatta.

Bayesiläisen intensiteetin estimoinnin tuloksena saatiin, että spatiaalisen rakenteen on oltava mukana mallissa, mikä oli myös ennalta odotettua. Molempien metsien kohdalla riippumattomuusmalli, jossa priorina oli yksinkertainen gamma-Poisson-malli, johti yliestimoituihin intensiteetteihin ja kontekstuaalisuuden puutteeseen. Liian suuret intensiteetin estimaatit johtuivat yli suurista gammajakauman muotoparametrien arvoista ($\beta_1 = 360$ ja $\beta_2 = 794$). Tämä puolestaan johtui aineistoista estimoiduista intensiteettien variansseista, jotka riippumattomuuden vuoksi olivat liian pieniä.

Kontekstuaalisuusmalli toimi molempien metsien kohdalla hyvin. Gammajakauman muotoparametrien arvot olivat oikeaa suuruusluokkaa (posteriorikeskiarvot $\beta_1 = 77.61$ ja $\beta_2 = 82.71$) ja estimoidut intensiteetit todellisuutta vastaavia. Priorissa tehty kernel-tasointu näkyi tuloksena saadun intensiteettikentän korreloituneisuutena.

Kummankaan metsän kohdalla puutiheydessä ei ollut havaittavissa selvää trendiä missään suunnassa. Tähän voi vaikuttaa osaltaan myös se, että tutkitut alueet olivat melko pieniä, vain 0.25 hehtaarin kokoisia.



Kuva 6.9 Vuorijärven parametrien α_1 , β_1 ja ruudun (3,14) impulssin simulointi-
ketjut ja simuloidut marginaaliposteriorit.

Taulukko 6.2 Vuorimäen puiden lukumäärät 16×16 -rasterissa. Ylemmässä taulukossa on alkuperäinen aineisto ja alemmassa taulukossa kontekstuaalisuusmallilla estimoidut puiden lukumäärät.

2	1	0	0	1	0	0	1	1	1	14	16	4	0	1	2
2	1	3	0	1	1	0	2	1	0	7	8	0	1	0	1
1	1	0	1	2	2	0	1	0	3	5	5	3	1	1	0
1	1	0	0	7	5	2	1	3	3	9	3	2	2	1	0
1	1	0	1	0	2	2	2	1	8	4	1	1	3	1	1
0	1	2	0	1	2	0	1	2	6	1	1	1	5	2	4
0	3	0	1	1	0	2	1	0	3	1	1	1	2	1	3
0	1	0	0	1	0	2	0	1	0	1	1	1	0	1	1
0	1	0	1	0	2	0	0	2	1	3	0	1	2	0	1
0	0	1	0	2	0	1	1	0	0	0	1	0	1	1	2
0	0	0	0	1	1	1	1	1	0	2	1	0	0	1	1
0	1	1	1	0	0	0	0	1	1	1	2	2	0	0	1
1	1	0	0	1	2	0	1	0	2	0	0	1	0	1	0
0	0	0	0	1	0	1	0	1	1	3	2	2	1	0	0
2	2	0	1	2	1	1	1	2	2	0	0	1	0	3	0
3	5	2	1	2	0	2	2	3	1	2	1	2	2	0	1

