

KLAIPĖDOS UNIVERSITETAS
Gamtos ir matematikos mokslų fakultetas
Statistikos katedra

**ERDVINIŲ TAŠKINIŲ PROCESŲ ANALIZĖ IR MODELIAVIMAS
ATVIROJO KODO PROGAMA R**

Statistikos ir operacijų tyrimo studijų programos
magistro baigiamasis darbas

Autorius
Vadovas

SOT-10 gr. stud. Salomėja Sovaitė
Prof. Kęstutis Dučinskas

Klaipėda, 2012

TURINYS

TURINYS.....	2
PAVEIKSLŲ SĄRAŠAS.....	3
SANTRAUKA.....	5
SUMMARY	5
ĮVADAS.....	6
I. ERDVINIAI TAŠKINIAI PROCESAI	8
I.1. Pagrindinės sąvokos	10
I.2. Puasono taškinis procesas.....	14
I.3. Erdvinio taškinio proceso analizės technikos	16
I.4. Kiti erdviniai taškiniai procesai	22
II. KLASTERINIAI TAŠKINIAI PROCESAI	25
II.1. Puasono grupės klasteriniai procesai	26
II.2. Neigiamas binominis procesas	27
III. ŽYMĖTI TAŠKINIAI PROCESAI	29
III.1. Multi-tipo Puasono modeliai.....	30
III.2. Žymėtų taškinių procesų atsitiktinumo testai.....	32
IV. ETP ANALIZĖ R APLINKOJE.....	34
IV.1. Įvadas į atvirojo kodo programą R.....	34
IV.2. Puasono proceso tyrimas	38
IV.3. Žymėto taškinio proceso analizė.....	56
IV.4. Klasterinio ETP analizė	61
IŠVADOS	78
LITERATŪRA	79
PRIEDAI	81
Priedas nr. 1 23 pav. Atstumų funkcijos F , G , J , K	81
Priedas nr. 2 38 pav. Kryžminė G_{ij} funkcija visoms žymių poroms	82
Priedas nr. 3 39 pav. Kryžminė J_{ij} funkcija visoms žymių poroms.....	83
Priedas nr. 4 40 pav. Kryžminė $G_i(r)$ funkcija nuo fiksuotos žymės iki kito bet kokio taško.....	84
Priedas nr. 5 47 pav. ETP X atstumų funkcijos	85

PAVEIKSLŲ SĄRAŠAS

1 pav. Geostatistiniai duomenys	8
2 pav. Regioniniai duomenys	8
3 pav. Taškiniai duomenys	9
4 pav. ETP taškai iš atrankos lango W	11
5 pav. Žemės drebėjimų epicentrai Los Andžele. Stebėjimų langas imtas nuo 32 iki 37 šiaurės platumos ir nuo -122 iki -114 vakarų ilgumos	14
6 pav. ETD pasiskirstymai	17
7 pav. Pasiskirstymų išsidėstymas pagal mažėjančią klasterizaciją	25
8 pav. Puasono klasterinis procesas	26
9 pav. Ranka suvestų duomenų taškinis vaizdas	36
10 pav. Nuskaitytų koordinačių taškinis vaizdas	37
11 pav. Erdvinis taškinis procesas p	39
12 pav. Taškų skaičius kvadratuose	40
13 pav. Kernel'io intensyvumo tankio įvertinys	41
14 pav. Trimatis intensyvumo atvaizdavimas	41
15 pav. Tuščių erdvių atstumai	42
16 pav. Tuščių erdvių atstumo pasiskirstymo funkcija F	43
17 pav. Poriniai taškų atstumai	43
18 pav. Porinių atstumų funkcija K	44
19 pav. Artimiausio kaimyno atstumai	44
20 pav. Artimiausio kaimyno atstumų pasiskirstymo funkcija G	45
21 pav. Porinė koreliacinė funkcija $g(r)$	46
22 pav. Kombinacinė funkcija J	47
23 pav. Atstumų funkcijos F, G, J, K	<i>Priedas nr.1</i> 81
24 pav. χ^2 testas	48
25 pav. Kolmogorov-Smirnov testas	49
26 pav. Monte Carlo testas	50
27 pav. Tankio funkcijos $fit1$ prognozė	52
28 pav. Tankio funkcijos $fit2$ prognozė	52
29 pav. Tankio funkcijos $fitx$ prognozė	52
30 pav. Keturių tipų paklaidos funkcijos $fit1$	53
31 pav. Keturių tipų paklaidos funkcijos $fit2$	54

32 pav. Keturių tipų paklaidos funkcijos <i>fitx</i>	54
33 pav. Žymėtas taškinis procesas <i>p</i>	57
34 pav. Žymių pasiskirstymas	58
35 pav. Žymių intensyvumo pasiskirstymas	58
36 pav. Kryžminė G_{ij} funkcija žymių porai „Beržas“ ir „Ažuolas“	59
37 pav. Kryžminė G_{ij} funkcija žymių porai „Klevas“ ir „Ažuolas“	60
38 pav. Kryžminė G_{ij} funkcija visoms žymių poroms	Priedas nr.2 82
39 pav. Kryžminė J_{ij} funkcija visoms žymių poroms	Priedas nr.3 83
40 pav. Kryžminė $G_i(r)$ funkcija nuo fiksuotos žymės iki kito bet kokio taško.....	Priedas nr.4 84
41 pav. Erdvinis taškinis procesas <i>X</i>	61
42 pav. Taškų skaičius kvadratuose.....	62
43 pav. Kernel'io intensyvumo įvertinys	63
44 pav. Tuščių erdvių atstumai.....	63
45 pav. Poriniai atstumai.....	64
46 pav. Artimiausio kaimyno atstumai	64
47 pav. ETP <i>X</i> atstumų funkcijos	Priedas nr. 5 85
48 pav. χ^2 testas.....	65
49 pav. Monte Carlo testas ETP <i>X</i>	66
50 pav. Keturių tipų paklaidos tankio funkcijos <i>fit1</i>	68
51 pav. Keturių tipų paklaidos tankio funkcijos <i>fit2</i>	68
52 pav. Keturių tipų paklaidos tankio funkcijos <i>fit3</i>	69
53 pav. Keturių tipų paklaidos tankio funkcijos <i>fit4</i>	69
54 pav. <i>K</i> funkcija įvertinta mažiausių kvadratų metodu su tankio funkcija <i>fit4</i>	70
55 pav. <i>K</i> funkcija įvertinta mažiausių kvadratų metodu su tankio funkcija <i>fit4</i>	71
56 pav. Žymėtas ETP <i>X</i>	72
57 pav. Žymėto ETP <i>X</i> žymių pasiskirstymas	73
58 pav. ETP <i>X</i> erdvinis žymių trendas	73
59 pav. Multi-tipo taškinis procesas <i>X</i>	74
60 pav. Kryžminė G_{cross} funkcija porai „Žemieji“-„Aukštieji“	75
61 pav. Kryžminė K_{cross} funkcija porai „Vidutiniai“-„Aukštieji“	76
62 pav. Kryžminė J_{cross} funkcija porai „Žemieji“-„Vidutiniai“	76
63 pav. Žymių koreliacinė funkcija	77

SANTRAUKA

Sovaitė S. Erdvinių taškinių procesų analizė ir modeliavimas atvirojo kodo programa R. Statistikos ir operacijų tyrimo magistro studijų programos baigiamasis darbas. Darbo vadovas prof. K. Dučinskas, Klaipėdos universitetas: Klaipėda, 2012. – 85 p.

Raktažodžiai: erdviniai taškiniai procesai, Puasono procesas, klasteriniai procesai, žymėti taškiniai procesai, modelių pritaikymas, pilno erdvinio atsitiktinumo tyrimas.

Šiame darbe pateikiamos teorinės erdvinių taškinių procesų analizavimo ir modeliavimo žinios, aptariamas Puasono procesas, klasteriniai bei žymėti taškiniai procesai. Taip pat atliekamas taškinių procesų, Puasono, žymėtų bei klasterinio, tyrimas/analizė atvirojo kodo programos R aplinkoje. Pateikiamos išvados apie atliktą analizę.

SUMMARY

Sovaitė S. Analysing and Modelling Spatial Point Patterns in R. Final work of Statistics and Operation Research master study programme. The leader of the work prof. K. Dučinskas, Klaipėda University: Klaipėda, 2012. – 85 p.

Main keywords: spatial point pattern, Poisson model, clustering, marked point patterns, modell fitting, complete spatial randomness.

In this final work of Statistics and Operation Research master study programme is given theoretical material for spatial point process analysis and modelling, discussed Poisson process, processes with clusters and marks. There are given some analysis of Poisson process and process with clustered data and marks using an open code statistic package R. Conclusions about the given analysis are represented.

IVADAS

Erdviniai taškiniai procesai – tai viena iš erdvinės statistikos rūšių. Erdvinių taškinių procesų analizė pirmą kartą buvo paminėta 1930-aisiais metais botanikų ir ekologų darbuose. Per pastaruosius dvidešimt metų, ji buvo pradėta naudoti ir kitose mokslo srityse, tokiose kaip astronomija, geografija, kriminalistika, ekologija ir kitos. Bendru atveju, taškinių procesų analizė gali būti panaudota apibūdinant bet kokio pobūdžio stebinio duomenis. **Praktinio erdvinių taškinių procesų pritaikomumo** pavyzdžiai galėtų būti labiausiai nusikalstamų vietovių analizė ieškant nusikaltimų „karštojo taško“, kur nusikalstamumas yra didžiausias. Kitas pavyzdys būtų tam tikra liga sergančių žmonių gyvenamų vietų analizė, norint atrasti ligos židinius ir pan. Tokio pobūdžio taškinių procesų analizės padeda strategams ir įvairioms institucijoms paskirstyti nepakankamus išteklius skirtingose srityse, kur labiausiai reikia. (Burden F.F., 2003, 1 p.)

Šiame darbe naudojamoje literatūroje galime pastebėti, jog dauguma literatūros šaltinių yra ne senesni negu du tūkstantieji metai. Galime teigti, jog tik per pastarąjį dešimtmetį, erdviniai taškiniai procesai buvo pradėti išsamiai analizuoti. Taškinių procesų užuomazgų galime atrasti dar 1965 m. NASA memorandumo dokumente (Clifford Cohen A. et al., 1965). 1979 m. mokslininkas B.D. Ripley analizavo atsitiktinumo testus taškiniams vaizdams. Perkopus du tūkstantuosius metus, pastebime, kad kur kas daugiau mokslininkų pradėjo analizuoti erdvinius taškinius procesus, įvairius jų analizavimo metodus, nulinės hipotezės patikrinimo testus, taškų pasiskirstymus ir kt. Vieni iš žymiausių mokslininkų, kurių darbų apie erdvinius taškinius procesus galima rasti daugiausiai yra A. Baddeley ir J.Moller.

Lietuvoje erdviniai taškiniai procesai pradėti analizuoti tik pastaraisiais keletą metų. Vienintelis žinomas erdvinės statistikos literatūros šaltinis yra mokomoji knyga K. Dučinsko ir J. Šaltytės-Benth „Erdvinė statistika“. Joje supažindinama su plačiausiai taikomais erdvinių duomenų statistiniais modeliais bei jų analizės ir prognozės metodais. Šis vadovėlis aprašo bendrai erdvinę statistiką, tačiau akcentuoja geostatistinių duomenų analizę, o apie taškinius vaizdus kalbama ne daug. Literatūros lietuvių kalba apie erdvinius taškinius procesus nėra, o vienintelis informacijos šaltinis buvo modulio „Atsitiktiniai taškiniai procesai“ paskaitų medžiaga. Mokslinių straipsnių šia tema taip pat nerasta, o Lietuvoje vienintelė vieta, kur pradėti analizuoti erdviniai taškiniai procesai, yra Klaipėdos universitetas. Šioje vietoje galėtume įvardinti šio darbo rašymo **problematiką**, jog visos teorinės žinios bei praktinis jų pritaikomumo aprašymas, analizuojant erdvinius taškinius procesus, yra versti iš anglų kalbos. Užsienio literatūroje mokslininkai tam tikrus terminus naudoja skirtingai, pvz. galime pastebėti, jog per pastaruosius penkerius metus

taškiniai vaizdai mokslininkų vis labiau buvo pradėti vadinti erdviniais taškiniais procesais, todėl reikia išsiginčyti į jų naudojamą terminologiją, norint taisyklingai suprasti pateikiamą informaciją.

Šio darbo pirmame skyriuje yra apibrėžta pagrindinė naudojama terminologija, remiantis A. Baddeley, J. Moller ir K. Dučinsko darbais. Visa pateikta teorinė dalis yra versta iš anglų kalbos, todėl galime drąsiai teigti, jog šio darbo *originalumas* yra tai, jog šis magistro baigiamasis darbas yra vienintelis mokslinis darbas lietuvių kalba apie erdviųjų taškinių procesų analizę.

Kad erdviniai taškiniai procesai tampa vis labiau *aktualūs* įvairiose srityse ir jų analizė yra vis labiau plėtojama, įrodo tai, jog 2012 m. sausio mėnesį Aalborgo universitete (Danija) buvo išleistas mokslinis straipsnis apie Cox tipo taškinius procesus, kuriems būdingi taškų klasteriai (Møller J. et. al., 2012). Iš tiesų, daugelis mokslininkų būtent su tokiais erdviniais taškiniais procesais susiduria, nes „idealus“ erdvinio taškinių procesų Puasono modelis vis rečiau sutinkamas realybėje. Kitaip sakant, vis dažniau yra atmetama nulinė hipotezė, jog taškinis procesas yra homogeninis Puasono procesas ir jam yra būdingas pilnas erdvinis atsitiktinumas. Klasteriniams taškiniams procesams šiuo metu skiriamas didelis dėmesys.

Šiame darbe šalia klasikinio Puasono modelio yra aprašyti ir klasteriniai procesai bei šių procesų praktinis realizavimas ir analizė yra atlikta atvirojo kodo programa R. Rašant šį darbą buvo išsikeltas toks *darbo tikslas*: aprašyti erdviųjų taškinių procesų teorinę dalį bei atlikti praktinį tyrimą, teorinio procesų aprašymo pagrindu. Šiam tikslui pasiekti buvo suformuluoti *uždaviniai*:

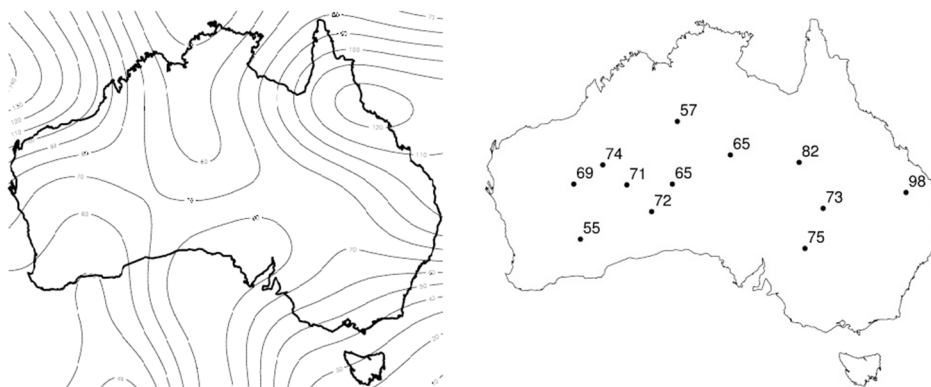
- apžvelgti kuo daugiau naujausios literatūros anglų kalba ir, apibrėžus terminologiją lietuvių kalba, aprašyti erdviųjų taškinių procesų teorinę dalį;
- išsiginčyti į klasikinį Puasono taškinį ir klasterinį taškinį procesus bei aprašyti jų panaudojimo galimybes erdviųjų taškinių duomenų analizėje;
- atlikti dviejų erdviųjų taškinių procesų - su vienodo pasiskirstymo taškais ir klasterinio proceso – tyrimą/statistinę analizę programos R aplinkoje;
- atlikti žymėtų taškinių procesų statistinę analizę (su kokybinėmis ir kiekybinėmis žymėmis).

I. ERDVINIAI TAŠKINIAI PROCESAI

Erdvinių procesų aprašomąją dalį pradėsime nuo erdvinių duomenų apžvalgos. Remiantis mokslininko A.Baddeley darbais (2008, 2 p.), erdviniai duomenys yra skirstomi į:

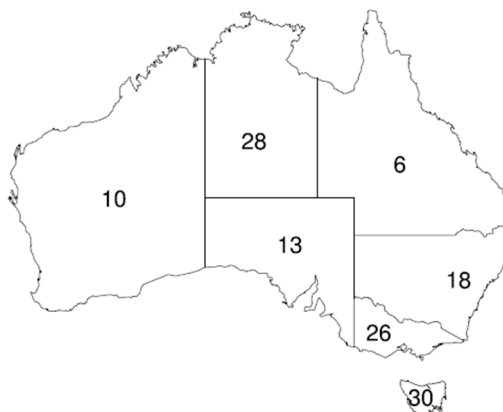
- geostatistinius duomenis (ang. *geostatistical data*),
- regioninius duomenis (ang. *regional data, lattice*),
- taškinių vaizdų duomenis (ang. *point pattern data*).

Geostatistinių duomenų atveju, yra stebima visa sritis, tačiau matavimai atliekami tik tam tikruose taškuose. Šių taškų matavimai ir yra erdvinio proceso duomenys.



1 pav. Geostatistiniai duomenys

Regioninių duomenų atveju, stebima sritis yra padalinama į fiksuotus regionus ir matavimai atliekami kiekvienam regionui. Šie regioniniai stebinio matavimai yra erdvinio proceso duomenys. Toks duomenų tipas dar vadinamas mozaikomis (ang. *tessellations*), kai erdvė suskirstoma į nepersiklojančias sritis. Mozaika gali būti „tikra“, pavyzdžiui, žemynas, padalintas į valstybes ar sritis. Mozaika gali būti visiškai dirbtinė, t.y. netikra, pavyzdžiui, *Dirichlet* mozaika apibrėžta aibe taškų.



2 pav. Regioniniai duomenys

Taškinių vaizdų atveju, yra stebimos tam tikro įvykio vietos-taškai erdvėje (ang. *locations, sites*), pavyzdžiui, meteoritų nukritimo, kriminalinių nusikaltimų ar medžių žuvimo vietos. Visos šios vietos yra pažymimos ir tai sudaro erdvinius vaizdus.



3 pav. Taškiniai duomenys

Erdvinė geostatistika nuo erdvinės taškinės statistikos skiriasi tuo, jog geostatistika įvykio vietas laiko kaip paaiškinamąją informaciją, o tiria stebinio reikšmę tose vietose. Tuo tarpu taškinė statistika laiko tiriamuoju objektu ir įvykio vietą, ir stebinio reikšmę tose vietose.

Šiame darbe nagrinėsime erdvinius procesus, kurių duomenys yra erdvinių taškinių vaizdų duomenys (sutr.ETD) (ang. *spatial Point Pattern Data*), o patį procesą vadinsime erdviu taškiniu procesu (sutr. ETP) (ang. *spatial Point Process*).

ETP analizės metodai

Kadangi ETP yra palyginus neseniai pradėti mokslininkų tirti, todėl nėra dar sukurtos bendros metodologijos, pagal kurią būtų galima juos vertinti. Tačiau įvairiose užsienio publikacijose, moksliniuose darbuose randami pasikartojantys metodai leidžia manyti, jog šie metodai yra pagrindas ETP analizėje. Šie metodai apima:

- glaustą statistiką: taikomajai literatūrai įtaką daro specialūs metodai, pagrįsti glaustu statistinio fakto vertinimu (pvz.: vidutinis atstumas nuo taško iki jo artimiausio kaimyno) su labai ne daug statistinės teorijos, kuri pagrįstą šį vertinimą;
- patikrinimą ar nagrinėjamas ETP yra Puasono ETP: taikomojoje literatūroje hipotezės testais nustatoma ar ETP yra visiškai atsitiktinis (t.y. pastovus Puasono ETP);
- modeliavimą: tikrai per praėjusį dešimtmetį pavyko suformuluoti ir pritaikyti realistinius modelius taškinių procesų duomenims. Tačiau modeliavimas nėra baigtas, dar reikia iširti modelių parinkimą, jų pritaikymo „naudą“ bei tinkamumo duomenims patikrinimą. (Baddeley A., 2008, 12 p.)

I.1. Pagrindinės sąvokos

Kaip jau buvo minėta, ETD nurodo įvykių vietas, mūsų tiriamoje srityje. Taškai gali nurodyti medžius, žemės drebėjimo epicentrus, nusikaltimų vietas ir t.t. Jie gali būti išdėstyti dvimatėje erdvėje, žemės paviršiuje arba trimatėje erdvėje. Šiuose taškuose gali būti stebimi kiti požymiai (ang. *covariates*) arba žymės (ang. *marks*). Plačiau apie tokio tipo taškus kalbėsime III skyriuje.

Apibrėšime pagrindines sąvokas, reikalingas ETP analizei.

Apibrėžimas 1. *Atsitiktiniu įvykiu vadinamas bet koks ETP stebinys.*

Apibrėžimas 2. *Bendras taškinis vaizdas (ang. *mapped point pattern*) – tai visi galimi įvykiai srityje S .*

Apibrėžimas 3. *Įvykiai, kurie užfiksuoti stebint S poaibiuose, vadinami imties taškiniu vaizdu (ang. *sampled (spatial) point pattern*).*

Dažnai tai vadinama ETP X realizacijomis, kurias žymėsime $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{S}$, $n \geq 0$. ETP realizacijos – tai įvairūs taškų iš S rinkiniai, kurie dar vadinami konfigūracijomis. Kalbant apie X skirstinį, reiškia, jog yra žinomos visos X realizacijos ir jų tikimybės. Šios realizacijos gali būti įvairių formų.

Apibrėžimas 4. *Stochastinis mechanizmas, kuris generuoja suskaičiuojamą aibę taškų (x_i) srityje $S \subset \mathbf{R}^2$ vadinamas erdviniu taškiniu procesu (ETP). Jį žymėsime X .*

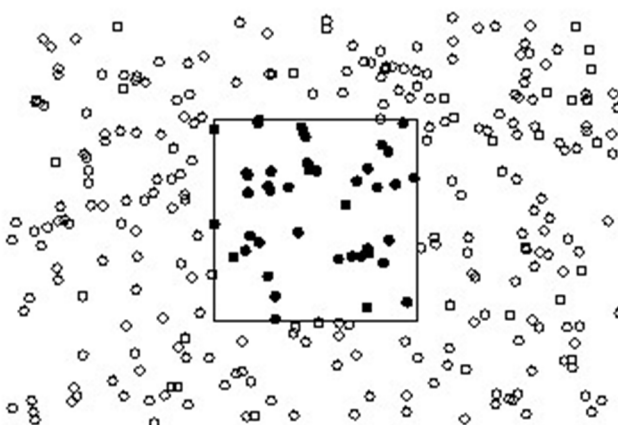
Taip ETP apibrėžia K.Dučinskas. Kiti autoriai ETP apibrėžia kaip atsitiktinį suskaičiuojamą S poaibį. A.Baddeley (2008, 13 p.) ETP aprašo taip: traktuojant taškinį vaizdą kaip taškinį procesą, darome prielaidą, kad vaizdas yra atsitiktinis, taip pat, kad šis vaizdas yra mus dominančio požymio stebėjimas. Taškinių proceso realizacija yra nesutvarkyta taškų aibė, todėl taškai neturi eilės numerio. J.Moller (2007, 7 p.) ETP aprašo panašiai: ETP X yra baigtinis $S \subset \mathbf{R}^2$ poaibis, o tokio ETP X realizacija yra erdvinis taškinis vaizdas $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{S}$, $n \geq 0$. Sakome, kad ETP yra apibrėžtas srityje S ir skaitome, jog $\mathbf{x} = \emptyset$ tuščiam taškiniam vaizdui.

Taškų skaičius, generuojamų ETP, žymimas $n(X)$. Tai yra atsitiktinis dydis ir galima skaičiuoti jo visas tikimybinės charakteristikas. Taškų, generuojamų ETP ir patenkančių į $B \subseteq S$, skaičius (tai taip pat yra atsitiktinis dydis) žymimas $N(B) = n(X_B)$, kur $X_B = X \cap B$. Praktikoje šis atsitiktinis dydis dažnai naudojamas kaip sąlyga sąlyginiuose skirstiniuose.

Apibrėžimas 5. *Kiekvienas ETP X taškas u gali turėti papildomos informacijos. Tokia informacija yra vadinama žyme, žymima m_u . ETP $\Phi = \{(u, m_u) : u \in X\}$ vadinamas žymėtu taškiniu procesu.*

Prielaida apie taškų atrankos langą W

ETP sritis S yra vadinama ETP atrankos langų. Šis langas gali būti įvairiausių formų, nuo taisyklingo stačiakampio iki rombo ar žvaigždės formos. ETP turi turėti langą, kuriame šie duomenys yra apibrėžti. „Standartinis modelis“ sako, kad taškinis procesas X , guli dvimatėje erdvėje, bet yra laikomas tikrai srities W viduje. Mūsų duomenys susideda iš nesutvarkytos aibės $x = \{x_1, \dots, x_n\}$, $x_i \in W, n \geq 0$ taškų x_i iš W . W yra fiksuotas ir žinomas. Svarbu žinoti langą W , kadangi turime žinoti, kurie taškai nėra stebėjimų objektas. Net paprastas taškų tankio įvertinimas priklauso nuo lango. Būtų neteisinga analizuoti taškinį procesą spėjant tinkamą langą. Netiesiogiai manoma, kad visi ETP X taškai lango W viduje buvo pažymėti be praleidimo. Dauguma modelių priima, kad atsitiktiniai taškai galbūt buvo priimti bet kokioje lango W vietoje be tolimesnio apribojimo. Galvojant apie metodologines problemas dažnai naudinga galvoti apie diskretišką taškinio proceso versiją. Tarkime, kad langas W yra suskaldytas į be galo daug „vaizdo elementų“. Kiekvienam vaizdo elementui skiriama vertė $I = 1$, jei jis turi savyje tašką X ir kitu atveju $I = 0$. Šis 0 ir 1 rinkinys sudaro duomenis, kurie turi būti modeliuojami.



4 pav. ETP taškai iš atrankos lango W

Nulinė ETP hipotezė

Pirmasis žingsnis analizuojant ETP yra patvirtinti arba paneigti nulinę hipotezę:

H_0 : ETP yra homogeninis Puasono procesas.

Atmetus nulinę hipotezę yra priimama alternatyvi hipotezė:

H_1 : ETP nėra homogeninis Puasono procesas (procesu duomenims būdingi klasteriai arba išretinimas).

Homogeninis Puasono procesas plačiau aprašytas I.2. poskyryje.

Nulinės hipotezės priėmimui/atmetimui yra naudojami šie metodai:

- a) intensyvumo (tankio) tyrimas;

b) pilno erdvinio atsitiktinumo nustatymas.

Intensyvumas yra vidutinis taškų tankumas (tikėtinas taškų skaičius srities vienetė). Intensyvumas gali būti pastovus (tuomet ETP yra homogeninis) arba gali priklausyti nuo stebimų taškų padėties srityje, būti nepastovus (tuomet ETP yra nehomogeninis).

ETP X vadinamas homogeniniu, jeigu jo skirstinys yra invariantiškas atžvilgiu poslinkių erdvėje R^2 .

K.Dučinsko ir J.Šaltytės-Benth knygoje „Erdvinė statistika“ (2003, 12 p.) griežtas homogeniškumas apibrėžiamas taip: atsitiktinis laukas $\{X(s): s \in D\}$ vadinamas griežtai stacionariu (griežtai homogenišku), jei bendras stebėjimų $(X(s_1), X(s_2), \dots, X(s_3))$ skirstinys yra toks pat kaip ir $(X(s_1 + h), X(s_2 + h), \dots, X(s_3 + h))$ skirstinys visiems baigtiniams n ir visiems s_1, s_2, \dots, s_n, h bei $s_1 + h, s_2 + h, \dots, s_n + h$ iš D , arba kitaip, kai visi baigtiniamieji skirstiniai yra invariantiški koordinatinių pradžios erdvėje atžvilgiu.

Praktikoje dažniausiai taikomos skaitinės charakteristikos yra vidurkiai ir kovariacijos. Todėl tikslinga nagrinėti platesnę nei griežtai stacionarių atsitiktinių laukų klasę.

Apibrėžimas 6. Atsitiktinis laukas $\{X(s): s \in D\}$ vadinamas stacionariu, jei:

$$E\{|X(s)|^2\} < \infty \text{ visiems } s \in D,$$

$$E\{X(s)\} \equiv \mu \text{ visiems } s \in D,$$

$$C(s_1, s_2) = C(s_1 - s_2) \text{ visiems } s_1, s_2 \in D.$$

Apibrėžimas 7. ETP X vadinamas izotropišku, jeigu jo skirstinys yra invariantiškas atžvilgiu pasukimo apie erdvės R^2 pradžios tašką.

ETP stacionarumas ir izotropiškumas yra svarbios homogeninio proceso prielaidos. Šios prielaidos buvo pradėtos tirti kartu tik vėliausioje ETP literatūroje, kur buvo tiriami mažų erdvių taškiniai procesai.

Intensyvumo funkciją galima interpretuoti kaip tikimybę įvykio, kad lygiai 1 taškas pateks į „be galo mažą“ sritį su centru taške u ir plotu du .

Kiekvienam srities S poaibiui B galime apibrėžti skaičiuojantį kintamąjį $N(B)$, kuris yra atsitiktinis dydis. Jo vidurkis apibrėžiamas

$$\mu(B) = EN(B), B \subset S \quad (1)$$

Vidurkio struktūra dažniausiai susiejama su neneigiama intensyvumo funkcija ρ , t.y.
$$\mu(B) = \int_B \rho(u) du.$$

Kovariacijų struktūra skaičiuojamiems atsitiktiniams dydžiams apibrėžiama antros eilės faktorialiniu momentu μ^2 , kuris apibrėžiamas tokiu būdu:

$$\mu^{(2)}(A) = E \sum_{u,v \in X} \mathbf{1}[(u, v) \in A] \quad (2)$$

čia

$$A \subseteq R^2 \times R^2, u \neq v;$$

$\mathbf{1}[\cdot]$ yra indikatorinė funkcija.

Dažnai μ^2 išreiškiama per žinomus antros eilės tankius $\rho^{(2)}$:

$$\mu^{(2)}(A) = \int \mathbf{1}[(u, v) \in A] \rho^2(u, v) du dv \quad (3)$$

Šis tankis gali būti interpretuojamas kaip tikimybė įvykio, kad 1 taškas pateks į „be galo mažą“ sritį su centru u ir plotu du ir 1 taškas pateks į „be galo mažą“ sritį su centru v ir plotu dv .

Apibrėžimas 8. Koreliacijų funkcija visiems $u, v \in S$, kuriems $\rho(u) \neq 0$ ir $\rho(v) \neq 0$ apibrėžiama:

$$g(u, v) = \frac{\rho^2(u, v)}{\rho(u) \cdot \rho(v)} \quad (4)$$

Jeigu taškai yra nepriklausomi vienas nuo kito, tuomet $\rho^{(2)}(u, v) = \rho(u) \cdot \rho(v)$, t.y. $g(u, v) = 1$.

Jeigu $g(u, v) > 1$, tuomet sakome, kad tarp taškų u, v yra trauka.

Jeigu $g(u, v) < 1$, tuomet sakome, kad tarp taškų u, v yra stūma.

Apibrėžimas 9. Taškinis procesas X vadinamas homogeniniu, jeigu bet kokiam dvimatės erdvės poaibiui A , tikėtinas taškų skaičius poaibyje A yra proporcingas A plotui:

$$E[N(X \cap A)] = \lambda |A| \quad (5)$$

Proporcingumo konstanta λ yra intensyvumas. Jei žinoma, kad taškinis procesas yra homogeninis, tada empirinis taškų tankis:

$$\hat{\lambda} = \frac{n(x)}{|A|} \quad (6)$$

yra nepaslinktas nežinomo parametro λ įvertis.

Esant nepastoviam intensyvumui, bendru atveju taškinio proceso intensyvumas, einant nuo vienos vietos iki kitos, kinta. Tarkime, tikėtinas taškų skaičius, patenkantis į mažą teritoriją du apie lokaciją u yra lygus $\lambda(u)du$. Tada $\lambda(u)$ yra proceso intensyvumo funkcija:

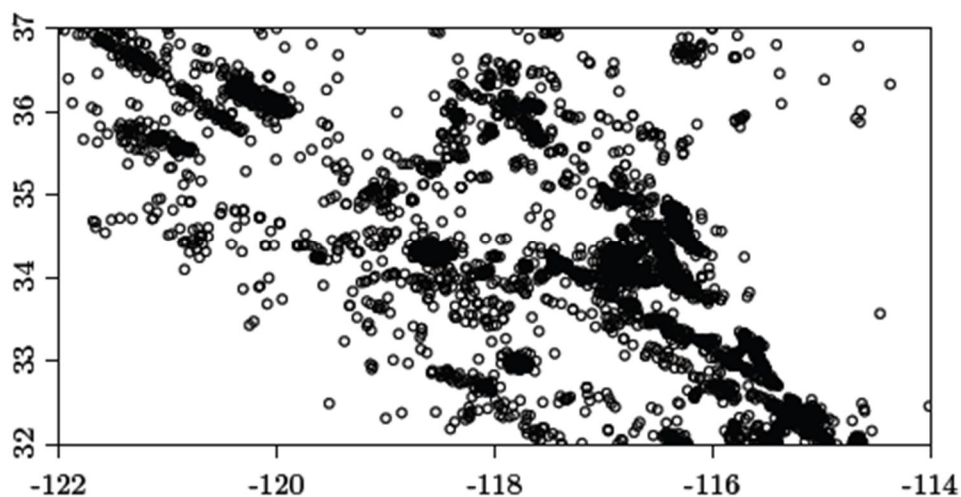
$$E[N(X \cap A)] = \int_A \lambda(u) du \quad (7)$$

visoms sritims A .

Jeigu įtariama, kad intensyvumas yra nepastovus, gali būti tiriama anizotropija. Informacijos apie taškams būdingą anizotropiją šaltinis buvo modulio „Geostatistika ir GIS“ paskaitų medžiaga.

Apibrėžimas 10. Sakome, kad ETD būdinga anizotropija, kai autokoreliacija viena kryptimi yra didesnė negu bet kuria kita kryptimi (nėra vienoda visomis kryptimis).

Anizotropija gali būti tiriama klasteriniams procesams, kuomet norima iširti ar nėra tam tikro taškų klasterių paplitimo viena ar kita kryptimi. Tokių duomenų pavyzdžiai gali būti augalų augimo vietos dėl žemės tam tikrų savybių, taip susidaro klasteriai, o dėl tam tikros vėjo krypties, augalų augimo vietoms gali būti būdinga anizotropija. Pavyzdys galėtų būti 6922 žemės drebėjimų epicentru vietos Los Andžele, užfiksuotos nuo 1984 m. sausio 1 d. iki 2004 m. birželio 17 d.



5 pav. Žemės drebėjimų epicentrai Los Andžele. Stebėjimų langas imtas nuo 32 iki 37 šiaurės platumos ir nuo -122 iki -114 vakarų ilgumos

Anizotropijos tyrimui atvirojo kodo programoje R yra naudojamas statistinis paketas *gstat*.

I.2. Puasono taškinis procesas

Kaip jau buvo minėta, bazinis taškinio proceso modelis yra homogeninis Puasono taškinis procesas plokštumoje su intensyvumu λ . Toks procesas vadinamas „pilnu erdvinio atsitiktinumumu“ (ang. *complete spatial randomness (CSR)*).

Jo pagrindinės ypatybės yra:

- taškų skaičius, gulinčių bet kokioje srityje A , turi Puasono skirstinį su vidurkiu $\lambda|A|$;
- tarkime, kad srityje A yra n taškų, šių taškų vietos yra nepriklausomai ir vienodai paskirsčiusios srityje A ;
- dviejų nesusikertančių sričių A ir B taškų aibės yra nepriklausomos.

Apibrėžimas 11. ETP vadinamas Puasono ETP su vidurkiu μ ir intensyvumo funkcija ρ , jeigu kiekvienai aprėžtai sričiai $B \subseteq S$ su $\mu(B) > 0$ galioja šios sąlygos:

- $N(B)$ yra atsitiktinis dydis pasiskirstęs pagal Puasono dėsnį su vidurkiu $\mu(B)$;
- taškai X_B su sąlyga $N(B)$ yra nepriklausomi atsitiktiniai dydžiai su tankiu proporcingu $\rho(u)$, $u \in B$.

Jeigu $\rho(u) = \rho, \forall u \in S$, tai sakome, kad Puasono ETP yra homogeninis.

Homogeninis Puasono procesas, kai intensyvumas $\lambda > 0$, turi ypatybes (A.Baddeley, 2008, 72 p.):

(PP1): skaičius $N(X \cap B)$ taškų, gulinčių bet kokioje srityje B yra Puasono atsitiktinis kintamasis;

(PP2): tikėtinas taškų skaičius, gulinčių B , yra $E[N(X \cap B)] = \lambda|B|$;

(PP3): jei B_1, B_2 yra nesusieti komplektai tada $N(X \cap B_1)$ ir $N(X \cap B_2)$ yra nepriklausomi atsitiktiniai kintamieji;

(PP4): duota, kad $N(X \cap B) = n$, turime, kad n taškai yra nepriklausomi ir vienodai pasiskirstę srityje B .

(PP2) ir (PP3) yra pakankamos sąlygos.

Nehomogeninis Puasono procesas su intensyvumo funkcija $\lambda(u), u \in R^2$ yra modifikuotas homogeninis Puasono procesas, kuriame ypatybės (PP2) ir (PP4) pakeistos:

(PP2'): taškų $N(X \cap B)$, esančių srityje B , tikėtinas skaičius yra

$$E[N(X \cap B)] = \int_B \lambda(u) du \quad (8)$$

(PP4'): duota, kad $N(X \cap B) = n$, n taškai yra nepriklausomi ir vienodai pasiskirstę, su tikimybinio tankiu

$$f(u) = \lambda(u)/I \quad (9)$$

kur

$$I = \int_B \lambda(u) du \quad (10)$$

Logaritminis tikėtinas homogeninis Puasono procesas su intensyvumu λ yra:

$$\log L(\lambda; x) = n(x) \log \lambda - \lambda |W| \quad (11)$$

kur $n(x)$ yra taškų skaičius rinkinyje x . Maksimalus tikėtinas λ įvertinys yra

$$\hat{\lambda} = \frac{n(x)}{|W|} \quad (12)$$

kuris yra taip pat nešališkas įvertinys. $\hat{\lambda}$ variacija yra

$$\text{var}[\hat{\lambda}] = \frac{\lambda}{|W|} \quad (13)$$

Nehomogeninio Puasono proceso su intensyvumo funkcija $\lambda_\theta(u)$, priklausančio nuo parametro θ , logaritminis θ tikėtinumas bus

$$\log L(\theta; x) = \sum_{i=1}^n \log \lambda_\theta(x_i) - \int_W \lambda_\theta(u) du \quad (14)$$

Deja, bendru atveju rasti analitinį maksimalaus tikėtimumo įvertį parametru θ yra neįmanoma. Vienas iš variantų yra taikyti Berman-Turner algoritmą. Šiuo atveju intensyvumo funkcija $\lambda_\theta(u)$ turi būti logtiesinė parametro θ atžvilgiu:

$$\log \lambda_\theta = \theta \cdot S(u) \quad (15)$$

kur $S(u)$ yra reali vietos u funkcija arba vektorius. Praktikoje $S(u)$ gali būti u erdvinės koordinatės, stebėtos kovariacijos arba abiejų mišinys.

I.3. Erdvinio taškinio proceso analizės technikos

Visą ETP analizę galima suskaidyti į atskirus etapus:

- 1) apžvalga;
 - 2) intensyvumo tyrimas;
 - 3) pilno erdvinio atsitiktinumą nustatymas;
 - 4) tankio funkcijos parinkimas procesui;
 - 5) ETP modelio glodinimas.
- } H_0 tikrinimas

} Modeliavimas

Pirmasis ETP analizės žingsnis yra proceso apžvalga. Galime procesą atvaizduoti grafiškai, patikrinti ar nėra besidubliuojančių taškų, jei tai žymėtas taškinis procesas, suskaidyti jį į subprocesus tam, kad geriau pamatyti atskirų žymių pasiskirstymą ir t.t.

Kitas žingsnis yra intensyvumo tyrimas. Intensyvumo funkcija gali būti įvertinta neparametriškai, naudojant kvadratų skaičiavimo arba branduolinio glodinimo (Kernel'io lyginimas) technikas.

Kvadratų skaičiavimo technika

Langas W padalijamas į lygius subregionus (kvadratus) A_1, A_2, \dots, A_m . Skaičiuojame taškus, patenkančius į kiekvieną kvadratą $n_j = n(x \cap A_j)$, $j = 1 \dots m$. Tai ir yra nepaslinkti atitinkamo intensyvumo mato $\Lambda(A_j)$ įverčiai.

Branduolio glodinimo technika

Dažniausiai naudojamas branduolinis intensyvumo funkcijos įvertis yra:

$$\tilde{\lambda}(u) = e(u) \sum_{i=1}^n k(u - x_i) \quad (16)$$

kur $k(u)$ yra branduolio funkcija, o

$$e(u)^{-1} = \int_W k(u - v) dv \quad (17)$$

vadinama poslinkio korekcija.

Taigi $\tilde{\lambda}(u)$ yra nepaslinktas priglodinto tikrojo intensyvumo funkcijos $\lambda(u)$ įvertis:

$$\tilde{\lambda}(u) = e(u) \int_W k(u - v) \lambda(v) dv \quad (18)$$

Tuo atveju, kai pastebimi keistos formos intensyvumai (pvz. žemės drebėjimų epicentra bus susikcentravę apie lūžių linijas), intensyvumo funkcija gali neegzistuoti. Tokiu atveju kalbame apie intensyvumo matą Λ , apibrėžiamą tokiu būdu:

$$\Lambda(A) = E[N(X \cap A)] \quad (19)$$

visiems $A \subset R^2$.