1.8	1.4	0.6	0.4	0.7	0.4	0.5	1.1	0.9	2.6	10.6	14.3	6.3	0.4	1.0	1.8
1.9	1.6	1.1	0.8	1.4	0.7	0.7	1.2	1.0	0.9	4.8	7.0	0.8	0.8	0.5	1.0
1.2	1.1	0.7	1.1	1.9	2.9	0.5	1.0	0.6	3.4	6.2	4.5	2.2	1.5	0.9	0.4
1.2	0.8	0.5	0.7	3.4	4.7	2.7	1.3	2.3	4.8	6.1	3.4	1.9	2.0	1.0	0.6
0.9	1.0	0.5	0.7	1.0	2.9	1.4	1.6	2.2	5.6	3.2	1.5	1.5	2.6	1.3	1.4
0.9	1.4	1.1	0.7	0.8	1.1	1.4	1.2	2.5	4.8	2.8	1.0	1.9	3.4	2.8	2.9
0.5	1.4	0.8	0.8	0.8	1.0	1.3	0.9	0.6	2.4	1.2	1.0	1.0	2.1	1.4	2.5
0.5	0.9	0.7	0.4	0.9	0.7	1.0	0.7	0.8	0.9	1.2	1.0	0.9	1.1	0.9	1.4
0.3	0.6	0.4	0.7	0.9	1.0	0.7	0.8	1.0	1.1	1.4	1.0	0.9	1.1	0.9	1.1
0.2	0.4	0.5	0.6	1.1	1.0	0.9	0.7	0.8	0.8	1.0	0.9	0.5	0.8	1.0	1.6
0.3	0.4	0.5	0.6	0.7	1.0	0.8	0.9	0.8	0.8	1.1	1.1	0.7	0.4	0.9	1.5
0.5	0.8	0.8	0.7	0.7	0.6	0.6	0.6	0.8	1.0	1.3	1.4	1.0	0.6	0.5	0.8
0.7	0.7	0.6	0.5	0.8	0.9	0.6	0.6	0.6	1.3	0.9	1.0	1.2	0.4	0.5	0.3
0.6	0.5	0.3	0.4	1.1	0.7	0.9	0.5	1.3	1.6	1.6	1.4	1.6	1.1	0.6	0.5
1.4	1.9	0.5	1.1	1.6	1.3	1.1	1.5	1.7	1.7	1.0	0.8	1.4	0.7	1.1	0.5
3.7	3.9	2.2	1.1	1.6	0.8	1.5	2.0	2.6	1.7	1.2	1.4	1.6	1.5	1.2	0.9

Taulukko 6.3 Susimäen puiden lukumäärät 16×16 -rasterissa. Ylemmässä taulukossa on alkuperäinen aineisto ja alemmassa kontekstuaalisuusmallilla estimoidut puiden lukumäärät.

0	1	1	1	0	0	0	3	0	1	3	1	3	1	5	2
3	2	1	2	2	2	1	0	1	2	0	0	1	1	1	4
1	2	0	0	5	1	0	1	2	3	0	2	2	3	1	1
2	1	2	5	2	0	2	3	0	2	0	4	1	2	1	1
3	4	1	1	2	0	0	1	0	0	0	1	3	2	0	1
3	1	0	1	5	1	2	1	1	5	2	1	2	0	0	0
2	4	1	1	2	2	1	1	3	1	2	0	3	4	1	1
1	2	2	1	1	4	2	1	1	6	2	3	2	0	2	1
3	2	3	2	1	0	3	2	1	2	2	3	0	1	2	1
1	3	4	0	1	1	2	2	5	0	0	0	1	2	2	1
1	3	3	1	2	2	1	0	2	5	2	2	2	0	1	1
3	1	0	0	5	2	3	1	3	1	2	1	2	4	0	2
1	1	2	1	0	1	1	0	6	2	1	1	1	0	0	1
3	4	2	2	3	5	4	4	1	4	3	3	2	0	0	1
4	2	5	3	5	4	0	6	2	0	1	0	1	2	1	2
0	2	6	3	4	5	2	2	0	1	0	0	3	0	0	0

0.8	1.4	1.0	1.0	1.2	0.7	0.8	1.2	1.2	1.5	1.6	1.8	2.0	2.1	2.9	3.4
1.9	2.0	1.5	1.6	2.2	1.4	0.9	1.0	1.1	1.7	0.8	1.0	1.2	1.7	2.0	2.6
1.6	1.7	1.0	1.8	2.6	1.2	0.8	1.4	1.7	2.4	1.4	1.5	2.1	1.9	1.6	1.5
2.3	1.7	1.7	2.6	2.7	1.2	1.3	1.7	1.2	1.4	0.9	2.1	2.7	2.1	1.1	1.2
3.2	2.6	1.4	1.6	2.3	0.7	0.8	1.3	0.7	0.8	0.9	1.6	2.1	1.5	0.7	0.8
2.6	2.0	0.9	1.6	2.9	1.9	1.2	1.3	1.4	2.6	1.5	1.3	2.0	1.1	0.7	0.5
2.4	2.4	1.4	1.2	2.4	1.9	1.8	1.3	2.1	3.3	2.6	1.8	2.5	2.1	1.3	1.1
1.9	2.2	2.0	1.3	1.8	2.2	2.3	1.6	1.7	3.4	3.0	2.3	1.9	1.2	1.6	1.2
2.0	2.8	3.0	1.6	1.1	1.4	2.1	1.8	1.9	1.8	2.0	1.7	1.1	1.2	1.8	1.4
1.8	2.7	2.9	1.3	1.3	1.2	1.9	2.1	2.6	1.5	1.3	1.0	1.0	1.4	1.8	1.5
1.9	2.4	2.1	1.5	2.1	2.2	1.5	1.7	2.8	2.7	2.0	1.6	1.8	1.1	1.3	1.3
1.8	1.6	1.2	0.9	2.4	2.6	2.0	1.3	2.7	2.3	1.7	1.8	2.6	1.7	0.8	1.3
2.2	1.5	1.4	1.1	1.1	2.6	1.3	1.5	2.9	2.4	2.1	1.2	1.5	0.6	0.6	1.1
3.0	3.0	2.6	2.1	3.2	3.7	2.9	3.4	2.7	2.8	2.8	2.1	1.5	0.6	0.6	1.1
2.6	2.9	4.3	3.4	3.8	4.3	2.7	3.5	2.5	0.7	1.4	1.2	1.8	1.3	0.9	1.2
0.9	2.8	4.4	3.8	4.4	3.9	2.2	2.4	0.8	0.6	0.5	0.6	1.5	0.8	0.6	0.7

Kiitosmaininnat

Pro gradu -tutkimukseni ohjaajana on toiminut dosentti Antti Penttinen. Tutkimusta on tukenut Suomen Akatemian projektit ”Spatiaalisen ympäristöaineiston analyysi” (ASED) n:o 37206 ja ”Paikkatiedon mallintaminen ja GIS” n:o 39372. Lisäksi haluan kiittää dosentti Timo Kuuluvaista aineiston luovuttamisesta käyttöni ja FT Harri Högmanderia kielipillisistä ja työn ulkoasuun liittyvistä parannusehdotuksista.

Kirjallisuutta

- Bartlett, M. S., (1964): Spectral analysis of two-dimensional point processes. *Biometrika*, **44**, 299–311.
- Besag, J. & Gleaves, J. T., (1973): On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical Institute*, **45**(1), 153–158.
- Besag, J., Green, P. J., Higdon, D. & Mengersen, K., (1995): Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3–66.
- Byth, K. & Ripley, B. D., (1980): On sampling spatial patterns by distance methods. *Biometrics*, **36**, 279–284.
- Cressie, N. A. C., (1991): *Statistics for Spatial Data*. John Wiley and Sons, New York.
- Diggle, P. J., (1983): *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Diggle, P. J., (1985): A kernel method for smoothing point process data. *Applied Statistics*, **34**(2), 138–147.
- Feller, W., (1971): *An Introduction to Probability Theory and its Applications*, volume 2, 2nd edition. John Wiley and Sons, New York.
- Fisher, R. A., Thornton, H. G. & Mackenzie, W. A., (1922): The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples. *Annals of Applied Botany*, **9**, 325–359.

- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B., (1995): *Bayesian Data Analysis*. Chapman and Hall, London.
- Green, P. J. & Silverman, B. W., (1994): *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London.
- Heikkinen, J. & Arjas, E., (1998): Nonparametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*. Painossa.
- Hogg, R. V. & Craig, A. T., (1978): *Introduction to Mathematical Statistics*. Collier Macmillan, London.
- Ickstadt, K. & Wolpert, R. L., (1996): Multiresolution assessment of forest inhomogeneity. *Case Studies in Bayesian Statistics*, **3**.
- Kuuluvainen, T., Penttinen, A., Leinonen, K. & Nygren, M., (1996): Statistical opportunities for comparing stand structural heterogeneity in managed and in primeval forests. An example from boreal spruce forest in southern Finland. *Silva Fennica*, **30**(2–3), 315–328.
- Lewis, P. A. W. & Shedler, G. S., (1979): Simulation of non-homogenous Poisson processes by thinning. *Naval Research Logistics Quarterly*, **26**, 403–413.
- Matérn, B., (1960): Spatial variation. *Meddelanden från Statens Skogsforskningsinstitut*, **49**(5), 1–144.
- Møller, J., Syversveen, A. R. & Waagepetersen, R. P., (1998): Log Gaussian Cox processes. *Scandinavian Journal of Statistics*.
- Ogata, Y., (1996): Evaluating spatial Bayesian models –Two computational methods. *Journal of Statistical Planning and Inference*, **51**, 1–18.
- Ogata, Y. & Tanemura, M., (1986): Likelihood estimation of interaction potentials and external fields of inhomogenous spatial point patterns. I. S. Francis, B. F. J. Manly & F. C. Lam, editoijat, *Pacific Statistical Congress*, North-Holland.
- Ripley, B. D., (1976): The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**, 255–266.
- Ripley, B. D., (1981): *Spatial Statistics*. John Wiley and Sons, New York.
- Silverman, B. W., (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

Stoyan, D., Kendall, W. S. & Mecke, J., (1995): *Stochastic Geometry and its Applications*, 2nd edition. John Wiley and Sons, Chichester.