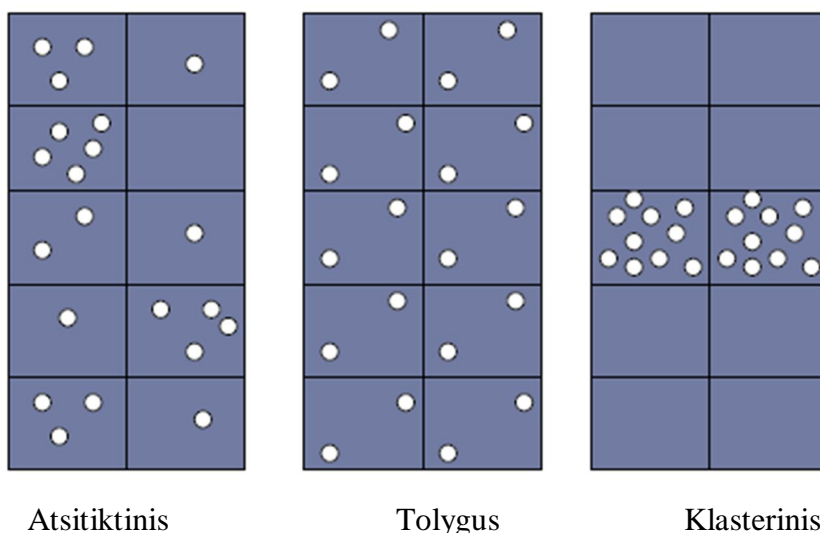
Jei intensyvumas nustatomas pastovus, iš to dar negalima daryti išvados, jog procesas yra Puasono procesas. Alternatyva yra ištirti taškų priklausomumą.

Taškų priklausomumui ištirti yra naudojamas atstumų tarp taškų skaičiavimo metodas, kurio pagalba galime nustatyti ar duomenys yra pasiskirstę atsitiktinai, tolygiai ar būdingi klasteriai.

Atstumų funkcijos

ETD pasiskirstymas gali būti:

- atsitiktinis (ang. *random*) – visi taškai turi vienodą galimybę atsirasti betkuriame vietoje ir ta vieta nėra paveikiama kitų taškų;
- tolygus, vienodas (ang. *uniform*) – kiekvienas taškas yra nutolęs nuo savo visų kaimynų kiek įmanoma didesniu atstumu;
- klasterinis (ang. *clustered*) – taškai susikcentravę į grupes ir yra daug tuščių erdvių, kuriose taškų labai mažai arba išvis nėra.



6 pav. ETD pasiskirstymai

Klasikinis būdas tarptaškinei priklausomybei ištirti yra atlikti atstumų tarp taškų tyrimą. Tuomet yra vertinami:

- porinis atstumas $s_{ij} = \|x_i - x_j\|$;
- artimiausio kaimyno atstumas $t_i = \min_{j \neq i} s_{ij}$;
- tuščių erdvių atstumas $d(u) = \min_i \|u - x_i\|$.

Atstumams apskaičiuoti naudojamos funkcijos atitinkamai:

- K funkcija
- G funkcija
- F funkcija

Svarbu prisiminti, kad:

1. Funkcijos K , G ir F yra apibrėžtos ir įvertintos darant prielaidą, kad taškinis procesas yra homogeninis (stacionarus).
2. Šios funkcijos pilnai necharakterizuoja proceso.
3. Jei procesas nėra homogeninis, nukrypimai tarp empirinių ir teoriškų funkcijų yra ne būtinai tarptaškinės sąveikos įrodymas, kadangi jie gali taip pat būti priskiriami intensyvumo pakitimams.

Apibrėžimas 12. Atstumas $s_{ij} = \|x_i - x_j\|$: $x_i, x_j \in X, x_i \neq x_j$ tarp visų skirtingų taškų porų vadinamas poriniu atstumu.

Stacionaraus proceso atveju, funkcija K apibrėžia tikėtiną taškų skaičių atkarpoje r :

$$K(r) = \frac{1}{\lambda} E[n(X \cap b(u, r) \setminus \{u\}) | u \in X] \quad (20)$$

Jos nepaslinktas ir populiariausias įvertis:

$$\hat{K}(r) = \frac{1}{\lambda^2 |W|} \sum_i \sum_{j \neq i} 1\{\|x_i - x_j\| \leq r\} e(x_i, x_j; r) \quad (21)$$

Homogeninio Puasono proceso atveju $K_{pois}(r) = \pi r^2$. Porinių atstumų funkcija K yra dažniausiai naudojama ETP tankio funkcijos įvertinimui.

Apibrėžimas 13. Atstumas $t_i = \min_{j \neq i} s_{ij}$ nuo taško iki jo artimiausio kaimyno vadinamas artimiausio kaimyno atstumu.

Remiantis prielaida, kad X yra stacionarus, galima apibrėžti artimiausių kaimynų atstumo pasiskirstymo funkciją:

$$G(r) = P\{d(u, X \setminus \{u\}) \leq r | u \in X\} \quad (22)$$

Empirinė stebėtų artimiausių kaimynų atstumų pasiskirstymo funkcija ir jos svartinis įvertis apibrėžiami atitinkamai:

$$G^*(r) = \frac{1}{n(x)} \sum_i 1\{t_i \leq r\} \quad (23)$$

$$\hat{G}(r) = \sum_i e(x_i, r) 1\{t_i \leq r\} \quad (24)$$

Homogeninio Puasono proceso atveju $G_{pois}(r) = 1 - \exp(-\lambda \pi r^2)$.

Apibrėžimas 14. Atstumas $d(u, x) = \min\{\|u - x_i\| : x_i \in X\}$ nuo fiksuotos vietos $u \in R^2$ iki artimiausio X taško vadinamas tuščios erdvės atstumu.

Jei ETP X yra stacionarus, tokiu atveju galima apibrėžti kumuliatyvinę tušių erdvių atstumų pasiskirstymo funkciją

$$F(r) = P\{d(u, X) \leq r\} \quad (25)$$

Empirinė stebėtų tuščių erdvių vietose $u_j, j = 1, \dots, m$ gardelės pasiskirstymo funkcija

$$F^*(r) = \frac{1}{m} \sum_j \mathbf{1}\{d(u_j, \mathbf{x}) \leq r\} \quad (26)$$

yra neigiamai pasislinkęs įvertis. Todėl praktikoje yra taikomos įvairios šio įverčio korekcijos. Populiariausias įvertis:

$$\hat{F}(r) = \sum_j e(u_j, r) \mathbf{1}\{d(u_j, \mathbf{x}) \leq r\} \quad (27)$$

kur $e(u_j, r)$ yra svoris įvedamas tokiu būdu, kad įvertis būtų nepaslinktas.

Homogeninio Puasono proceso su intensyvumu λ atveju, taškų patenkančių į $b(u, r)$ kiekis yra Puasono su vidurkiu $\mu = |\lambda|(b(u, r)) = \lambda\pi r^2$, todėl tikimybė, kad atsitiktiniame regione nėra taškų bus $\exp(-\mu) = \exp(-\lambda\pi r^2)$. Taigi Puasono proceso atveju

$$F_{pois}(r) = 1 - \exp(-\lambda\pi r^2) \quad (28)$$

Realiai mes lyginame $\hat{F}(r)$ su $F_{pois}(r)$. Jeigu $\hat{F}(r) > F_{pois}(r)$, galime daryti išvadą, kad atstumai yra mažesni nei Puasono proceso atveju; jeigu $\hat{F}(r) < F_{pois}(r)$ - duomenims būdingi klasteriai.

Kartais iškyla problemų darant išvadas, nes funkcijos G , K ir F gali duoti skirtingus rezultatus. Dažnai naudojama K funkcijos transformacija yra L funkcija:

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \quad (29)$$

kuri transformuoja Puasono K funkciją į tiesę $L_{pois}(r) = r$. Kovariacinės šaknies traukimas stabilizuoja dispersiją ir išvados būna tikslesnės. L funkcijos rezultatų interpretacija yra tokia pati kaip ir K funkcijos.

Kita apibendrinanti funkcija yra porinė koreliacinė funkcija $g(r)$:

$$g(r) = \frac{K'(r)}{2\pi r} \quad (30)$$

Įprasta funkcijų F ir G kombinacija yra J funkcija, apibrėžta visiems $r \geq 0$ ir $F(r) < 1$.

$$J(r) = \frac{1-G(r)}{1-F(r)} \quad (31)$$

Homogeninio Puasono proceso atveju $J_{pois}(r) \equiv 1$.

Pilnas erdvinis atsitiktinumas ir jo nustatymo testai

I.2. poskyryje aptarėme, jog bazinis taškinio proceso modelis yra homogeninis Puasono taškinis procesas plokštumoje su intensyvumu λ . Jam būdingas pilnas erdvinis atsitiktinumas

(CSR). Esant (CSR), taškai yra nepriklausomi ir turi tą patį pasiskirstymą visoje srityje. Pilno erdvinio atsitiktinumo tyrimui naudojama keletas testų.

χ^2 testas

Klasikinis testas nulinei (CSR) hipotezei patikrinti yra χ^2 testas. Kaip žinome iš anksčiau, langas W yra padalytas į lygias subsritis („kvadratus“) B_1, \dots, B_m . Skaičiuojame taškų skaičių, gulinčių kiekviename kvadrato, $n_j = n(x \cap B_j)$, $j = 1, \dots, m$. Esant nulinei (CSR) hipotezei, n_j yra tolygiai ir nepriklausomai pasiskirstę Puasono atsitiktiniai dydžiai su ta pačia tikėtina verte. Tokiu atveju yra taikomas Pirsono χ^2 „atitikimo gerumo“ testas.

Pirsono χ^2 testo statistinis faktas sutampa iki pastovaus faktoriaus:

$$\chi^2 = \frac{\sum_j (n_j - \frac{n}{m})^2}{n/m} \quad (32)$$

čia

$n = \sum_j n_j$ yra taškų bendras skaičius;

m – kvadratų skaičius.

Kolmogorov-Smirnov testas

Daug stipresnis (CSR) testas yra *Kolmogorov-Smirnov* testas, kuriame lyginame turimą ir tikėtiną funkcijos T verčių išsibarstymą. Apibrėžiame realiai įvertintą funkciją $T(x,y)$, apibrėžtą visose lango vietose (x,y) . Įvertiname šią funkciją kiekviename iš duotų taškų. Tada lyginame funkcijos T reikšmių empirinį išsibarstymą su tikėtinu funkcijos T išsibarsymu (CSR).

Monte Carlo testas

Pilnas erdvinis atsitiktinumas taip pat gali būti įvertintas *Monte Carlo* testo pagalba. Testas remiasi simuliacijomis pagal nulinę hipotezę. Pats principas buvo sukurtas Barnard'o ir Dwass'o. Testas buvo pritaikytas erdvinėje statistikoje Ripley'o ir Besag'o. *Monte Carlo* testai yra atsitiktinių testų ypatingieji atvejai, kurie yra paprastai naudojami nparametrinėje statistikoje. Tarkime, atskaitos kreivė yra teorinė K funkcijos kreivė pilno erdvinio atsitiktinumo atveju. Tuomet yra generuojama M nepriklausomų pilno erdvinio atsitiktinumo taškinių vaizdų nagrinėjamame regione W . Kiekvienai realizacijai skaičiuojame K funkcijos įvertį $\widehat{K}^j(r)$, $j = 1, \dots, M$. Apibrėžiame apatinę ir viršutinę sugeneruotų kreivių ribas (t.y. taškinės (ang. *pointwise*) kreivės):

$$L(r) = \min_j \widehat{K}^j(r)$$

$$U(r) = \max_j \widehat{K}^j(r)$$

Bet kokiai fiksuotai reikšmei r , generuotoms kreivėms, apibrėžiame tikimybę, kad $\widehat{K}(r)$ yra už intervalo $[L(r), U(r)]$ ribų. Jeigu duomenys pasiskirstę pagal tolygų Puasono dėsnį, tai $\widehat{K}(r)$ ir $\widehat{K}^1(r), \dots, \widehat{K}^M(r)$ yra statistiškai ekvivalentūs ir nepriklausomi, todėl atitinkama tikimybė yra $\frac{2}{(M+1)}$.

Vadinasi testo, kuris atmeta nulinę hipotezę reikšmingumo lygmuo yra $\alpha = \frac{2}{(M+1)}$.

Kartais vietoje pataškinio minimumo ir maksimumo galima naudoti $k -$ tąją didžiausią ir $k -$ tąją mažiausią reikšmes, tada reikšmingumo lygmuo būtų $\alpha = \frac{2k}{(M+1)}$.

Aparto *Monte Carlo* testo problema yra ta, kad sprendimas priklauso nuo r , kurį mes patys turime pasirinkti iš anksto. Norint išspręsti šią problemą, galima sukonstruoti apibendrintas kritines ribas, kurios pasižymi savybe, kad teisingos nulinės hipotezės atveju, tikimybė, kad $\widehat{K}(r)$ yra už šių ribų yra lygiai 5%. Paprasčiausias būdas tai padaryti yra kiekvienam įverčiui $\widehat{K}(r)$ apskaičiuoti maksimalų nuokrypį nuo Puasono K funkcijos: $D = \max_r |\widehat{K}(r) - K_{pois}(r)|$.

Šis dydis apskaičiuojamas kiekvienai sugeneruotai duomenų aibe, tokiu būdu nustatant D_{max} . Tada apatinė ir viršutinė ribos yra

$$L(r) = \pi r^2 - D_{max}$$

$$U(r) = \pi r^2 + D_{max}$$

Įvertinta K funkcija peržengia šias ribas tik tuo atveju, jeigu D reikšmė peržengia D_{max} . Esant teisingai nulinei hipotezei, taip atsitinka su tikimybe $\frac{1}{M+1}$.

Taigi, siekiant reikšmingumo lygmens $\alpha = 0,05$, mažiausias simuliacijų skaičius $M = 19$. Iš tikrųjų, *Monte Carlo* testas gali būti pritaikytas bet kokiam taškinio proceso modeliui, tenkinančiam nulinę hipotezę. Tiesiog turime sukurti realizavimo imitaciją su nuline hipoteze ir apskaičiuoti apibendrinančią funkciją kiekvienai realizavimo imitacijai. (Baddeley A., 2008, 120 p.)

ETP modeliavimas

Ištyrus proceso intensyvumą ir pilną erdvinį atsitiktinumą bei priėmus ar atmetus nulinę hipotezę, kitas žingsnis ETP analizėje yra geriausios tankio funkcijos parinkimas bei ETP modelio priglodinimas procesui. Šis analizės etapas dar vadinamas „modeliavimu“. Procesui yra pritaikoma keletas skirtingų tankio funkcijų ir parenkama geriausiai tinkanti. Tankio funkcijų tinkamumas erdviniams taškiniams duomenims gali būti įvertintas remiantis Akaikės informacijos kriterijumi (AIK), paklaidų analizės pagalba ar minimalaus kontrasto metodu.

Parinkus geriausią tankio funkciją, tiriamam ETP yra glodinamas tam tikras teorinis ETP modelis. Modelis parenkamas pagal tai, kokia tankio funkcija buvo nustatyta, kaip geriausiai aprašanti proceso intensyvumą.

Paklaidų analizė

Tiriamam ETP, su įvertintu tankiu $\hat{\lambda}(u)$, prognozuojamas taškų skaičius, patenkantis į bet kokią sritį B yra $\int_B \hat{\lambda}(u) du$. Taigi paklaidos srityje B gali būti apskaičiuojamos:

$$R(B) = n(x \cap B) - \int_B \hat{\lambda}(u) du \quad (33)$$

Atvirojo kodo programoje R, naudojant *spatstat* paketą, galima apskaičiuoti ir grafiškai atvaizduoti suglodintas paklaidas bei paklaidų keturių tipų grafikus, iš kurių galima daryti išvadas apie geriausiai tinkamą tankio funkciją erdviniams taškiniais duomenims.

Minimalaus kontrasto metodas

Klasikinėje statistikoje gerai žinomas momentų metodas, kai nežinomą parametą θ galima įvertinti sulyginus empirinį analogą su teoriniu:

$$E_{\theta}[A(x)] = A(x) \quad (34)$$

čia $A(x)$ yra stebėtiems duomenims įvertinta funkcija, o kairėje pusėje – nagrinėjamos funkcijos teorinis vidurkis. Analogas taškiniuose procesuose yra K funkcijos sulyginimas su jos teorine reikšme.

Tai galima padaryti minimizuojant nagrinėjamų funkcijų skirtumą intervale $[a;b]$:

$$D(\theta) = \int_a^b |\hat{K}(r)^q - K_{\theta}(r)^q|^p dr \quad (35)$$

čia

$$0 < a < b,$$

p, q yra indeksai.

Šis metodas yra žinomas kaip minimalaus kontrasto metodas. Dažniausiai taikomas duomenims, kuriems būdingi klasteriai.

I.4. Kiti erdviniai taškiniai procesai

Be Puasono homogeninių ar nehomogeninių erdviųjų taškinių procesų, galima išskirti atskiras ETP rūšis:

- išretinti ETP
- klasteriniai ETP
- žymėti ETP
- Gibbs'o ETP

Pagal šias rūšis yra sudaromi ETP modeliai (pvz. Puasono ETP modelis, Gibbs'o ETP modelis ir t.t.).

„Retinimas“ reiškia kai kurių taškų pašalinimą iš taškinio proceso. Nepriklausomas retinimas – kai taško išlikimas yra nepriklausomas nuo kitų taškų. Kai yra taikomas nepriklausomas retinimas

Puasono procesui, rezultate gauname išsaugotų taškų procesą, kuris taip pat yra Puasono tipo. Norint gauti ne Puasono procesą, reikia naudoti tam tikrą priklausomą retinimo funkciją.

Klasterinius procesus plačiau apžvelgsime II skyriuje, o žymėtus procesus plačiau aptarsime III skyriuje.

Atskira taškinių procesų rūšis yra tikimybiniai erdviniai taškiniai procesai, juos trumpai apžvelgsime.

Gibbs'o taškiniai procesai

Vienas iš būdų kaip galima sukonstruoti statistinį modelį, tai yra užrašyti jo tikimybinį tankį. Tokio užrašymo privalumai yra:

- funkcinė tankio forma atspindi jo tikimybinės savybes;
- tankio funkcijos nariai ir koeficientai dažnai gali būti interpretuojami kaip modelio komponentai;
- lengva įvesti narius, kurie nusako modelio priklausomybę nuo kovariacijų ir pan.

Žinoma, šis metodas gali būti taikomas tik tuo atveju, kai tankio funkciją įmanoma užrašyti ir ji yra interpretuotina. Erdviniai taškiniai vaizdai, sudaryti remiantis jų tikimybinio tankiu yra vadinami Gibbs'o procesais.

Tikimybinis tankis bus funkcija $f(x)$ apibrėžta kiekvienam taškų $x = (x_1, \dots, x_n)$ rinkiniui iš apibrėžto lango W ($x_i \in W$), bet kokiam $n \geq 0$. Taškų kiekis n nėra apibrėžtas ir gali būti lygus nuliui. Tačiau bendru atveju tikimybinis taškinių proceso tankis yra interpretuojamas kaip paprasčiausio duomenų rinkinio tikimybinis tankis. Taškinis procesas X lange W tikimybinį tankį turi tik tokiu atveju, jei bet kokiai neneigiamai integruojamai funkcijai h :

$$E[h(X)] = e^{-|W|}h(\emptyset)f(\emptyset) + e^{-|W|} \sum_{n=1}^{\infty} \frac{1}{n!} \int_W \dots \int_W h(\{x_1, \dots, x_n\})f(\{x_1, \dots, x_n\}) dx_1 \dots dx_n \quad (36)$$

Taigi, tikimybė, kad X yra sudarytas iš lygiai n taškų, bus lygi:

$$p_n = P\{n(X) = n\} = \frac{e^{-|W|}}{n!} \int_W \dots \int_W h(\{x_1, \dots, x_n\})f(\{x_1, \dots, x_n\}) dx_1 \dots dx_n \quad (37)$$

kai $n \geq 1$ ir $p_0 = P\{n(X) = 0\} = e^{-|W|}f(\emptyset)$.

Homogeninio Puasono proceso, su intensyvumu lygiu 1, tikimybinis tankis yra $f(x) = 1$. Homogeninio Puasono proceso lange W , su intensyvumu λ , tikimybinis tankis yra $f(x) = \alpha \lambda^{n(x)}$, čia $n(x)$ yra taškų rinkinyje x kiekis, o konstanta $\alpha = e^{(1-\lambda)|W|}$.

Neghomogeninio Puasono proceso lange W su intensyvumo funkcija $\lambda(u)$ tankis:

$$f(x) = \alpha \prod_{i=1}^n \lambda(x_i) \quad (38)$$

kur konstanta $\alpha = \exp \left[\int_W (1 - \lambda(u)) du \right]$.

Pagrindinis įrankis Gibbs'ο taškinio proceso analizėje yra sąlyginis tankis $\lambda(u, X)$. Šis tankis apibrėžia sąlyginę tikimybę, kad rasime vietoje u tašką, žinant visą informaciją apie likusį procesą. Sąlyginė tikimybė, kad šis taškas bus aptiktas lokacijos u kaimynystėje, yra $\lambda(u, X)du$. Taškiniam procesui, apibrėžtam lange W , sąlyginis tankis vietoje u yra susijęs su tikimybinio tankiu tokiu būdu:

$$\lambda(u, x) = f(x \cup \{u\}) / f(x) \quad (39)$$

Homogeninio Puasono proceso su intensyvumu λ sąlyginis tankis $\lambda(u, x) = \lambda$. O nehomogeninio $\lambda(u, x) = \lambda(u)$. Sąlyginis Puasono proceso tankis nepriklauso nuo x konfigūracijos, nes šiuo atveju taškai yra nepriklausomi.

Gibbs'ο ETP gali būti sugeneruoti Markovo grandinių *Monte Carlo* (MCMC) algoritmu pagalba. Šie algoritmai generuoja Markovo grandinę, kurios būsenos yra taškiniai vaizdai. Grandinė yra sudaroma taip, kad pasiskirstymas būtų toks, kokį mes siekiame sugeneruoti. Jeigu grandinė būtų „paleidžiama“ begalinį skaičių kartų, būsenos konverguotų į norimą taškinio proceso pasiskirstymą. Praktikoje grandinė yra „paleidžiama“ daug kartų.

Sąlyginis proceso tankis $\lambda_\theta(u, x)$ turi būti logtiesinis parametro θ atžvilgiu: $\log \lambda_\theta(u, x) = \theta \cdot S(u, x)$. Parametrai θ yra vadinami reguliariais, o visi kiti parametrai nereguliariais.

Lyginant su Puasono modeliu, Gibbs'ο modelį patikrinti yra daug sudėtingiau. Nėra teorijos, leidžiančios pritaikyti χ^2 ar *Kolmogorov – Smirnov* testus. Prognozuojamas vidutinis taškų skaičius tam tikrame regione Gibbs'ο modelio atveju taip pat nėra žinomas. Todėl dažniausiai vertinimas remiasi K ir G funkcijomis. Funkcija *envelope* kaip pirmąjį argumentą gali priimti įvertintą Gibbs'ο modelį ir pagal jį generuos, sudarydama taip vadinamą „voką“. Gibbs'ο proceso paklaidos srityje $B \in R^2$ yra apibrėžiamos tokiu būdu:

$$R(B) = n(x \cap B) - \int_B \hat{\lambda}(u, x) du \quad (40)$$

Jeigu įvertintas modelis yra teisingas, paklaidų vidurkis turi būti lygus nuliui. Grafiškai paklaidas galima analizuoti remiantis lygiai tokiais pačiais metodais, kaip ir Puasono modelio atveju (pvz. sudaryti keturių tipų grafiką).

II. KLAŠTERINIAI TAŠKINIAI PROCESAI

Kaip jau buvo minėta, klasikinis taškinių proceso modelis yra homogeninis Puasono taškinis procesas plokštumoje su intensyvumu λ . Toks procesas dažnai vadinamas pilnu erdviniu atsitiktinumu, tačiau šis Puasono procesas realybėje pasitaiko labai retai. Vis dažniau nustatomi nukrypimai nuo šio „idealaus“ proceso modelio.

Sakoma, kad taškinis procesas, kuris nėra Puasono procesas, parodo „sąveiką“ ar priklausomybę tarp taškų.


Kaip jau buvo aprašyta, taškinių procesų duomenims gali būti būdingi šie pasiskirstymai:

- atsitiktinis,
- tolygus,
- klasterinis.

Pasak mokslininko Andrew David Cliff (1981, 90 p.), kiek taškų pasiskirstymas yra nutolęs nuo atsitiktinio Puasono pasiskirstymo, galima įvertinti skaičiuojant dispersijos indeksą φ . Jis yra naudojamas ištirti nukrypimo stiprumą nuo homogeninio Puasono proceso. Jo išraiška yra tokia:

$$\varphi = \frac{\text{imties dispersija}}{\text{imties vidurkis}}$$

Iš tikrųjų, jei skaičiuosime dispersijos ir vidurkio santykį, pasiskirstymus, nutolusius nuo paprasto Puasono pasiskirstymo, žemiau pateiktoje schemoje (7 pav.), galima išdėstyti taip:

Klasteriniai taškiniai procesai	Neigiamas binominis	($\varphi > 1$)	 Mažėjanti klasterizacija
	Puasono-Paskalio		
	Niumano A tipas		
Atsitiktiniai taškiniai procesai	Paprastas Puasono	($\varphi = 1$)	
Tolygūs taškiniai procesai	Binominis	($\varphi < 1$)	

7 pav. Pasiskirstymų išsidėstymas pagal mažėjančią klasterizaciją

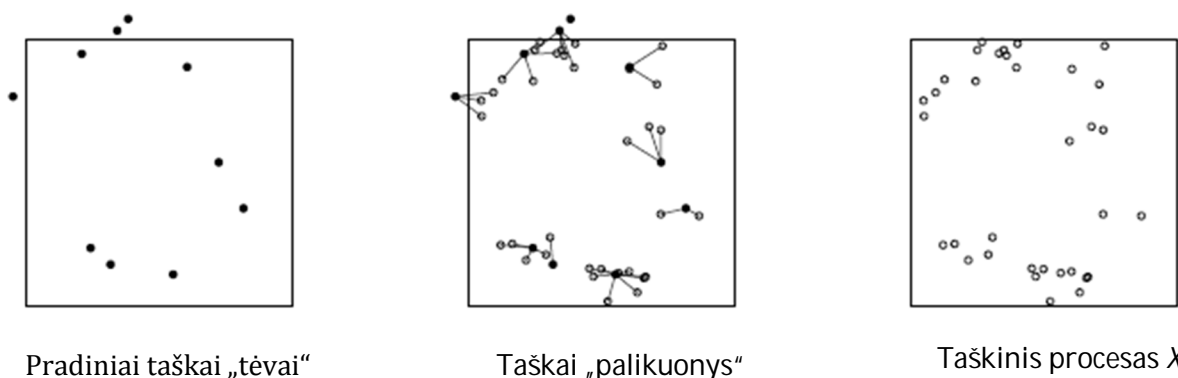
Visi šie pasiskirstymai rodo klasterizacijos pokytį.

Kaip matome, klasterizacijos atveju, dispersija yra didesnė negu vidurkis, kadangi dispersijos indeksas φ yra didesnis už vienetą. Ši situacija dažnai vadinama perviršyta dispersija (ang. *overdispersion*). Puasono proceso atveju dispersija lygi vidurkiui ($\varphi = 1$). Pastaroji schema turėtų būti traktuojama su atsargumu, kadangi ne mažai pasiskirstymų yra įtraukti į dar labiau bendresnes pasiskirstymų šeimas, kurios gali turėti ir klasterizacijos ir tolygumo elementų.

Aptarsime klasterinius procesus, kurie yra gauti iš Puasono proceso ir vis dar išsaugo kai kurias paprastas Puasono proceso ypatybes.

II.1. Puasono grupės klasteriniai procesai

Puasono grupės klasterinius procesus aprašo A.Baddeley (2008, 98 p.). Puasono klasterinis procesas pradamas kurti nuo proceso Y iš „tėvinių“ (pradinių) taškų. Kiekvienas „tėvinis“ taškas $y_i \in Y$ sukuria baigtinę aibę Z_i „palikuonių“ taškų pagal tam tikrą atsitiktinį mechanizmą. Aibė, apimanti visus „palikuonių“ taškus, suformuoja taškinį procesą X . Būtent X ir yra laikomas stebėjimų objektu.



8 pav. Puasono klasterinis procesas

Galime išskirti keletą Puasono grupės klasterinių procesų.

- *Mat'ern* klasterinis procesas, kuriame visų pirma sugeneruojamas Puasono procesas „tėvinių“ taškų su intensyvumu (κ) (*kappa*). Tuomet kiekvienas taškas „tėvas“ yra pakeičiamas atsitiktinių taškų klasteriu, kuris turi Puasono (μ) taškų skaičių ir kur taškai yra nepriklausomai ir tolygiai pasiskirstę r spindulio diske, kurio centras yra taškas „tėvas“.

Realizuojant šį procesą atvirojo kodo programoje R, modelyje turi būti nurodomas Puasono proceso klasterių centrų intensyvumas (κ). Jis gali būti teigiamas skaičius, funkcija arba vaizdo taškų objektas (nuotrauka). Taip pat nurodomas klasterių spindulys (r), vidutinis taškų skaičius klasteriuose (μ) (teigiamas skaičius arba intensyvumo funkcija) bei langas (*win*).

- *Thomas* klasterinis procesas, kuriame visų pirma yra sugeneruojamas vienas Puasono procesas „tėvinių“ taškų su intensyvumu (κ). Tuomet kiekvienas „tėvas“ taškas yra pakeičiamas atsitiktinių taškų klasteriu, kur klasterio taškų pasiskirstymas yra Puasono (μ), o jų pakeitimas yra pasiskirstęs pagal izotropinį Gauso $N(0, \sigma^2 I)$ pasiskirstymą nuo „tėvinio“ taško.

Realizuojant šį procesą atvirojo kodo programoje R, modelyje turi būti nurodomas Puasono proceso klasterių centrų intensyvumas (κ), taškų perkėlimo iš klasterio centro standartinis nuokrypis (σ), tikėtinas skaičius taškų klasteryje (μ) ir langas (*win*).

- *Gauss-Poisson* klasterinis procesas, kuriame visų pirma sugeneruojamas Puasono procesas „tėvinių“ taškų su intensyvumu (κ). Tuomet kiekvienas taškas yra arba išsaugomas (su tikimybe $1-p_2$) arba pakeičiamas taškų pora nustatytame atstume r (su tikimybe p_2). Dviejų taškų klasteriuose, linija, jungianti tuos du taškus, turi vienarūšę atsitiktinę orientaciją.

Realizuojant šį procesą atvirojo kodo programoje R, modelyje turi būti nurodomas Puasono proceso klasterių centrų intensyvumas (κ). Jis gali būti teigiamas skaičius, funkcija arba taškinių vaizdų objektas (nuotrauka). Taip pat nurodomas klasterio, kuris susidaro iš tiksliai 2 taškų, diametras (r), tikimybė, jog klasteris susideda iš 2 taškų (p_2) bei langas (win).

- *Neyman-Scott* klasterinis procesas, kuriame visų pirma sugeneruojamas Puasono procesas „tėvinių“ taškų su intensyvumu (κ), kur (κ) gali būti teigiamas skaičius, funkcija $\kappa(x,y)$ arba taškinių vaizdų objektas klasės „*im*“. Kiekvienas taškas „tėvas“ yra pakeičiamas atsitiktinio klasterio taškais, sukurtais *rcluster* funkcijos pagalba. Šie klasteriai yra apjungiami, kas ir duoda rezultate *Neyman-Scott* procesą.

Realizuojant šį procesą atvirojo kodo programoje R, modelyje turi būti nurodomas Puasono proceso klasterių centrų intensyvumas (κ). Taip pat nurodomas maksimalus atsitiktinio klasterio spindulys ($rmax$), funkcija, kuri generuoja atsitiktinius klasterius (*rcluster*) bei langas (win).

„Cox“ procesai

„Cox“ taškinis procesas yra iš esmės Puasono procesas su atsitiktine intensyvumo funkcija. Tarkime $\Lambda(u)$ yra atsitiktinė funkcija su neneigiamomis reikšmėmis, apibrėžta visose vietose $u \in R^2$. Apibrėžtai Λ tarkime, kad X yra Puasono procesas su intensyvumo funkcija Λ . Toks erdvinis taškinis procesas X yra „Cox“ procesas.

„Cox“ modelis yra „atsitiktinių efektų“ modelio analogas. Jo dispersija yra visada perviršyta, lyginant su Puasono procesu. Kitaip sakant, taškų, krintančių į tam tikrą regioną, skaičiaus dispersija yra didesnė negu vidurkis. Ypač naudinga „Cox“ procesų klasė yra „*log-Gaussian*“ (sutr. *LGCP*), kur $\log\Lambda(u)$ yra Gauso atsitiktinė funkcija.

II.2. Neigiamas binominis procesas

Kai dispersijos indeksas yra didesnis už vienetą, tuomet duomenims yra būdingi klasteriai. Dėl šios priežasties neigiamas binominis pasiskirstymas, kitaip dar vadinamas *Puasono-gamma* pasiskirstymu, tampa labiausiai naudojamu iš visų Puasono pasiskirstymo modifikacijų. Neigiamas binominis pasiskirstymas yra laikomas geriausiu pasiskirstymu, galinčiu valdyti viršijančią dispersiją ir tampa vis plačiau naudojamu tiriant būtent tokio tipo duomenis. Neigiamą binominį pasiskirstymą nagrinėjo Zhang Y. (2006, 3 p.) tirdamas nelaimingų atsitikimų duomenis.

Neigiamas binominis pasiskirstymas turi du parametrus: vidurkį μ ir dispersijos parametą k , kuris dažnai laikomas fiksuotu viršijančios dispersijos matu.

Imties X , kuri pasiskirsčiusi pagal neigiamą binominį pasiskirstymą ($X \sim NB(\mu, k)$), dispersija yra lygi $\sigma^2 = \mu + \mu^2/k$. Tikimybė, jog kintamasis X įgyja reikšmę x yra:

$$Pr[X = x] = \frac{\Gamma(x+k)}{x! \Gamma(k)} \left(\frac{\mu}{\mu+k}\right)^x \left(1 + \frac{\mu}{k}\right)^{-k} = \frac{(x+k-1)(x+k-2)\dots(k+1)k}{x!} \left(\frac{\mu}{\mu+k}\right)^x \left(1 + \frac{\mu}{k}\right)^{-k} \quad (41)$$

čia

$$\mu, k > 0;$$

$$x = 0, 1, 2, \dots;$$

$$\Gamma(\cdot) \text{ žymi gama funkciją } \Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt.$$

Kai kuriuose literatūros šaltiniuose, dispersijos parametras k yra pažymimas $a=1/k$.

Neigiamo binominio pasiskirstymo savybes taip pat aiškina ir Jain S.K. (1958, 2 p.) Apibrėškime $p=\mu/k$. Jeigu k artėja prie ∞ ir $p \rightarrow 0$, tuomet perviršyta dispersija mažėja ir taip pasiskirstymas tam tikru laiku tampa Puasono pasiskirstymu. Kita vertus, jei leisime N (stebėjimų skaičiui) artėti į ∞ ir $k \rightarrow 0$ ir nepaisysime nulio klasės, turėsime Fišerio logaritminės eilutės formą. Trūki neigiamo binominio pasiskirstymo forma gali atsirasti tuomet, kai stebėjimų skaičius nulio klasėje negali būti izoliuotas.

Pagrindinis žingsnis naudojant neigiamą binominį pasiskirstymą yra parametro k įvertinimas. Naudojami keli būdai parametro k įvertinimui:

- a) Momentų metodas. Naudojami pirmieji du momentai \bar{r} ir s^2 . \hat{k} bus priimtinas tuomet, kai $s^2 > \bar{r}$, kadangi $k > 0$.

$$\hat{k} = \bar{r}^2 / (s^2 - \bar{r}) \quad (42)$$

- b) Nulio klasės metodas. Naudojamas nulio klasės dažnis $f_0 = N/q^k$, kur $q = 1 + p$, kai $(1+p)^k = N/f_0$. k įvertis yra gaunamas lygties iteraciniu sprendimu:

$$\hat{k} \log\left(1 + \frac{\bar{r}}{\hat{k}}\right) = \log(N/f_0) \quad (43)$$

- c) Maksimalaus tikėtinumo (ang. *maximum likelihood*, sutr. *ML*) metodas. ML funkcija:

$$l(k, \mu) = \frac{1}{n} \sum_{i=1}^n \log\left\{\frac{\Gamma(x_i+k)}{\Gamma(k)}\right\} + \bar{r} \log(\mu) - (\bar{r} + k) \log\{1 + \mu/k\} \quad (44)$$

Atlikus pakeitimus bei įvedus gradiento elementus galėsime apskaičiuoti maksimalaus tikėtinumo k įvertį.

III. ŽYMĖTI TAŠKINIAI PROCESAI

Visada taškui galima priskirti, pridėti tam tikros papildomos informacijos, pavyzdžiui, medžių augimo vietos gali turėti medžio rūšies žymę (žymė yra kokybinė) ar medžio aukščio matą (žymė yra kiekybinė). Kiekvienas taškas gali turėti tik vienmatę žymę. Taškinius procesus su žymėmis galima analizuoti tiesiog ignoruojant žymes, tuomet bus atlikta paprasta taškinio proceso analizė, neatsižvelgiant į taškų papildomą informaciją, tačiau yra ne mažai įdomių klausimų, kuriuos galima užduoti būtent žymėtiems taškiniams procesams, tokių kaip „ar yra ryšys tarp tam tikrų dviejų, ar daugiau, medžių rūšių?“ ir t.t.

Žymėti taškiniai procesai su kokybinėmis žymėmis yra tokie, kuriuose įvykius - taškus galime sugrupuoti į atskiras grupes pagal jų žymes. Dažniausiai šių skirtingų grupių skaičius būna mažas, t.y. žymės dažniausiai būna dviejų tipų, rečiau pasitaiko taškų su didesniu žymių rūšių skaičiumi.

Pažymėtas taškinis procesas erdvėje S su žymėmis, priklausančiomis aibei M , yra matematiškai apibrėžtas kaip taškinis procesas Dekarto sandaugoje $S \times M$. Galimų žymių erdvė M gali būti bet kas. Dažniausiai žymė yra kategoriškas kintamasis (kad taškai būtų sugrupuoti į „tipus“) ar realus skaičius. Multivariacinės žymės, susidedančios iš kelių tokių kintamųjų, taip pat yra paplitusios. Žymėto taškinio proceso struktūra yra netvarkinga aibė

$$y = \{(x_1, m_1), \dots, (x_n, m_n)\}, x_i \in W, m_i \in M$$

čia

x_i yra vietos;

m_i yra atitinkamos žymės.

Tikimybė, kad taškas u turi žymę m yra

$$p(m|u) = \frac{\lambda(u,m)}{\lambda(u)} \quad (45)$$

kur $\lambda(u) = \sum_m \lambda(u,m)$ yra intensyvumo funkcija visų tipų taškams.

Žymėtuose taškiniuose procesuose gali būti skirtingų tipų taškų sąveikos arba to pačio tipo taškų sąveikos. Tarkime, kad tipo i taškai turi tolyginį intensyvumą λ_i , visiems i . Tuomet, dvejiems tipams i ir j nukrypimo funkcija G bus

$$G_{ij}(r) = P(R_{ij} \leq r) \quad (46)$$

kur R_{ij} yra atstumas nuo tam tikro taško iš tipo i iki artimiausio taško iš tipo j .

Sąlyginis intensyvumas $\lambda(u,m|x)$ yra iš esmės sąlyginė tikimybė rasti tašką iš tipo m vietovėje u , kai duota baigtinė informacija apie likusį procesą x .

III.1. Multi-tipo Puasono modeliai

Erdviniai taškiniai procesai, kurių duomenys turi kokybines žymes, vadinami multi-tipo (ang. *multitype*) taškiniais procesais. Analizuojant multi-tipo taškinius procesus yra taikomi tam tikri metodai, o priglodinant žymėtam procesui tam tikrą modelį, ne visi modeliai tinka, kuriuos galima priglodinti procesams be žymių. Aprašysime keletą žymėtų Puasono procesų, kuriais dažniausiai remiamasi analizuojant multi-tipo erdvinis taškinius procesus.

Pastovus Puasono procesas srityje R^2 su žymėmis srityje M gali būti apibrėžtas keletu būdų:

- atsitiktinai pažymėtas Puasono procesas (*Poisson [X], i.i.d. [M |X]*): Puasono taškinis procesas X su intensyvumu β yra sugeneruojamas pirmiausia. Tada kiekvienas taškas x_i yra pažymėtas atsitiktine žyme m_i , nepriklausomai nuo kitų taškų, su pasiskirstymu $P\{M_i = m\} = p_m, m \in M$;

- nepriklausomas Puasono procesas (*i.i.d. [M], Poisson [X|M]*): kiekvienai galimai žymei $m \in M$ yra sukurtas Puasono procesas X_m su intensyvumu β_m . X_m taškai yra pažymimi žyme m . Tada visi X_m procesai su skirtingomis žymėmis yra sudedami, kad duotų pažymėtą taškinį procesą;

- pažymėtas Puasono taškinis procesas (*jointly Poisson [X, M]*): yra sugeneruojamas Puasono procesas srityje $R^2 \times M$ su intensyvumo funkcija $\lambda(u, m) = \beta_m$ srityje u ir žymėmis m .

Atitinkamai šiems multi-tipo Puasono procesams gali būti formuluojamos hipotezės, kurias galima patikrinti:

- atsitiktinis žymėjimas;
- komponentų nepriklausomumas;
- pilnas erdvinis atsitiktinumas ir nepriklausomumas.