Liite 1

Matérnin klusteriprosessin ja gamma-Poisson-rasteriprosessin simulointialgoritmit Splus:lla ohjelmoituina.

```
function(a = 1, b = 1, R = 0.1, roo = 15, eta = 250)
{
  # Funktio simuloi Maternin klusteriprosessin.
  # Parametrit: a = alueen leveys
  #             b = alueen korkeus
  #             R = klusterin säde
  #             roo = intensiteetti klustereiden keskipisteille
  #             eta = intensiteetti klusterin sisällä
  #
  lkm <- rpois(1, a * b * roo)
  x0 <- runif(lkm, 0, a)
  y0 <- runif(lkm, 0, b)
  pi <- 3.141592654
  n <- rpois(lkm, pi * R^2 * eta)
  otosx <- vector()
  otosy <- vector()
  for(i in 1:lkm) {
    k <- 0
    while(k != n[i]) {
      x1 <- runif(1, x0[i] - R, x0[i] + R)
      y1 <- runif(1, y0[i] - R, y0[i] + R)
      r <- sqrt((x1 - x0[i])^2 + (y1 - y0[i])^2)
      if(r < R & x1 < a & y1 < b & x1 > 0 & y1 > 0) {
        otosx <- c(otosx, x1)
```

```

        otosy <- c(otosy, y1)
        k <- k + 1
    }
}
list(x = otosx, x = otosy)
}

```

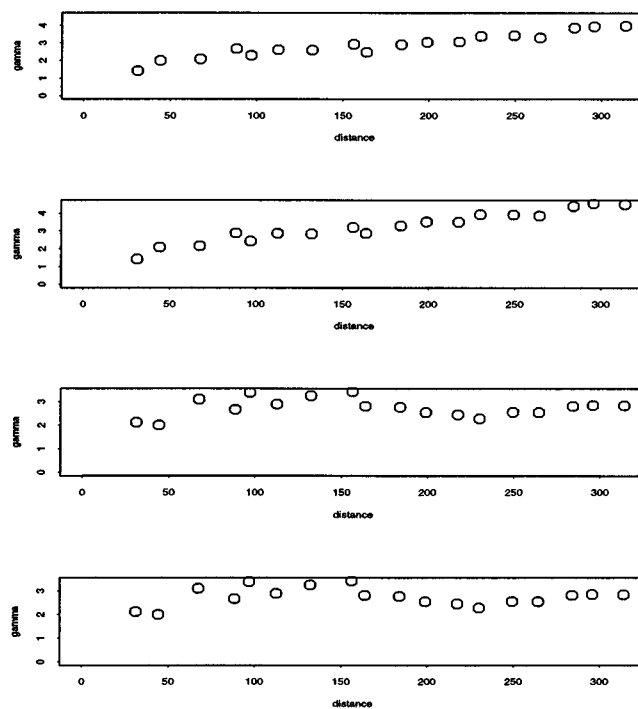
```

function(lamdat)
{
# Funktio simuloi Gamma-Poisson -prosesssin.
# Parametrit: lamdat = intensiteettihila (=lista)
#
n <- length(lamdat[[1]]) - 1
vali <- lamdat[[1]][2] - lamdat[[1]][1]
lkm <- matrix(rpois(length(lamdat[[3]]), lamdat[[3]]), nrow = n, ncol = n)
otosx <- vector()
otosy <- vector()
for(i in 1:n) {
    for(j in 1:n) {
        x <- runif(lkm[j, i], vali * (i - 1), vali * i)
        y <- runif(lkm[j, i], vali * (j - 1), vali * j)
        otosx <- c(otosx, x)
        otosy <- c(otosy, y)
    }
}
list(x = otosx, y = otosy)
}

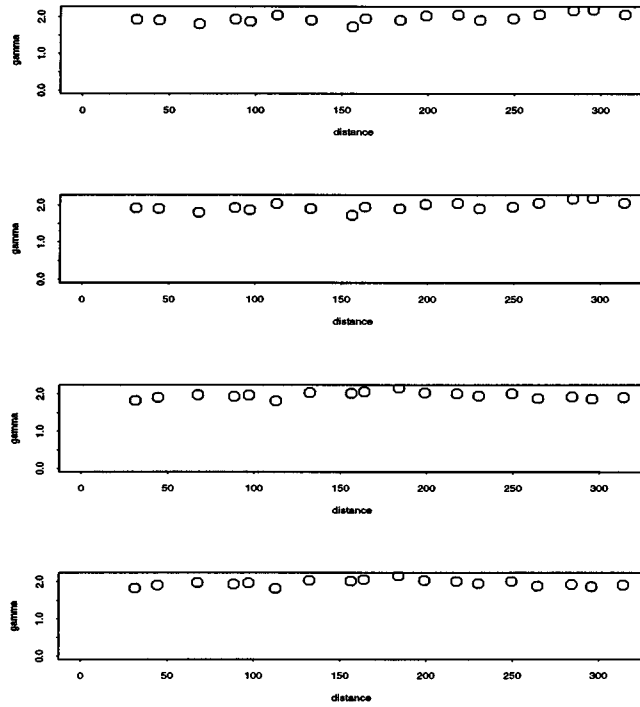
```