Kai žymės yra kokybinės, dažniausiai yra atliekama žvalgomoji žymėtų taškinų procesų analizė. Duomenis galima grafiškai perskelti į subprocesus grupuojant pagal žymes, ištirti šių subprocesų intensyvumus, sąlyginį tankį kitų žymių atžvilgiu. Skaitinių žymių atveju, kai žymės yra loginis ar kiekybinis kintamasis, dažniausiai yra brėžiama histograma arba vertinama branduolinė tankio funkcija. Skaitinės žymės yra suskirstomos į intervalus. Siekiant įvertinti žymių erdvinį trendą, galima suformuoti glodinančią branduolinę regresijos funkciją. Tada suglodinta žymė taške $u \in R^2$ bus

$$\hat{m}(u) = \frac{\sum m_i \kappa(u-x_i)}{\sum \kappa(u-x_i)} \quad (47)$$

čia

κ – glodinimo funkcija;

m_i - žymė taške x_i .

Taip pat galima nustatyti žymių erdvinį homogeniškumą/nehomogeniškumą pasinaudojus proceso perskėlimo į subprocesus funkcija, pagal žymių intervalus.

Multi-tipo taškinių procesų atveju galima ištirti taško artimiausių kaimynų žymes. Taip pat analizėje naudojamos apibendrinančios atstumų funkcijos, kurių pagalba galima nustatyti žymių priklausomumą.

Tarkime X yra stacionarus taškinis procesas. X_j - tam tikro tipo j taškų poaibis su intensyvumu λ_j . Tada bet kuriai tipų i ir j porai:

- $F_j(r)$ yra tuščių erdvių funkcija poaibiui X_j ;
- $G_{ij}(r)$ atstumo nuo i tipo taško iki artimiausio j tipo taško pasiskirstymo funkcija;
- $K_{ij}(r)$ yra tikėtinas j tipo taškų skaičius atstumu r iki i tipo taško.
- $L_{ij}(r)$ yra atitinkamai L funkcija

$$L_{ij}(r) = \sqrt{\frac{K_{ij}(r)}{\pi}} \quad (48)$$

- $g_{ij}(r)$ yra porinės koreliacijos funkcija

$$g_{ij}(r) = \frac{K_{ij}(r)}{2\pi r} \quad (49)$$

- $J_{ij}(r)$ apibrėžiama tokiu būdu

$$J_{ij}(r) = \frac{1-G_{ij}(r)}{1-F_j(r)} \quad (50)$$

Funkcijos G_{ij} , K_{ij} , L_{ij} , g_{ij} , J_{ij} yra vadinamos „kryžminėmis“ funkcijomis. Jos gali būti įvertintos atitinkamai funkcijų G_{cross} , K_{cross} , L_{cross} , $pcfcross$ bei J_{cross} pagalba R sistemoje.

Kryžminių funkcijų interpretacija yra labai panaši, kaip ir nežymėtų taškinių procesų atveju, tačiau ne visai tokia pati:

- Jeigu X_j yra tolygus Puasono procesas (CSR), tada $F_j(r) = 1 - \exp(-\lambda_j \pi r^2)$;
- Jeigu X_j yra tolygus Puasono procesas (CSR) ir nepriklauso nuo X_i , tada $G_{ij}(r) = 1 - \exp(-\lambda_j \pi r^2)$;
- Jeigu X_j ir X_i yra nepriklausomi, tada $K_{ij}(r) = \pi r^2$ atitinkamai $L_{ij}(r) = r$;
- Jeigu X_j ir X_i yra priklausomi, tada $g_{ij}(r) = 1$ ir $J_{ij}(r) = 1$.

Galima įvertinti ir atstumus nuo tam tikro vieno tipo taško iki taško su betkuria žyme. Tam naudojamos funkcijos:

- $G_i(r)$ atstumo nuo i tipo taško iki artimiausio kito tipo taško pasiskirstymo funkcija;
- $K_i(r)$ yra tikėtinas taškų skaičius atstumu r iki i tipo taško;
- $L_i(r)$ yra atitinkamai L funkcija

$$L_i(r) = \sqrt{\frac{K_i(r)}{\pi}} \quad (51)$$

- $J_i(r)$ apibrėžiama tokiu būdu

$$J_{ij}(r) = \frac{1-G_i(r)}{1-F(r)} \quad (52)$$

Šias funkcijas galima apskaičiuoti naudojant *Gdot*, *Kdot*, *Ldot* bei *Jdot* komandas R sistemoje.

Žymių koreliacinė funkcija

Koreliacinė funkcija $\rho_f(r)$ stacionaraus taškinio proceso Y atveju yra priklausomybės tarp dviejų taškų, esančių atstumu r , matas. Bendru atveju koreliacinė funkcija gali būti apibrėžiama tokiu būdu:

$$\rho_f(r) = \frac{E[f(M_1, M_2)]}{E[f(M, M')]} \quad (53)$$

čia

M_1, M_2 - taškų, esančių atstumu r , žymės;

M, M' - žymių pasiskirstymo nepriklausomos realizacijos.

Funkcija f gali būti bet kokia neneigiama funkcija. Dažniausiai nagrinėjami tokie variantai:

- Kiekybinėms žymėms: $f(m_1, m_2) = m_1 m_2$;
- Kokybinėms žymėms: $f(m_1, m_2) = 1\{m_1 = m_2\}$;
- Žymėms, įgyjančioms reikšmes intervale $[0, 2\pi]$: $f(m_1, m_2) = \sin(m_1 - m_2)$;

Funkcija $\rho_f(r)$ nėra „koreliacija“ statistine prasme, nes ji gali įgyti bet kokią neneigiamą, realią reikšmę. 1 reiškia, kad koreliacijos nėra. Reikšmių didesnių ar mažesnių už 1 interpretacija priklauso nuo funkcijos f pasirinkimo. Žymių koreliacija gali būti apskaičiuota komandos *markcorr* R programoje pagalba.

III.2. Žymėtų taškinių procesų atsitiktinumo testai

Homogeninio žymėto Puasono taškinio proceso atveju, nulinė hipotezė gali būti patikrinta tiesioginio generavimo pagalba naudojant viršutinį ir apatinį vokus.

Taip pat galima patikrinti ir kitokias hipotezes:

- *Atsitiktinis žymėjimas*: žymės yra sąlygiškai nepriklausomos ir vienodai pasiskirsčiusios.

Šiuo atveju taškinis vaizdas yra generuojamas iš aibės, laikant, kad taškų lokacijos yra fiksuotos ir atsitiktinai perskirstant žymes (ang. *randomly resampling*) gali būti su pasikartojimu ir be.

Atsitiktinio žymėjimo atveju:

- $G_i(r) = G_i(r)$
- $K_i(r) = K_i(r)$
- $J_i(r) = J_i(r)$

kur G , K ir J yra taškinio proceso be žymių funkcijos. Šiuo atveju kryžminės funkcijos yra labai komplikotos. Siekiant sukonstruoti atsitiktinio žymėjimo testo funkciją, dažniausiai yra naudojami skirtumai $K_i(r) - K(r)$.

- *Komponenčių nepriklausomumas*: kiekvienos žymės m subprocesas (poaibis) X_m yra nepriklausomas taškinis procesas.

Šiuo atveju vaizdas yra generuojamas dalijant visą duomenų aibę į vieno tipo taškų poaibius ir juos keičiant nepriklausomai vienas nuo kito.

Nepriklausomumo atveju:

- $G_{ij}(r) = F_j(r)$
- $K_{ij}(r) = \pi r^2$
- $J_{ij}(r) = 1$

Komponenčių nepriklausomumui nustatyti dažniausiai naudojamos kryžminės funkcijos, nes kitų aptartų funkcijų išraiškos dažniausiai būna labai sudėtingos.

- *Pilnas erdvinis atsitiktinumas ir nepriklausomumas*: tolygus žymėtas Puasono procesas R^2 su žymėmis aibėje M yra pilnai erdvėje atsitiktinis ir nepriklausomas taškinis procesas. Hipotezė gali būti patikrinta tiesioginio generavimo pagalba naudojant vokus.

Nehomogeninis žymėtas Puasono procesas Y , su kokybinėmis žymėmis iš baigtinės aibės M , pasižymi tokiais savybėmis:

- vietos X , apibrėžtos panaikinant žymes, išlieka nehomogeninis Puasono procesas erdvėje R^2 su tankio funkcija

$$\beta(u) = \sum \lambda(u, m) \quad (54)$$

- jeigu žymės yra pridedamos, atsižvelgiant į vietas X , jos yra nepriklausomos, o jų sąlyginis pasiskirstymas taške x_i bus

$$P\{M_i = m\} = \frac{\lambda(x_i, m)}{\beta(x_i)} \quad (55)$$

- subprocesas X_m , sudarytas iš taškų su žymėmis m , yra nehomogeninis Puasono procesas su intensyvumu

$$\beta_m(u) = \lambda(u, m) \quad (56)$$

- subprocesai X_m su skirtingomis žymėmis yra nepriklausomi procesai.

IV. ETP ANALIZĖ R APLINKOJE

IV.1. Įvadas į atvirojo kodo programą R

Taškinių procesų analizei naudosime atvirojo kodo programą R. R yra nemokama programinė įranga su atviro šaltinio licencija. Šį statistinį paketą galima parsisiųsti iš r-project.org ir lengvai įdiegti kompiuteryje. Be to, be pagrindinės R sistemos, R žiniatinklio svetainė taip pat siūlo daugelį pridėdamų modulių (bibliotekas (ang. *libraries*) ar papildomus paketus (ang. *packages*). Juos galima parsisiųsti iš cran.r-project.org.

***Spatstat* paketas**

Spatstat yra R paketas, kurį parašė Adrian Baddeley ir Rolf Turner, skirtas analizuoti erdvinius duomenis. Dabartinės *spatstat* versijos yra skirtos dirbti pagrindė su dvimačiais erdviniais taškiniais procesais. Paketą sudaro:

- taškinių procesų sukūrimas, manipuliavimas ir braižymas;
- tiriamoji duomenų analizė;
- taškinio proceso modelių imitavimas;
- parametrinis modelio pritaikymas (ang. *model-fitting*);
- hipotezės testai, diagnostika.

Spatstat yra vienas iš didžiausių paketų pasiekiamų R programoje, su daugiau kaip 300 vartotojų lygmens funkcijų ir 500 puslapių „vadovu“. Paketui sukurtas internetinis puslapis www.spatstat.org, kuriame galima rasti visą informaciją apie šį paketą. *Spatstat* yra suprojektuotas, kad apimtų visus tiriamosios analizės standartinius tipus taškiniais procesams.

Duomenų klasės

Aptarsime skirtingas erdvių duomenų rūšis (taškų struktūrą, langus, vaizdo elementų (pikselių) atvaizdus ir kt.) bei jų valdymą pakete *spatstat*.

R programa naudoja kalbą „orientuotą į objektą“. Duomenys, turintys tam tikrą struktūrą (pavyzdžiui laiko eilutė) yra traktuojami kaip vienas „objektas“. Kiekvienas objektas R programoje, priklausomai nuo jo struktūros, yra identifikuojamas kaip priklausantis tam tikram tipui ar klasei.

Taškinio proceso duomenų valdymui, *spatstat* pakete yra apibrėžtos tam tikros objektų klasės:

- *ppp*: plokštuminis taško procesas;
- *owin*: erdvinė sritis („stebėjimo langas“);
- *im*: vaizdo elementų atvaizdas;
- *psp*: tiesės atkarpos struktūra;
- *tess*: mozaika.

Didžioji dalis funkcijų *spatstat* pakete dirba su minėtų klasių objektais. Norint panaudoti šias

funkcijas reikia perkelti turimus nesutvarkytus duomenis į R programą ir paskui paversti juos į tinkamo formato objektą.

Klasės „ppp“ objektai

Taškiniai procesai *spatstat* pakete yra pristatomi kaip klasės *ppp* objektai. Į juos įeina: taškų koordinatės, laisvai pasirenkamos prie taškų pridėtos „žymių“ vertės ir analizuojamos srities erdvinio lango apibūdinimas. Taškinio proceso objektas *P* turi šiuos komponentus:

- *P\$n* yra taškų skaičius (kuris gali būti lygus nuliui);
- *P\$x* yra skaitmeninis vektorius, aprašantis taškų *x* koordinates. Jo ilgis lygus *P\$n* (ir gali būti lygus nuliui);
- *P\$y* yra skaitmeninis vektorius, aprašantis taškų *y* koordinates. Jo ilgis taip pat lygus *P\$n*;
- *P\$marks* aprašo žymes. Tai yra arba nulis, arba vektorius, kurio ilgis yra *P\$n*, aprašantis žymės vertes. *P\$marks* vertės gali būti bet kokio tipo (žodinio, skaitinio, loginio, kompleksinio). Jei žymės yra kokybiniai kintamieji, tai *P\$marks* yra faktorius;
- *P>window* yra klasės „*owin*“ objektas („stebėjimo langas“) nustatantis analizavimo sritį ar erdvinį „langą“.

Ypatingai pastabėsime, kad, kuriant naują taškinio proceso objektą, turime nustatyti erdvinę sritį ar langą, kuriame procesas bus apibrėžtas. *Spatstat* pakete stebėjimo langas yra neatimama taškinio proceso dalis. Taškinio proceso duomenys susideda iš žinių apie vietas, kur taškai nebuvo pastebėti ir buvo stebimi.

Norint pradėti darbą su *spatstat* paketu reikia paleisti R programą ir į kodo langą įrašyti:

```
> library(spatstat)
```

Atsakymas turėtų būti kažkas panašaus į:

```
Loading required package: mgcv  
This is mgcv 1.7-9. For overview type 'help("mgcv-package")'.  
Loading required package: deldir  
deldir 0.0-16  
Please note: The process for determining duplicated points  
has changed from that used in version 0.0-9 (and previously).  
spatstat 1.25-0
```

Parašius „*help(spatstat)*“ programa pateiks visą informaciją apie paketą bei galimų komandų sąrašą. Norint pamatyti informaciją apie konkrečią komandą reikia įvesti „*help(command)*“. Norint susidaryti bendrą įspūdį apie paketą *spatstat* galima paleisti paketo demonstraciją naudojant „*demo(spatstat)*“ .

Norint analizuoti turimus taškinio proceso duomenis pakete *spatstat*, turime perkelti neapdorotus duomenis į R programą ir paversti juos į objektą iš klasės „*ppp*“.

Paketo „vadovas“ duoda vieną pagrindinį būdą, t.y. sudėti taškų koordinates x ir y į 2 vektorius (x ir y atitinkamai), jei yra žymių, pridėtų prie taško duomenų, jas sudėti į vektorių m atitinkamai taškams (*pastaba*: tikrai viena žymės vertė yra leistina vienam taškui) bei sukurti ETP.

Vertė, gražinta funkcijos *ppp*, yra klasės „*ppp*“ objektas, taškinio proceso reprezentavimas, t.y. taškinis vaizdas.

Jei langas nėra stačiakampis, tai reikia panaudoti komandą:

```
> ppp(x, y, langas = W),
```

kur W yra apibrėžtas atskirai.

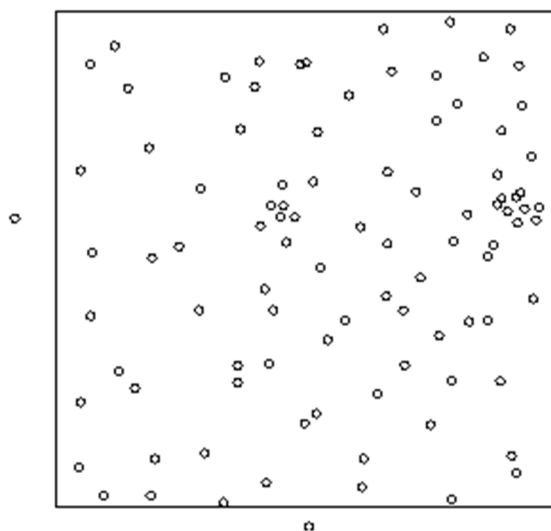
Koordinatės gali būti įvestos į R programą įvairiais būdais, pavyzdžiui įrašant jas tiesiogiai.

Norint sukurti duomenis ir kai reikalingas nedidelis duomenų kiekis, vartotojas gali pats sudėti duomenų taškus lange.

Tam naudojama funkcija:

```
> X<-clickppp(100)
```

SPUSTELĖKITE RANKA 100 KARTŲ



9 pav. Ranka suvestų duomenų taškinis vaizdas

Įvedus šią funkciją į programą, ji automatiškai išspausdina tuščią langą, kuriame vartotojas kursoriaus pagalba taškus lange sudėlioja pats.

Kaip matome, taškų kiekį gali nustatyti pats vartotojas. Mūsų atveju, baigus taškų žymėjimą grafike, programa iškart duoda įspėjantį atsakymą, kad yra du taškai už galimos srities ribų ir yra besidubliuojančių taškų, kurie tam tikrų funkcijų pagalba turėtų būti eliminuoti iš tyrimo.

Taip pat galime naudoti funkciją *file="filename"* norint peržiūrėti duomenų srautą. Jei duomenys yra tvarkingai suvesti į lentelę, kur kiekvienas taškas yra atskirtas nuo kito, galime naudoti *read.table* komandą.

Visų pirma nuskaitome lentelę (duomenys pateikti el. laikmenoje pavadinimu „duomenu_lentele“, txt formatu) su x ir y koordinatėmis:

```
> duom<-read.table("duomenu_lentele.txt",header=TRUE)
```

Priskiriame stulpelių reikšmes taškų koordinatėms:

```
> x<-duom$X
```

```
> y<-duom$Y
```

Sukuriame taškinį vaizdą:

```
> P<-ppp(x,y,c(0,5),c(0,5))
```

```
> P
```

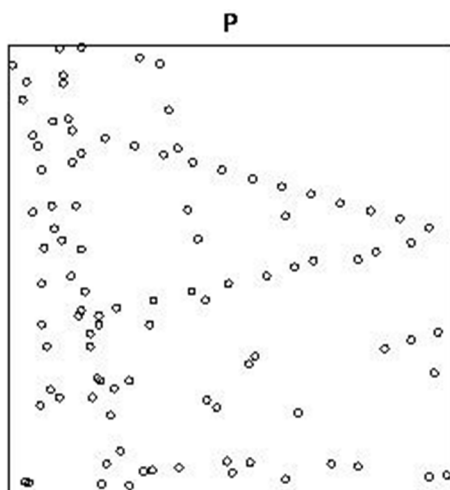
```
planar point pattern: 100 points
```

```
window: rectangle = [0, 5] x [0, 5] units
```

kurį pažymime raide P. Intervalas $c(0,5)$ nurodo x ir y verčių intervalus.

Gautą taškinį vaizdą nubraižome:

```
> plot(P)
```



10 pav. Nuskaitytų koordinatžių taškinis vaizdas

Duomenų tikrinimas

Besidubliuojantys taškai yra labai dažnas reiškinys duomenyse (t.y. kur du įrašai faile siejasi su ta pačia (x, y) vieta), todėl yra apdairu tikrinti duomenis, kad nebūtų netikslumų. Tai galima atlikti:

- Išspausdinus to paties rango vertes bei žymes ir patikrinus ar nėra padarytų klaidų įvedant duomenis, ir ar koordinatės nebuvo suapvalintos.

- Kai tik koordinatės yra įvestos į R programą kaip dviejų stulpelių matrica ar duomenų rėmas D , galima duomenis tikrinti komandos `any(duplicated(D))` pagalba. Jei duomenys jau yra taškinio vaizdo formoje X , galima taip pat surinkti komandą `any(duplicated(X))` duomenų patikrinimui.

```
> any(duplicated(P))
```

```
[1] FALSE
```

Programos išspausdintas atsakymas parodo, kad pasikartojančių taškų nėra. Jei tokių taškų būtų, naudotume funkciją *unique(P)* ir juos pašalintume.

Vienetai

Taškinis procesas X gali teikti informaciją apie ilgio vienetus, kuriais yra aprašytos x ir y koordinatės. Ši informacija yra laisvai pasirenkama; tai paprasčiausiai įgalina paketą išspausdinti geresnius pranešimus ir paaiškinti kampus brėžiniuose. Matavimo vienetas taip pat gali būti duotas kaip kažkoks standartinio vieneto kartotinis.

Pastaba: vieneto pavadinimas taikomas tik koordinatėms, bet ne žymėms ar taškiniam procesui.

Matavimo vienetų priskirimą galima padaryti:

```
> unitname(P) <- c("metre", "metres")
```

Jei norima paversti vienus vienetus į kitus (pavyzdžiui metrus į kilometrus), naudojame komandą *rescale*. Jei norima pakeisti koordinatas naudojant linijines transformacijas, tuomet naudojame komandą *affine*.

Kiti būdai sukurti taškinius vaizdus

Norint sukurti taškinį vaizdą, galime:

- sukurti jį iš neapdorotų duomenų naudojant funkciją *ppp*;
- konvertuoti duomenis iš kitų formatų, naudojant *as.ppp*;
- nuskaityti duomenis iš failo, naudojantis komanda *scanpp*;
- transformuoti jau turimą taškinį vaizdą naudojant įvairius įrankius;
- sukurti atsitiktinį vaizdą naudojant vieną iš imitavimo šablonų;
- panaudoti standartinio taškinio vaizdo duomenis, esančius pakete.