Liite 2

Vuorijärven ja Susimäen variogrammit alkuperäiselle aineistolle ja median polish -jäännöksille.



Kuva 6.10 Vuorijärven variogrammit: kaksi ylintä, joista ylempi on median polish -jäännöksille ja alempi alkuperäiselle aineistolle, on laskettu pohjois-etelä-suunnassa ja kaksi alinta vastaavasti itä-länsi-suunnassa.



Kuva 6.11 Susimäen variogrammit: kaksi ylintä, joista ylempi on median polish -jäännöksille ja alempi alkuperäiselle aineistolle, on laskettu pohjois-etelä-suunnassa ja kaksi alinta vastaavasti itä-länsi-suunnassa.

Liite 3

Metropolish–Hastings-algoritmin c-kielinen toteutus kontekstuaalisuusmallin posteriorin simuloinnille. Ohessa on pääohjelma ja tärkeimmät aliohjelmat. Tasoituksen vuoksi aluetta on laajennettu siten, että jokaisella reunapiksellä on myös neljä naapuripikseliä. Alueeseen kuuluvat pikselit tunnistetaan taustan avulla: tutkittavaan alueeseen kuuluvan pikselin tausta-arvo on yksi ja reunapikseleiden nolla. Aluetta käsitellään yhden indeksin avulla, jolloin impulssit ovat vektorissa.

```
main()
/* pääohjelma */
{
    int j, t=1;
    alustus();
    while(t<=KIERR) {
        paivita_impulssit();
        paivita_alfa();
        paivita_beta();
        talleta(t);
        t++;
    }
    estimointi();
    lue_tiedostoon();
}
```

```
-----

void alustus(void)
/* aliohjelma tekee tarvittavat alustukset */
```

```

{
/* taustan alustus */
int px, py, x, y, j, i;
FILE *f;
for(i=0; i<PKOKO; i++) {
    px = i/PSIVU;
    py = i/PSIVU;
    x = px + 1;
    y = py + 1;
    j = y * SIVU + x;
    tausta[j] = 1;
}
alfa=ALKUARVO1;
beta=ALKUARVO2;
srand48(time(NULL));    /* satunnaisgeneraattorin alustus */

/* impulssien alustus */
for(j=0; j<KOKO; j++) {
    if(tausta[j]) impulssi[j] = 0.001;
}

/* datan lukeminen */
f = fopen(DATA_TIEDOSTO,"r");
if(!f) {
    printf("Tiedosto \"DATA_TIEDOSTO\" ei aukea!\n");
    exit(1);
}
for(j=0; j<KOKO; j++) {
    if(tausta[j]) fscanf(f,"%d", &data[j]);
}
fclose(f);
}

-----

void paivita_impulssit(void)
/* aliohjelma paivittaa impulssitaulukon (=impulssivektorin) */
{
    int t, m, j;
    double paino, hyv_tn, apu, tulo;
    double uusi_impulssi, vanha_lamda, uusi_lamda;
    for (j=0; j<KOKO; j++) {