Sukurto erdvinio taškinio vaizdo bendrąją statistinę informaciją galima sužinoti panaudojus komandą *summary(P)*:

```
> summary(P)
Planar point pattern: 100 points
Average intensity 4 points per square metre
Window: rectangle = [0, 5] x [0, 5] metres
Window area = 25 square metres
Unit of length: 1 metre
```

Iš pateiktos informacijos matome, kad koordinatės yra aprašytos metrais, būdingas taškų pasiskirstymo intensyvumas yra 4 taškai kvadratiname metre.

IV.2. Puasono proceso tyrimas

Atvirojo kodo programoje R galima sukurti įvairius taškinius procesus (pvz. nehomogeninis Puasono procesas, Gibbs'o procesas ir kt.) generuojant erdvinius duomenis. Kiekvienai taškinų

procesų rūšiai generuoti yra naudojamos skirtingos komandos. Sukursime ETP naudodami standartinę duomenų generavimo funkciją ir atliksime jo analizę.

Kaip buvo minėta teorijoje, taškinis vaizdas laikomas ETP tuomet, kada jis yra apibrėžiamas tam tikroje srityje ir kai atsitiktiniai erdviniai taškai yra ETP stebiniai. Taigi, sukursime taškinį vaizdą iš taškų, kurie atvaizduos medžius, sergančius šaknų puvinio liga. ETP sritį apibrėšime tam tikrą miško teritoriją, pvz. Girulių miškuose. Taigi, apibrėšime srities langą. Lango forma gali būti be galo įvairi. Langą apibrėšime daugiakampio formos.

```
>langas<-owin(poly=list(x=c(1,1.5,3,3.5,3.5,3,1.5,1),y=c(2,1.5,1.5,2,3.5,4,4,3.5)))
```

```
>langas
```

```
window: polygonal boundary
```

```
enclosing rectangle: [1, 3.5] x [1.5, 4] units
```

Apsibrėžėme langą, kurio x ašis yra intervalas [1;3,5], o y ašis - intervalas [1,5;4].

Pasinaudoję funkcija *rpoint*, kuri automatiškai sugeneruoja atsitiktinius taškus ir nurodę taškų skaičių, gauname taškinį vaizdą, kurį laikysime medžių, sergančių šaknų puvinio liga, erdvinio taškiniu procesu.

```
>p<-rpoint(700,win=langas)
```

```
>p
```

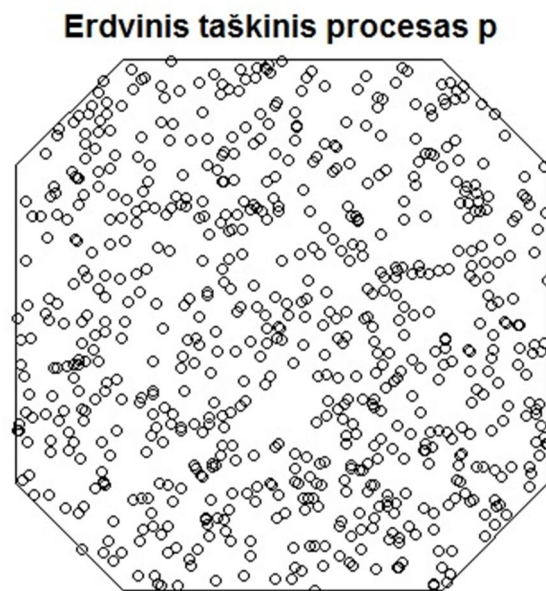
```
planar point pattern: 700 points
```

```
window: polygonal boundary
```

```
enclosing rectangle: [1, 3.5] x [1.5, 4] units
```

Pasirinkome sugeneruoti 700 taškų. ETP atvaizduojame grafiškai:

```
> plot(p)
```



11 pav. Erdvinis taškinis procesas p

Patikriname ar duomenyse nėra pasikartojančių taškų.

```
> any(duplicated(p))
```

```
[1] FALSE
```

Jų nėra, todėl be jokių papildomų veiksmų galima šį taškinį procesą tirti toliau (sukurto ETP p duomenys pateikti el. laikmenoje pavadinimu „duomenys_p“, txt formatu).

Visų pirma ištirsime ETP intensyvumą, nes intensyvumo tyrimas – vienas iš pirmųjų žingsnių taškinį procesų analizėje.

Apskaičiuojame ETP intensyvumą:

```
> summary(p)
```

```
Planar point pattern: 700 points
```

```
Average intensity 122 points per square unit
```

```
Window: polygonal boundary
```

```
single connected closed polygon with 8 vertices
```

```
enclosing rectangle: [1, 3.5]x[1.5, 4]units
```

```
Window area = 5.75 square units
```

```
> summary(p)$intensity
```

```
[1] 121.7391
```

Programa parodė, jog į kvadratinį ploto vienetą, patenka apytiksliai 122 taškai. Iš sukurto proceso grafinio vaizdo matome, jog intensyvumas visame lange pasiskirstęs gana vienodai, todėl galime įtarti, jog jis yra stacionarus. Intensyvumą tirsime anksčiau aprašytomis technikomis: kvadratų skaičiavimo bei branduolio glodinimo technika (Kernel'io intensyvumo įverčio skaičiavimas).

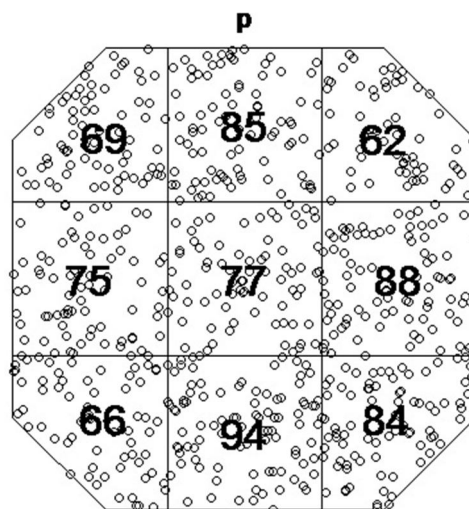
```
> Q<-quadratcount(p,nx=3, ny=3)
```

```
> Q
```

	x		
y	[1,1.83]	(1.83,2.67]	(2.67,3.5]
(3.17,4]	69	85	62
(2.33,3.17]	75	77	88
[1.5,2.33]	66	94	84

```
> plot(p,cex=1)
```

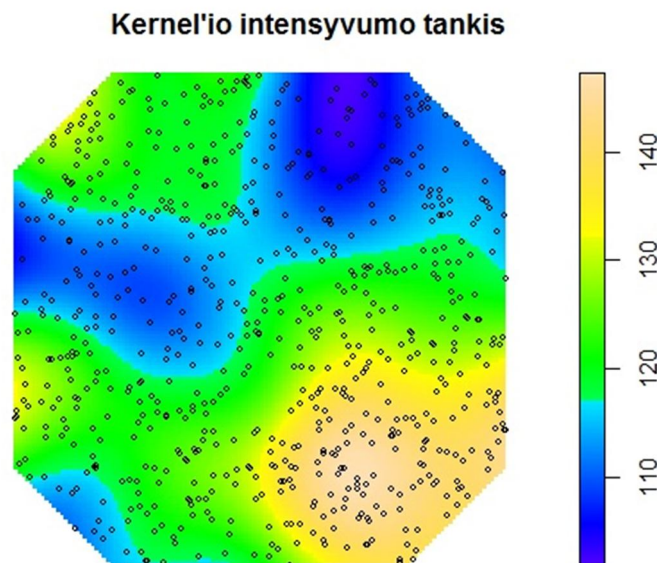
```
> plot(Q,add=TRUE,cex=2)
```



12 pav. Taškų skaičius kvadratuose

Kadangi langas yra šešiakampio formos, todėl parinkome jį suskaidyti į tris eilutes ir tris stulpelius, taip gaunant atitinkamus vienodo dydžio narvelius, kad būtų galima juos palyginti. Matome, kad narveliuose taškų skaičius skiriasi, tačiau skirtumai nėra ženklūs. Daugiakampio apatiniame dešiniajame kampe taškų skaičius, lyginant su visu langu, didžiausias. Taškų pasiskirstymas narveliuose priklauso ir nuo mūsų parinktų narvelių dydžių. Kad geriau pamatytume intensyvumo pasiskirstymą braižome Kernel'io intensyvumo tankio įvertinį.

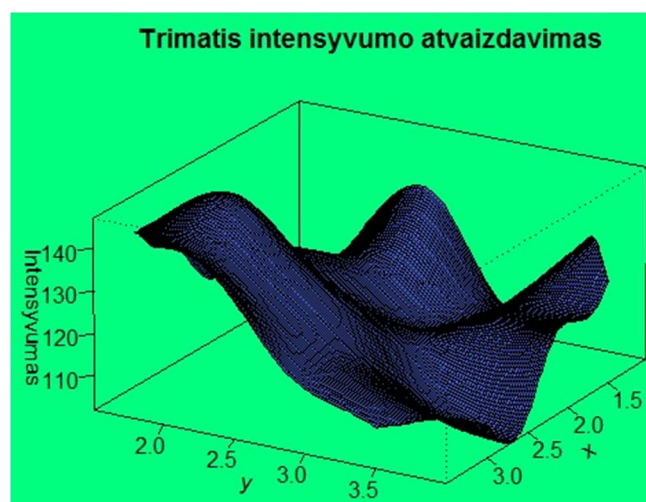
```
> den<-density(p)
> plot(den, main="Kernel'io intensyvumo tankis")
> plot(p,add=T,cex=0.5)
```



13 pav. Kernel'io intensyvumo tankio įvertinys

Matome, kad taškų tankis yra didesnis grafiko apatinėje dešiniojoje pusėje. Taip pat taškinio proceso intensyvumo tankis gali būti pavaizduotas trimatį vaizdu (ang. *perspective plot*).

```
> persp(den, zlab="Intensyvumas", main="Trimatis intensyvumo atvaizdavimas", theta = 120, phi = 20, colmap="royalblue", scale = FALSE, ltheta = 270, lphi=150, shade = 0.75, border = TRUE, box = TRUE,ticktype="detailed")
```



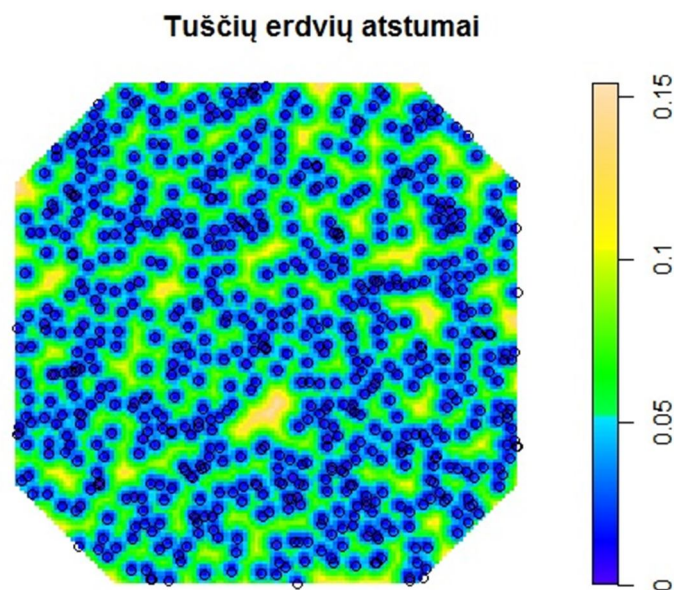
14 pav. Trimatis intensyvumo atvaizdavimas

Atlikę intensyvumo vertinimą kvadratų metodu bei Kernelio įverčio atvaizdavimu grafiškai, galime daryti išvadą, jog intensyvumas yra palyginus gana vienodas, didelių nukrypimų atskiruose regionuose nematyti.

Apskaičiuosime atstumus tarp taškų, kad galėtume nustatyti taškų išsidėstymo tipą. Atstumų funkcijų grafikus sudarysime naudodami 95% pasikliautinįjį intervalą, taip matydami funkcijos reikšmingumą.

Visų pirma apskaičiuojame tuščius erdvinius atstumus, t.y. atstumus nuo fiksuotos vietos lange iki artimiausio taško.

```
> esd<-distmap(p)
> plot(esd,main="Tuščių erdvių atstumai")
> plot(p,add=TRUE)
```

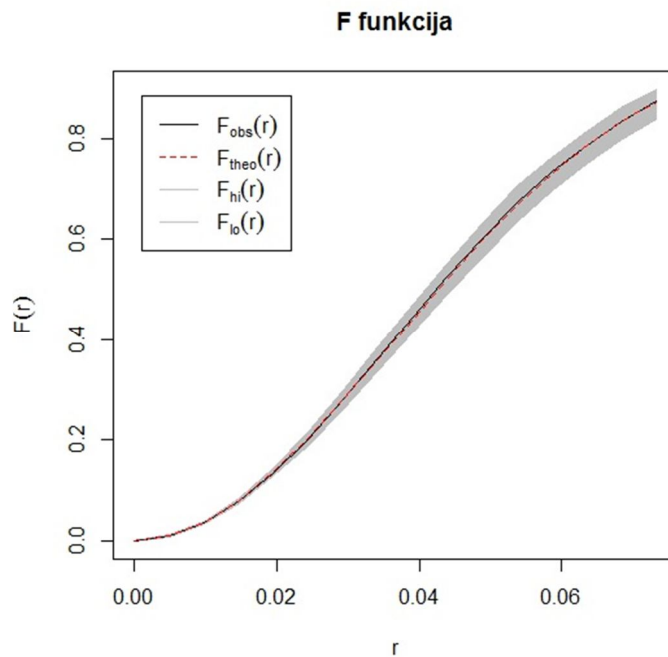


15 pav. Tuščių erdvių atstumai

Apskaičiuosime tuščių erdvių atstumo pasiskirstymo funkciją F .

```
> F<-envelope(p,Fest,nsim=59,rank=1)
> plot(F, main="F funkcija")
```

lty	col	key	label	meaning
obs	1	1	obs $F[obs](r)$	observed value of $F(r)$ for data pattern
theo	2	2	theo $F[theo](r)$	theoretical value of $F(r)$ for CSR
hi	1	8	hi $F[hi](r)$	upper pointwise envelope of $F(r)$ from simulations
lo	1	8	lo $F[lo](r)$	lower pointwise envelope of $F(r)$ from simulations

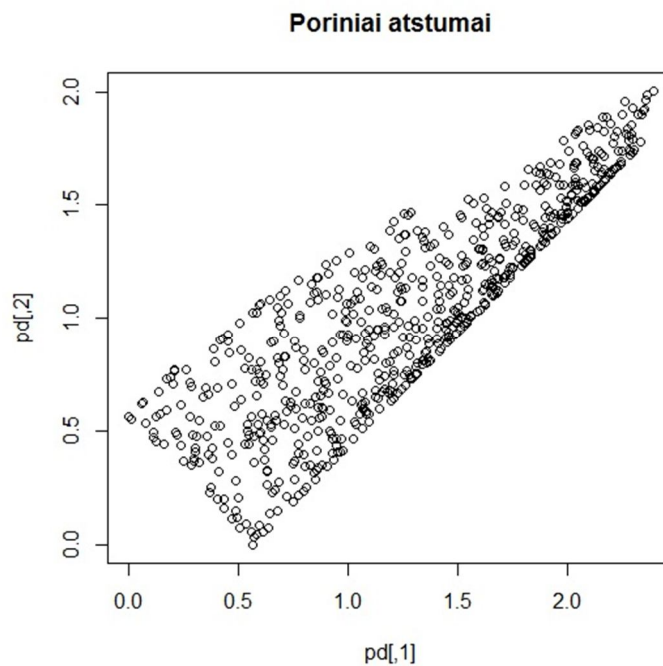


16 pav. Tuščių erdvių atstumo pasiskirstymo funkcija F

Kaip matome, apskaičiuotas F funkcijos įvertis sutampa su teorine Puasono proceso F funkcija ir neišeina iš pasikliautinio intervalo ribų. Jeigu apskaičiuotas F funkcijos grafikas išeitų iš voko ribų, kirstadamas viršutiniją ribą, tuomet sakytume, jog duomenims būdingas reguliarumas, jei kirstų apatinę ribą, tuomet duomenims būtų būdingi klasteriai. Mūsų atveju F funkcijos grafikas iš voko neišeina, todėl sakome, jog taškams būdingas atsitiktinumas.

Apskaičiuosime atstumus tarp visų taškų porų.

```
> pd<-pairdist(p)
> plot(pd, main="Poriniai atstumai")
```



17 pav. Poriniai taškų atstumai

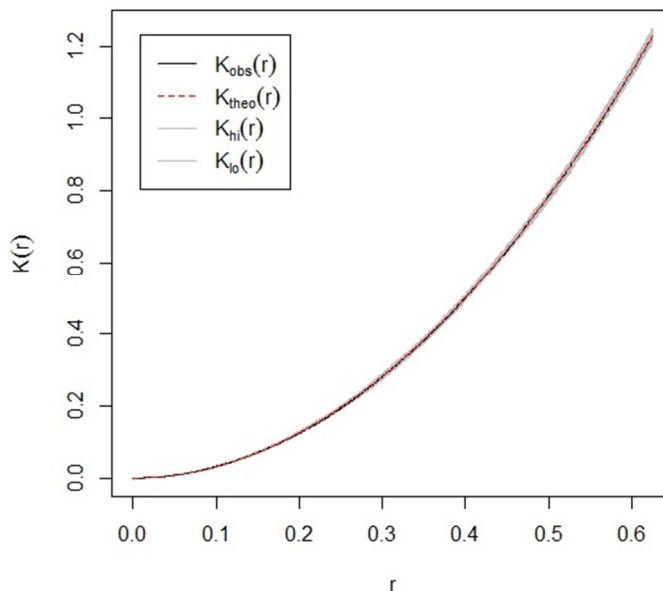
Skaičiuojame porinių atstumų pasiskirstymo funkciją K .

```
>K<-envelope(p,Kest,nsim=59,rank=1)
```

```
>plot(K, main="K funkcija")
```

	lty	col	key	label	meaning
obs	1	1	obs	$K[obs](r)$	observed value of $K(r)$ for data pattern
theo	2	2	theo	$K[theo](r)$	theoretical value of $K(r)$ for CSR
hi	1	8	hi	$K[hi](r)$	upper pointwise envelope of $K(r)$ from simulations
lo	1	8	lo	$K[lo](r)$	lower pointwise envelope of $K(r)$ from simulations

K funkcija



18 pav. Porinių atstumų funkcija K

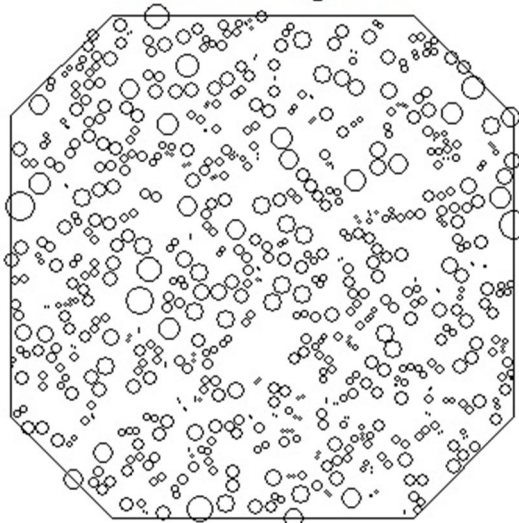
Iš grafiko matome, jog visos K funkcijos korekcijos ir apskaičiuotas K funkcijos įvertinys sutampa su teorine Puasono verte.

Apskaičiuosime artimiausio kaimyno atstumus.

```
>nndist(p)
```

```
>plot(P%mark%(nndist(p)/2),markscale=1,main="Stienen'o diagrama")
```

Stienen'o diagrama

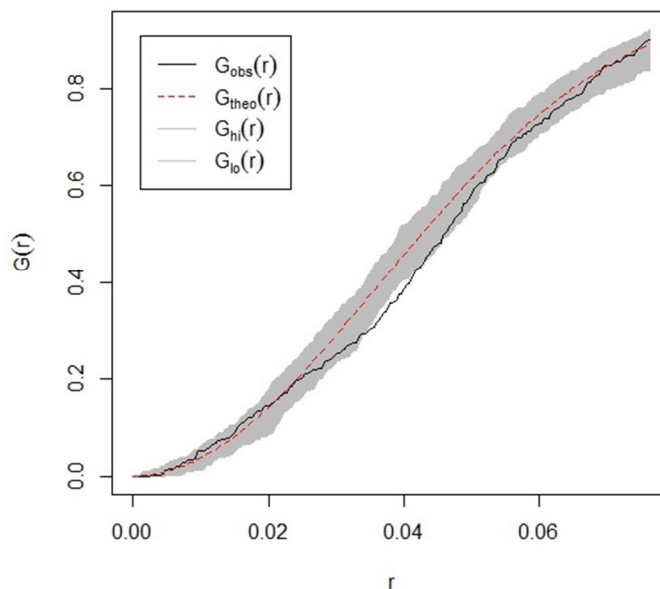


19 pav. Artimiausio kaimyno atstumai

Skaičiuojame artimiausio kaimyno atstumų pasiskirstymo funkciją G .

```
> G<-Gest(p)
> plot(G, main="G funkcija")
  lty col key label meaning
obs 1 1 obs G[obs](r) observed value of G(r) for data pattern
theo 2 2 theo G[theo](r) theoretical value of G(r) for CSR
hi 1 8 hi G[hi](r) upper pointwise envelope of G(r) from simulations
lo 1 8 lo G[lo](r) lower pointwise envelope of G(r) from simulations
```

G funkcija



20 pav. Artimiausio kaimyno atstumų pasiskirstymo funkcija G

Matome, jog artimiausio kaimyno atstumų pasiskirstymo funkcija nukrypsta nuo teorinės Puasono funkcijos ir šiek tiek išeina iš reikšmingumo intervalo ribų. Funkcijos nukrypimas žemiau Puasono teorinės funkcijos ir žemiau apatinės voko ribos rodo, jog taškams būdingas reguliarumas. Todėl galime numanyti, jog taškams, kurie išsidėstę atstumu ~ 0.04 , yra būdingas reguliarumas.

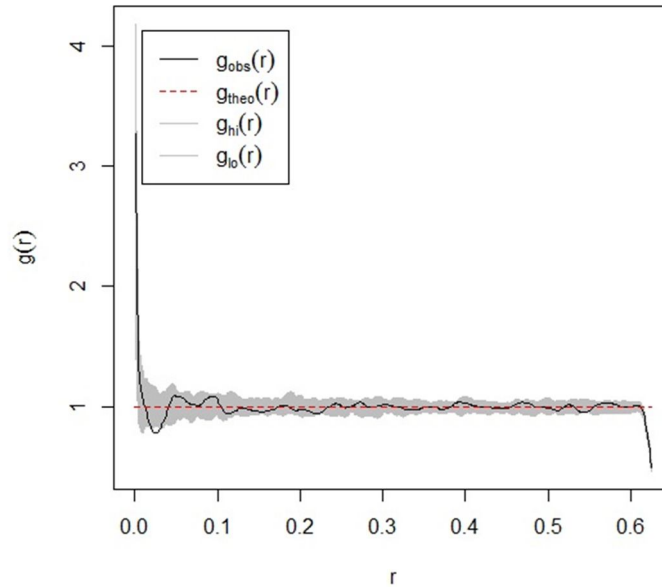
Kaip jau buvo minėta teorijoje, kartais atstumų funkcijas sunku vertinti, kadangi jos gali duoti skirtingus rezultatus. Mūsų atveju, F ir K atstumų funkcijos duoda vienodą rezultatą, t.y. sutampa su Puasono teorinėmis funkcijų reikšmėmis, o funkcija G šiek tiek nukrypsta. Dėl tikslesnių išvadų apskaičiuosime kai kurias atstumų funkcijų transformacijas.

Apskaičiuosime vieną iš K funkcijos transformacijų, t.y. porinę koreliacinę funkciją

$$g(r) = \frac{K(r)}{2\pi r}.$$

```
> g<-envelope(p,pcf,nsim=59,rank=1)
> plot(g, main="Porine koreliacine funkcija")
  lty col key label meaning
obs 1 1 obs g[obs](r) observed value of g(r) for data pattern
theo 2 2 theo g[theo](r) theoretical value of g(r) for CSR
hi 1 8 hi g[hi](r) upper pointwise envelope of g(r) from simulations
lo 1 8 lo g[lo](r) lower pointwise envelope of g(r) from simulations
```

Porinė koreliacinė funkcija



21 pav. Porinė koreliacinė funkcija $g(r)$

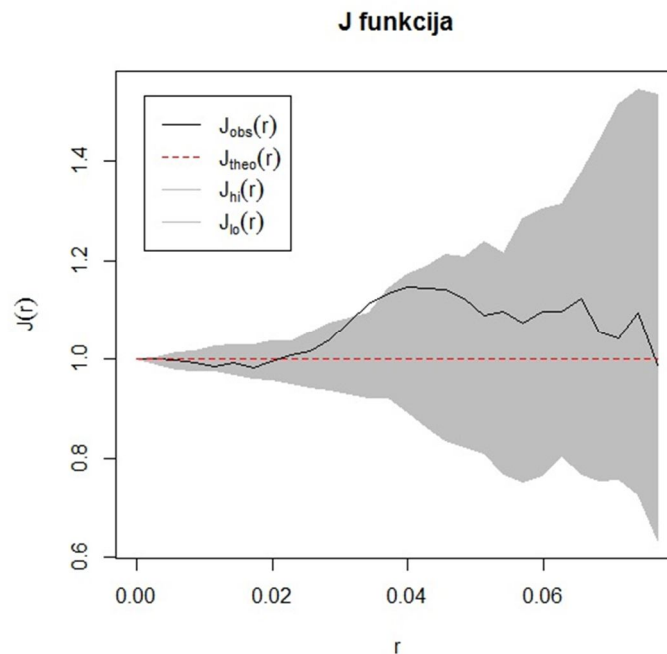
Šios funkcijos interpretacija yra tokia pati kaip ir K funkcijos. Matome, jog teorinė Puasono $g(r)$ funkcija yra lygi 1, o įvertinta mūsų ETP $g(r)$ funkcija taip pat krenta apie 1, varijuodama labai ne stipriai. Jeigu $g(r)$ funkcijos reikšmė būtų >1 , sakytume, jog duomenims būdingi klasteriai, jeigu $g(r)$ būtų <1 , sakytume, jog duomenims būdingas reguliarumas. Šiuo atveju, sakome, jog duomenims būdingas atsitiktinumas.

Apskaičiuosime funkcijų F ir G kombinacinę funkciją J .

```
> J<-envelope(p,Jest,nsim=59,rank=1)
```

```
> plot(J,main="J funkcija")
```

	lty	col	key	label	meaning
obs	1	1	obs	$J[obs](r)$	observed value of $J(r)$ for data pattern
theo	2	2	theo	$J[theo](r)$	theoretical value of $J(r)$ for CSR
hi	1	8	hi	$J[hi](r)$	upper pointwise envelope of $J(r)$ from simulations
lo	1	8	lo	$J[lo](r)$	lower pointwise envelope of $J(r)$ from simulations



22 pav. Kombinacinė funkcija J

Iš J funkcijos grafiko matome, jog funkcijos grafikas šiek tiek išeina iš pasikliautinio intervalo ribų ir ties atstumu ~ 0.04 labiausiai skiriasi nuo Puasono teorinės J funkcijos reikšmės. Kadangi matėme, jog G funkcija ties ~ 0.04 taip pat skiriasi nuo Puasono teorinės funkcijos, tai galime teigti, jog būtent šis G funkcijos nuokrypis ir įtakoja J funkcijos reikšmes.

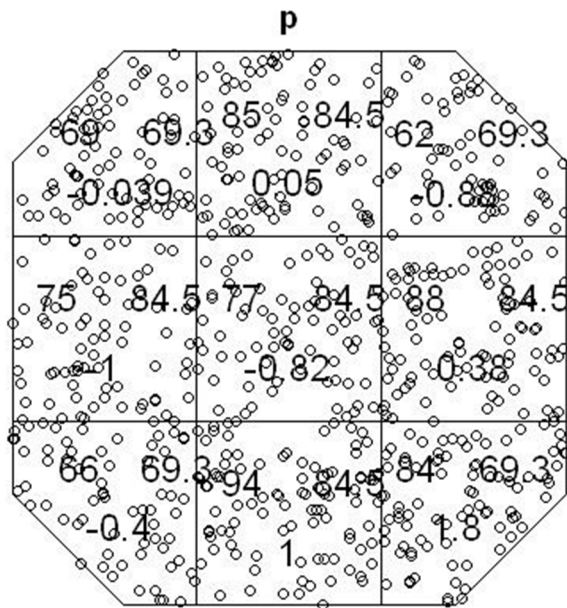
Visas atstumų funkcijas galima pavaizduoti kartu komandos `plot(allstats(p))` pagalba.

(Grafikas pateiktas *priede nr. 1, 23 pav.* Atstumų funkcijos F, G, J, K).

Įvertinę ETP intensyvumą bei apskaičiavę atstumų funkcijas, matome, jog klasterizacijos ir stipraus reguliarumo atvejų nebuvo iširta, o tai reiškia, jog duomenys yra pasiskirstę atsitiktinai. Taip pat intensyvumas visoje ETP srityje yra gana vienodas, todėl darome išvadą, jog turime vienarūšį ETP intensyvumą.

Kitas žingsnis analizuojant ETP yra nulinės hipotezės patikrinimas. Visų pirma, pilną erdvinį atsitiktinumą patikrinsime χ^2 testo pagalba.

```
> QTest<-quadrat.test(p,nx=3,ny=3)
> QTest
  Chi-squared test of CSR using quadrat counts
data: p
X-squared = 6.6058, df = 8, p-value = 0.5797
Quadrats: 9 tiles (levels of a pixel image)
> plot(p)
> plot(QTest,add=T,cex=2)
```



24 pav. χ^2 testas

Šiame paveiksle narveliuose esantis kairiajame kampe skaičius – nustatytas taškų skaičius, dešiniame kampe – spėtas taškų skaičius, apatinė reikšmė – Pirsono likutinė reikšmė, apskaičiuota pagal formulę:

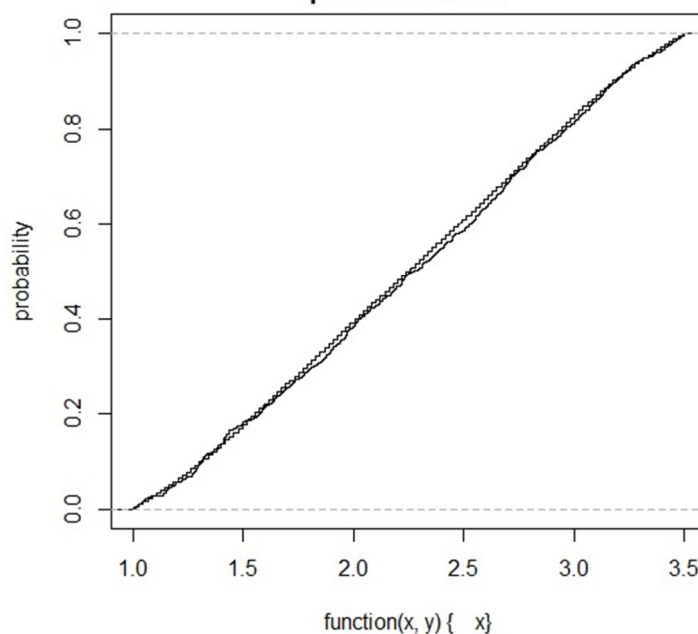
$$\text{Pirsono liekana} = (\text{nustatyta reikšmė} - \text{spėta reikšmė}) / \sqrt{\text{spėta reikšmė}}$$

Kaip matome Chi-kvadrato testas neatmeta nulinės hipotezės, kadangi Chi kvadrato reikšmė 6.6058 yra daugiau už reikšmę iš lentelės su 8 laisvės laipsniais, t.y. $6.6058 > 2.73$, o $p\text{-value}$ yra $0.5797 > 0.05$ (reikšmingumo lygmenį pasirenkame $\alpha = 0.05$), todėl galime teigti, kad taškai yra pasiskirstę nepriklausomai ir atsitiktinai pagal Puasono modelį, kitaip tariant jiems yra būdingas pilnas erdvinis atsitiktinumas.

Nulinę hipotezę patikrinsime atlikdami *Kolmogorov-Smirnov* testą. Pasirenkame taškų išsibarstymo funkciją $T(x,y)=x$.

```
> KS<-kstest(p,function(x,y) {x})
> plot(KS)
```

Spatial Kolmogorov-Smirnov test of CSR
 based on distribution of covariate "function(x, y) { x}"
 p-value= 0.7901



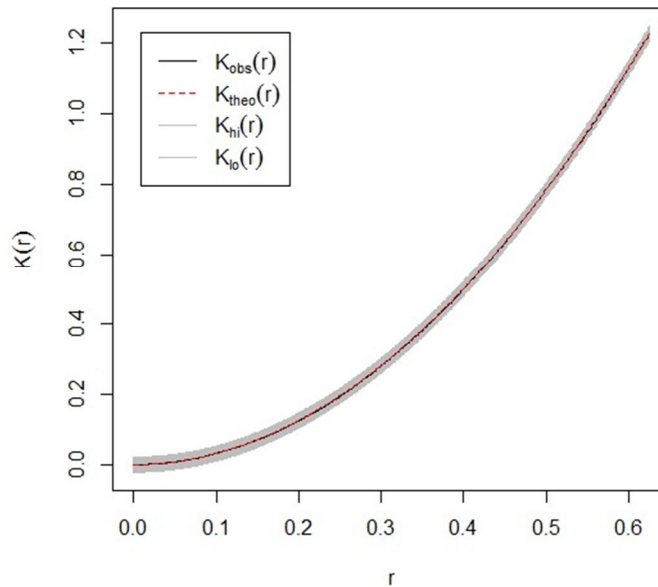
25 pav. Kolmogorov-Smirnov testas

```
> pval<-KS$p.value
> pval
[1] 0.7900867
```

Iš grafiko matome, kad išmatuota ir tikėtina išsibarstymo funkcijos reikšmės galima sakyti sutampa, o testo pateikta *p-value* yra lygi 0.7901. Tai tik patvirtina mūsų nulinės hipotezės priėmimą. Pilną erdvinį atsitiktinumą ištirsime apibendrinto *Monte Carlo* testo pagalba.

```
> KE<- envelope(p, Kest, nsim = 59,rank=1, global=TRUE)
> plot(KE)
  lty col key label meaning
obs 1 1 obs K[obs](r) observed value of K(r) for data pattern
theo 2 2 theo K[theo](r) theoretical value of K(r) for CSR
hi 1 8 hi K[hi](r) upper critical boundary for K(r)
lo 1 8 lo K[lo](r) lower critical boundary for K(r)
```

Monte Carlo testas



26 pav. Monte Carlo testas

Matome, jog K funkcijos įverčiai neperžengia kritinių ribų ir sutampa su teorine K funkcijos reikšme, todėl galime patvirtinti, jog ETD yra būdingas pilnas erdvinis atsitiktinumas, kitaip sakant priimame nulinę hipotezę, jog procesas yra atsitiktinai ir nepriklausomai pasiskirstęs pagal Puasono modelį.

Kitas žingsnis ETP analizėje yra geriausios tankio funkcijos duomenims parinkimas. Parenkame keturias trendo funkcijas:

- ~ 1 , nebūdingas trendas,
- $\sim x+y$, pirmojo laipsnio trendas,
- $\sim x+y+I(x^2)+I(y^2)+I(x*y)$, antrojo laipsnio trendas,
- $\sim x$, trendas pagal x koordinatę.

Aprašome šias funkcijas ir palyginsime jas naudoami Akaikės informacijos kriterijų.

```
>fit1<-ppm(p, ~1)
>fit1
Stationary Poisson process
Uniform intensity: 121.739130434783
      Estimate      S.E.      Ztest      CI95.lo      CI95.hi
log(lambda)  4.80188  0.03779645      4.727801      4.87596

>fit2<-ppm(p, ~x+y)
>fit2
Nonstationary Poisson process
Trend formula: ~x + y
Fitted coefficients for trend formula:
(Intercept)      x      y
4.90100692  0.06123793 -0.08741010
      Estimate      S.E.      Ztest      CI95.lo      CI95.hi
(Intercept)  4.90100692  0.20052921      4.50797689      5.29403696
      x      0.06123793  0.05561531      -0.04776607      0.17024194
```

```

y      -0.08741010 0.05563312      -0.19644900  0.02162881
>fit3<-ppm(p, ~polynom(x,y,2))
>fit3
Nonstationary Poisson process
Trend formula: ~polynom(x, y, 2)
Fitted coefficients for trend formula:
      (Intercept)      polynom(x, y, 2)[x]      polynom(x, y, 2)[y]
      4.055664440      0.356092936      0.282453887
polynom(x, y, 2)[x^2] polynom(x, y, 2)[x.y] polynom(x, y, 2)[y^2]
      0.020160602      -0.141503564      -0.009069363

```

	Estimate	S.E.	Ztest	CI95.lo	CI95.hi
(Intercept)	4.055664440	1.04226849	2.0128557	6.0984731	6.0984731
polynom(x, y, 2)[x]	0.356092936	0.48335340	-0.5912623	1.3034482	1.3034482
polynom(x, y, 2)[y]	0.282453887	0.54326725	-0.7823303	1.3472381	1.3472381
polynom(x, y, 2)[x^2]	0.020160602	0.08887303	-0.1540273	0.1943485	0.1943485
polynom(x, y, 2)[x.y]	-0.141503564	0.09049787	-0.3188761	0.0358690	0.0358690
polynom(x, y, 2)[y^2]	-0.009069363	0.08948644	-0.1844596	0.1663208	0.1663208

```

>fitx<-ppm(p,~x)
>fitx

```

Nonstationary Poisson process

Trend formula: ~x

Fitted coefficients for trend formula:

```

(Intercept)      x
4.66169536 0.06120652

```

	Estimate	S.E.	Ztest	CI95.lo	CI95.hi
(Intercept)	4.66169536	0.13174900	4.40347206	4.919919	4.919919
x	0.06120652	0.05560689	-0.04778098	0.170194	0.170194

Funkcijų Akaikės kriterijai:

```

> AIC(fit1,fit2,fit3,fitx)

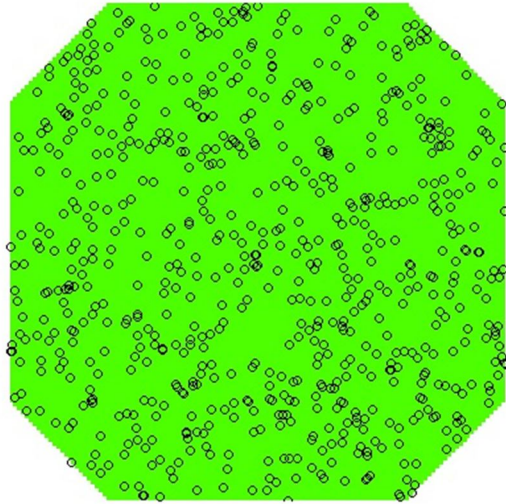
```

	df	AIC
fit1	1	-5320.633
fit2	3	-5309.720
fit3	6	-5306.256
fitx	2	-5309.249

Matome, jog mažiausias Akaikės informacijos kriterijus yra funkcijos *fit1*, tai yra kuomet trendas duomenims nėra būdingas.

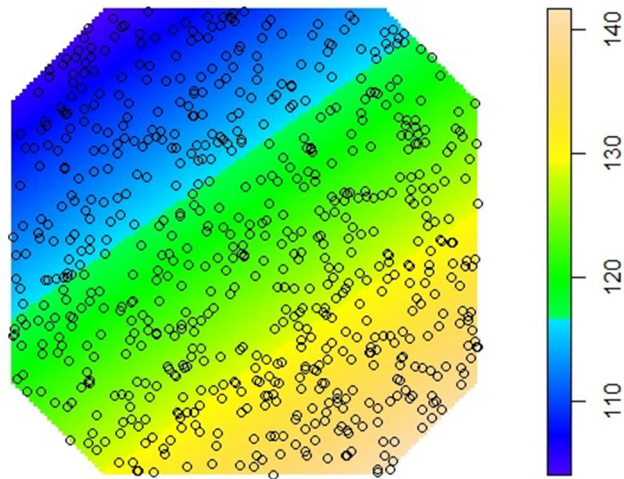
Pasinaudoję paklaidų analize patikrinsime, ar tikrai ši funkcija geriausiai aprašo tankį. Naudosime tris tinkamiausias tankio funkcijas pagal Akaikės informacijos kriterijų, t.y. *fit1*, *fit2* ir *fitx*. Komanda *plot(predict)* prognozuoja taškų išsidėstymą (spalvinė gama), o komanda *plot(p,add=T)* ant spalvų uždeda tikrus proceso taškus. Taip galime palyginti prognozuotus ir tikrus duomenis.

predict(fit1)



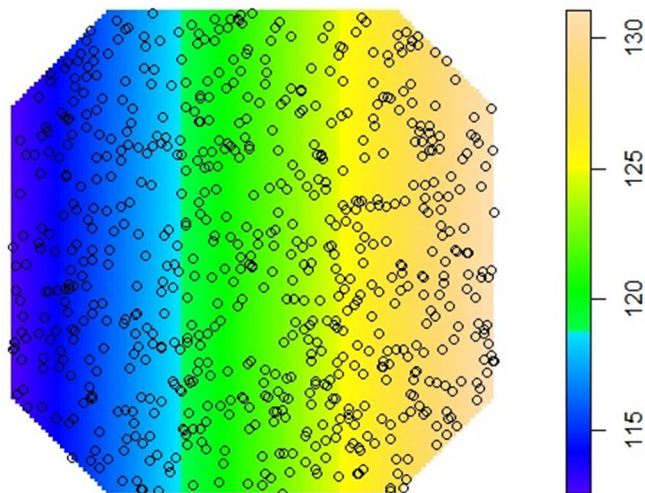
27 pav. Tankio funkcijos *fit1* prognozė

predict(fit2)



28 pav. Tankio funkcijos *fit2* prognozė

predict(fitx)



29 pav. Tankio funkcijos *fitx* prognozė

Iš grafikų matome, jog prognozuojamas (spalvinė gama) taškų išsidėstymas nesant trendui yra visoje srityje vienodas kaip ir buvo nurodyta. Prognozė pirmojo laipsnio trendo atveju nėra tiksli, nes proceso srities viršutinio kairiojo krašto prognozė neatspindi tikrųjų taškų išsibarstymo. Taip pat netiksliai prognozuojama su trendo funkcija pagal x koordinatę atveju. Viršutinis dešinysis srities kraštas prognozuojamas itin netiksliai.