```

```

if(tausta[j]) {
    uusi_impulssi = impulssi[j]*exp(C1*(2*runif()-1));
    m = naapurusto(j);
    paino = 2*m-(m-2);
    vanha_lamda = laske_lamda(j);
    uusi_lamda = vanha_lamda+(1-m/paino)*(uusi_impulssi-impulssi[j]);
    apu = pow((uusi_lamda/vanha_lamda),data[j])
        *exp(-ALA*(uusi_lamda-vanha_lamda));
    tulo = apu;
    if(tausta[j-1]) {
        m = naapurusto(j-1);
        paino = 2*m-(m-2);
        vanha_lamda = laske_lamda(j-1);
        uusi_lamda = vanha_lamda+(1/paino)*(uusi_impulssi-impulssi[j]);
        apu = pow((uusi_lamda/vanha_lamda),data[j-1])
            *exp(-ALA*(uusi_lamda-vanha_lamda));
        tulo = tulo*apu;
    }
    if(tausta[j+1]) {
        m = naapurusto(j+1);
        paino = 2*m-(m-2);
        vanha_lamda = laske_lamda(j+1);
        uusi_lamda = vanha_lamda+(1/paino)*(uusi_impulssi-impulssi[j]);
        apu = pow((uusi_lamda/vanha_lamda),data[j+1])
            *exp(-ALA*(uusi_lamda-vanha_lamda));
        tulo = tulo*apu;
    }
}
if(tausta[j-SIVU]) {
    m = naapurusto(j-SIVU);
    paino = 2*m-(m-2);
    vanha_lamda = laske_lamda(j-SIVU);
    uusi_lamda = vanha_lamda+(1/paino)*(uusi_impulssi-impulssi[j]);
    apu = pow((uusi_lamda/vanha_lamda),data[j-SIVU])
        *exp(-ALA*(uusi_lamda-vanha_lamda));
    tulo = tulo*apu;
}
if(tausta[j+SIVU]) {
    m = naapurusto(j+SIVU);
    paino = 2*m-(m-2);
    vanha_lamda = laske_lamda(j+SIVU);
    uusi_lamda = vanha_lamda+(1/paino)*(uusi_impulssi-impulssi[j]);
}

```

```

        apu = pow((uusi_lamda/vanha_lamda),data[j+SIVU])
            *exp(-ALA*(uusi_lamda-vanha_lamda));
        tulo = tulo*apu;
    }
    hyv_tn = pow((uusi_impulssi/impulssi[j]),(alfa-1))
        *exp(-beta*(uusi_impulssi-impulssi[j]))*tulo
        *(uusi_impulssi/impulssi[j]);
    if(runif()<=hyv_tn) {
        impulssi[j] = uusi_impulssi;
        hyvaksytty[j]++;
    }
}
}
}
}

```

```

void paivita_alfa(void)
/* aliohjelma päivittää alfan */
{
    int i;
    double hyv_tn, uusi_alfa, apu = 1.00;
    uusi_alfa = alfa*exp(C2*(2*runif()-1));
    for(i=0; i<KOKO; i++) {
        if(tausta[i]) apu = (exp(lgamma(alfa))/exp(lgamma(uusi_alfa)))
            *pow(beta*impulssi[i],uusi_alfa-alfa)*apu;
    }
    hyv_tn = apu*(uusi_alfa/alfa)*pow(uusi_alfa/alfa,A-1)
        *exp(-B*(uusi_alfa-alfa));
    if(runif()<=hyv_tn) {
        alfa = uusi_alfa;
        hyvak_alfa++;
    }
}

```

```

void paivita_beta(void)
/* aliohjelma päivittää betan */
{
    int i;
    double hyv_tn, uusi_beta, apu=0;
    uusi_beta = beta*exp(C3*(2*runif()-1));

```

```

for(i=0;i<KOKO; i++) {
    apu = impulssi[i]+apu;
}
hyv_tn = pow(uusi_beta/beta,PKOKO*alfa+C-1)*exp(-(uusi_beta-beta)
            *(apu+D))*(uusi_beta/beta);
if(runif()<=hyv_tn) {
    beta = uusi_beta;
    hyvak_beta++;
}
}

```

```

-----

double laske_lamda(int i)
/* aliohjelma laskee tasoitettun intensiteetin pikselissä i */
{
    int m;
    double s, paino;
    m = naapurusto(i);
    s = naapurisumma(i);
    paino = 2*m-(m-2);
    return s/paino + (1-m/paino)*impulssi[i];
}

```

```

-----

double naapurisumma(int i)
/* aliohjelma laskee pikselin i naapurien impulssien summan */
{
    return impulssi[i-1]+impulssi[i+1]+impulssi[i-SIVU]+impulssi[i+SIVU];
}

```

```

-----

int naapurusto(int i)
/* aliohjelma laskee pikselin i naapurien lukumäärän */
{
    return tausta[i-1]+tausta[i+1]+tausta[i-SIVU]+tausta[i+SIVU];
}

```