Analizuosime keturių tipų grafiškai vaizduojamas paklaidas.

```
> diagnose.ppm(fit1,main="Keturių tipų paklaidos funkcijos fit1")
```

Model diagnostics (raw residuals)

Diagnostics available:

four-panel plot

mark plot

smoothed residual field

x cumulative residuals

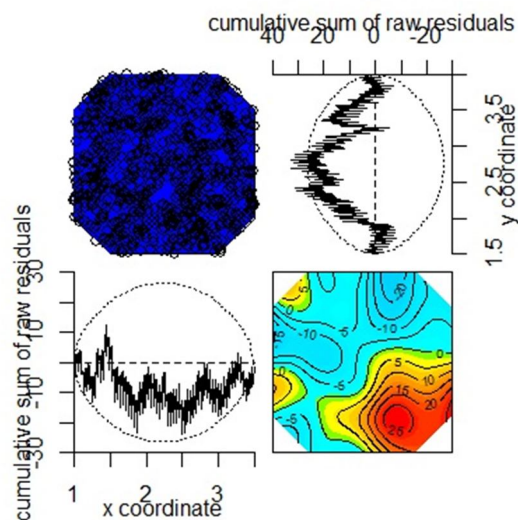
y cumulative residuals

sum of all residuals

sum of raw residuals in entire window = -0.4978

range of smoothed field = [-24.24,28.22]

Keturių tipų paklaidos funkcijos fit1



30 pav. Keturių tipų paklaidos funkcijos fit1

```
> diagnose.ppm(fit2,main="Keturių tipų paklaidos funkcijos fit2")
```

Model diagnostics (raw residuals)

Diagnostics available:

four-panel plot

mark plot

smoothed residual field

x cumulative residuals

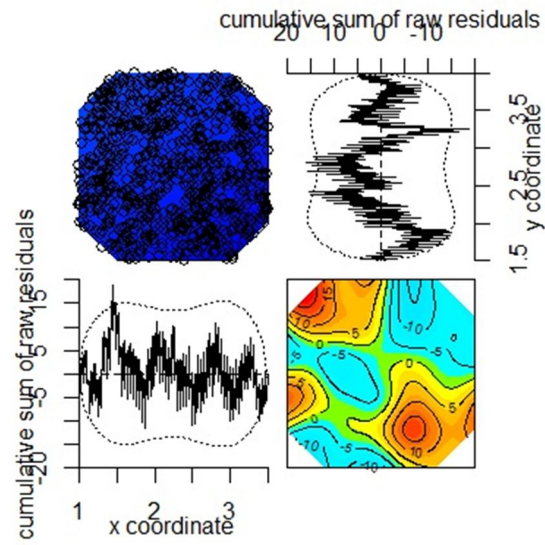
y cumulative residuals

sum of all residuals

sum of raw residuals in entire window = 1

range of smoothed field = [-16.15,23.76]

Keturių tipų paklaidos funkcijos fit2



31 pav. Keturių tipų paklaidos funkcijos fit2

```
> diagnose.ppm(fitx,main="Keturių tipų paklaidos funkcijos fitx")
```

Model diagnostics (raw residuals)

Diagnostics available:

four-panel plot

mark plot

smoothed residual field

x cumulative residuals

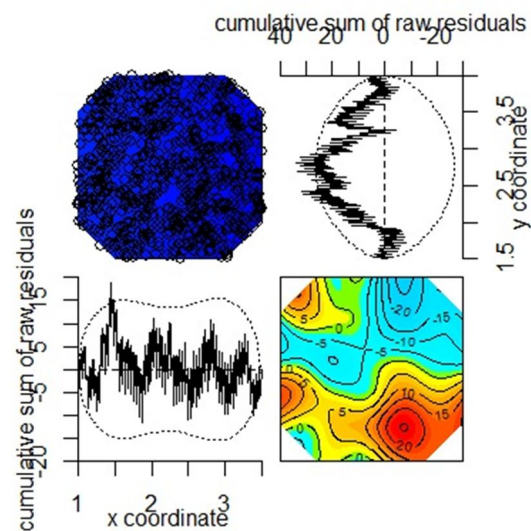
y cumulative residuals

sum of all residuals

sum of raw residuals in entire window = 1

range of smoothed field = [-27.02,25.22]

Keturių tipų paklaidos funkcijos fitx



32 pav. Keturių tipų paklaidos funkcijos fitx

Iš programos pateiktų grafikų matome, jog funkcija *fit1* yra labiausiai tinkama taškiniam procesui, nes sugludintų paklaidų suma yra mažiausia, t.y. -0.4978. Palyginus funkcijų *fit2* bei *fitx* liekanų sumas, jos abi yra lygios 1, taip kaip ir šių funkcijų Akaikės kriterijai skiriasi labai nežymiai. Vis dėlto, arčiausiai nulio liekanų suma yra funkcijos *fit1*, todėl priimame, kad duomenų tankis visoje srityje yra vienodas, kitaip sakant, geriausiai tankį aprašanti funkcija yra *fit1*.

Pagal nustatytą geriausiai tankį aprašančią funkciją, medžių sergančių šaknų puvinimo liga erdviniam taškiniam procesui priglodinsime tam tikrą ETP modelį. Programa automatiškai parinks geriausiai tinkantį duomenims modelį, nurodžius įvertintą tankio funkciją.

```

> fitmodel<-ppm(p,~1)
> summary(fitmodel)
Point process model
Fitting method: maximum pseudolikelihood (Berman-Turner approximation)
Model was fitted analytically
Call:
ppm(Q = p, trend = ~1)
Edge correction: "border"
 [border correction distance r = 0 ]
-----
Quadrature scheme = data + dummy + weights
Data pattern:
Planar point pattern: 700 points
Average intensity 122 points per square unit
Window: polygonal boundary
single connected closed polygon with 8 vertices
enclosing rectangle: [1, 3.5]x[1.5, 4]units
Window area = 5.75 square units
Dummy quadrature points:
( 64 x 64 grid, plus 4 corner points)
Planar point pattern: 3784 points
Average intensity 658 points per square unit
Window: polygonal boundary
single connected closed polygon with 8 vertices
enclosing rectangle: [1, 3.5]x[1.5, 4]units
Window area = 5.75 square units
Quadrature weights:
(counting weights based on 64 x 64 array of rectangular tiles)
All weights:
  range: [0, 0.00153]  total: 5.75
Weights on data points:
  range: [0, 0.000763]  total: 0.503
Weights on dummy points:
  range: [0.000381, 0.00153]  total: 5.25
-----
FITTED MODEL:
Stationary Poisson process
---- Intensity: ----
Uniform intensity:
[1] 121.7391
      Estimate   S.E. Ztest CI95.lo CI95.hi
log(lambda) 4.80188 0.03779645  4.727801 4.87596
----- gory details -----
Fitted regular parameters (theta):
log(lambda)
  4.80188

```

```
Fitted exp(theta):  
log(lambda)  
121.7391
```

Kaip matome, nurodžius tankio funkciją *fit1*, duomenims buvo priglodintas stacionarus Puasono procesas, todėl galime patvirtinti visas prieš tai padarytas išvadas apie intensyvumo vienaarūšiškumą bei taškų atsitiktinumą. Modelio tinkamumas procesui vertinamas naudojant tuos pačius vertinimo metodus bei testus, kurie jau buvo pademonstruoti, kaip ir vertinant tankio funkciją (*Monte Carlo* testas, paklaidų analizė ir t.t.). Kadangi taškinio vaizdo duomenis generavo programa, o R programa dažniausiai, nenurodant tam tikro šabloninio proceso generavimui, sukuria atsitiktinius duomenis, kurie yra būdingi Puasono tipo procesams, šio modelio tinkamumo neanalizuosime, o laikysime, kad homogeninis Puasono procesas geriausiai aprašo sergančių medžių duomenis. Tai patvirtina ir išvados, atliktos analizės metu.

IV.3. Žymėto taškinio proceso analizė

Analizuojamo erdvinio taškinio proceso duomenims pridėsime žymes. Parenkame kategorines žymes, t.y. kokybines, tris medžių rūšis: beržas, ąžuolas, klevas. Kadangi priskiriant jau sukurtam taškiniam procesui žymes reikia žinoti tikslų taškų skaičių, atliekame trumpą apžvalgą.

```
> summary(p)  
Planar point pattern: 700 points  
Average intensity 122 points per square unit  
Window: polygonal boundary  
single connected closed polygon with 8 vertices  
enclosing rectangle: [1, 3.5]x[1.5, 4]units  
Window area = 5.75 square units
```

Žinodami, jog turime 700 taškų, jiems priskiriame žymes.

```
> zym1 <- sample(c("Beržas", "Ąžuolas", "Klevas"), 700, replace=TRUE)
```

Žymes turime apsirašyti kaip faktorių.

```
> zym1 <- factor(zym1)  
> zym1 <- factor(zym1)  
> marks(p) <- zym1  
> p  
marked planar point pattern: 700 points  
multitype, with levels = Ąžuolas Beržas Klevas  
window: polygonal boundary  
enclosing rectangle: [1, 3.5] x [1.5, 4] units
```

Matome, jog taškams buvo priskirtos žymės. Patikriname ar tikrai taškinis procesas suprantamas kaip multi-tipo.

```
> is.multitype(p)  
[1] TRUE
```

Norėdami pamatyti taškams priskirtas žymių kategorijas, duomenis apsirašome kaip masyvą.

```
> as.data.frame(p)  
x y marks
```

```

1 3.303922 2.755815 Ažuolas
2 2.785509 2.975879 Klevas
3 2.765666 1.766998 Beržas
4 3.027418 2.246466 Beržas

```

<...>

```

697 1.400789 2.054682 Ažuolas
698 1.544486 3.953658 Klevas
699 2.959988 1.864459 Klevas
700 2.575295 2.072849 Beržas

```

Atliekame multi-tipo taškinio proceso apžvalgą.

```

> summary(p)
Marked planar point pattern: 700 points
Average intensity 122 points per square unit
Multitype:
  frequency  proportion  intensity
Ažuolas    242    0.346    42.1
Beržas     204    0.291    35.5
Klevas     254    0.363    44.2
Window: polygonal boundary
single connected closed polygon with 8 vertices
enclosing rectangle: [1, 3.5]x[1.5, 4]units
Window area = 5.75 square units

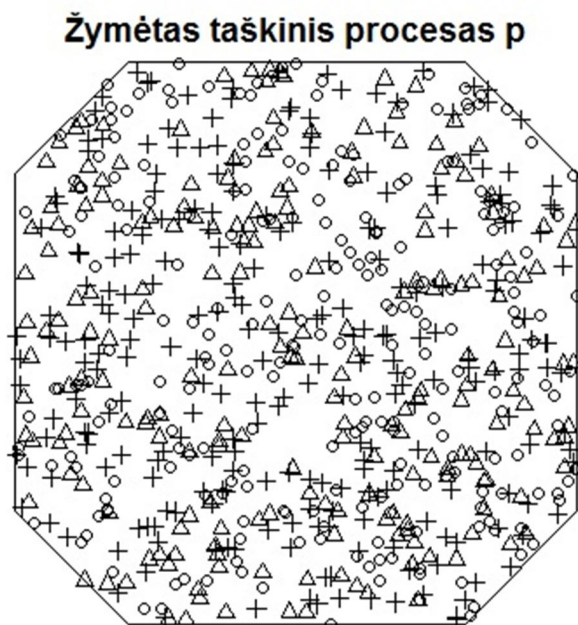
```

Matome, jog 242 taškams buvo priskirta žymė „Ažuolas“, 204 taškams žymė „Beržas“ ir 254 taškams žymė „Klevas“. Grafiškai atvaizduojame žymėtą taškinį procesą.

```

> plot(p)
Ažuolas Beržas Klevas
  1     2     3

```



33 pav. Žymėtas taškinis procesas p

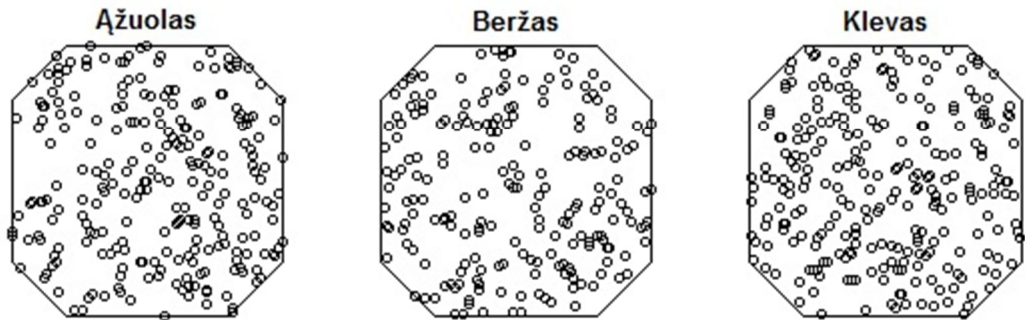
Matome, jog kiekviena žymės kategorija atvaizduojama skirtingu simboliu. Atskiriame žymes, kad geriau pamatytume kiekvienos žymės pasiskirstymą.

```

> plot(split(p))

```

Proceso p žymių pasiskirstymas

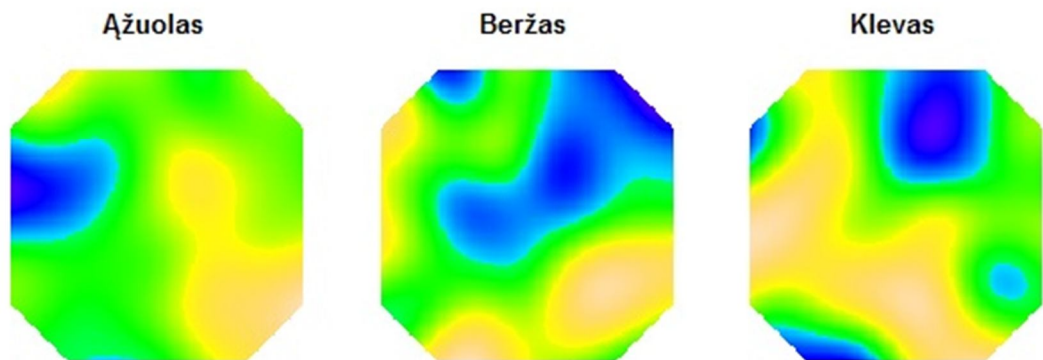


34 pav. Žymių pasiskirstymas

Atskirų žymių intensyvumo pasiskirstymą atvaizduojame grafiškai.

```
> plot(density(split(p)), ribbon = FALSE, main="Žymių intensyvumo pasiskirstymas")
```

Žymių intensyvumo pasiskirstymas



35 pav. Žymių intensyvumo pasiskirstymas

Apžvelgsime artimiausių kaimynų, esančių per 0.1 atstumą, žymes. Pateiksime pirmųjų penkiolikos taškų kaimynus.

```
> M <- marktable(p, R = 0.1)
```

```
> M[1:15, ]
```

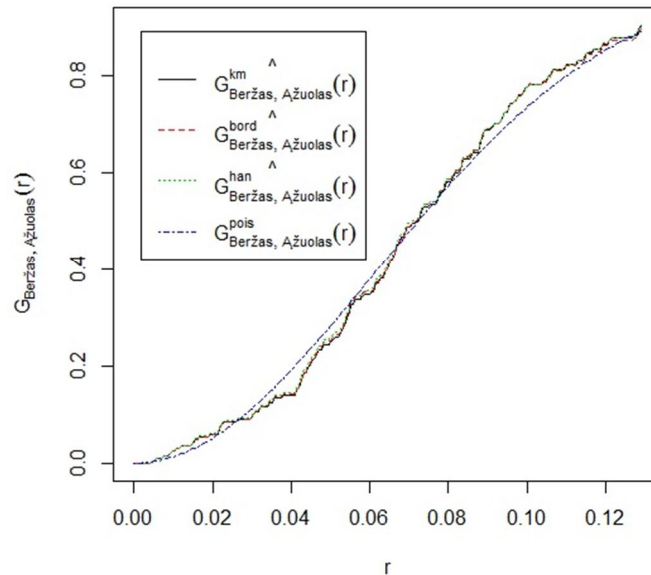
point	mark		
	Ažuolas	Beržas	Klevas
1	0	0	4
2	4	3	1
3	3	2	1
4	1	2	1
5	0	1	1
6	1	1	0
7	2	0	3
8	0	0	1
9	0	3	3
10	1	0	1

11	0	1	0
12	2	1	2
13	2	1	2
14	2	1	0
15	5	2	2

Matome, jog artimiausi kaimynai turi visų trijų tipų žymes, „Beržo“ tipo kaimynų yra mažiausiai. Apskaičiuosime apibendrinančią atstumų kryžminę funkciją G_{ij} kelioms medžių poroms.

```
> plot(Gcross(p,"Beržas","Ąžuolas"))
      meaning
km      Kaplan-Meier estimate of G["Beržas", "Ąžuolas"](r)
rs      border corrected estimate of G["Beržas", "Ąžuolas"](r)
han     Hanisch estimate of G["Beržas", "Ąžuolas"](r)
theo    theoretical Poisson G["Beržas", "Ąžuolas"](r)
```

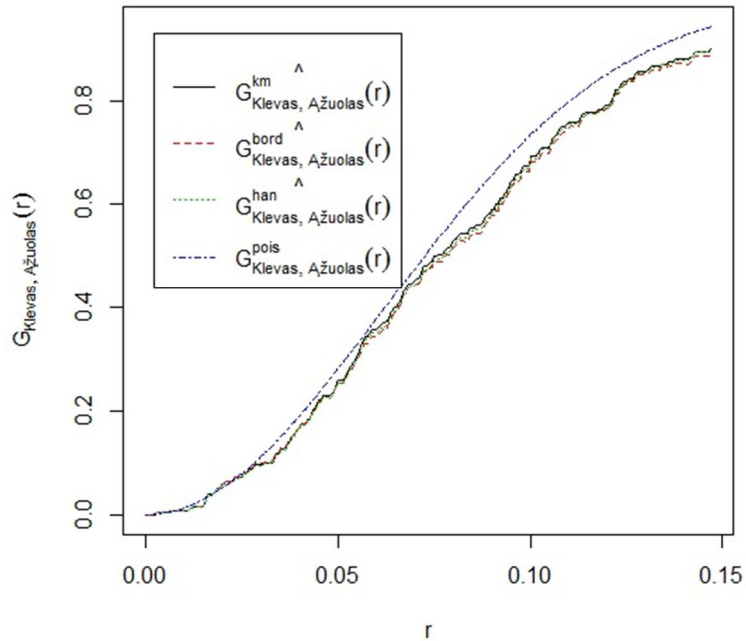
Kryžminė funkcija G žymių porai Beržas ir Ąžuolas



36 pav. Kryžminė G_{ij} funkcija žymių porai „Beržas“ ir „Ąžuolas“

```
> plot(Gcross(p,"Klevas","Ąžuolas"),main="Kryžminė funkcija G žymių porai Klevas ir Ąžuolas")
      meaning
km      Kaplan-Meier estimate of G["Klevas", "Ąžuolas"](r)
rs      border corrected estimate of G["Klevas", "Ąžuolas"](r)
han     Hanisch estimate of G["Klevas", "Ąžuolas"](r)
theo    theoretical Poisson G["Klevas", "Ąžuolas"](r)
```

Kryžminė funkcija G žymių porai Klevas ir Ažuolas



37 pav. Kryžminė G_{ij} funkcija žymių porai „Klevas“ ir „Ažuolas“

Atvaizduosime kryžmines G_{ij} ir J_{ij} funkcijas visoms žymių poroms, naudodami komandą $plot(alltypes(p, "G"))$ ir $plot(alltypes(p, "J"))$. Funkcijų grafikai pateikti atitinkamai *priede nr. 2, 38 pav.* Kryžminė G_{ij} funkcija visoms žymių poroms ir *priede nr. 3, 39 pav.* Kryžminė J_{ij} funkcija visoms žymių poroms.

Iš J_{ij} ir G_{ij} funkcijų grafikų, galime daryti išvadą, jog galima įžiūrėti tam tikrą žymių priklausomumą mažu atstumu tarp taškų. Tačiau, kuomet atstumas didesnis už 0.02, J_{ij} funkcijos grafikai rodo visišką žymių nepriklausomybę. Jei $J_{ij}=1$, tuomet sakome, jog žymės yra priklausomos viena nuo kitos.

Grafiškai pavaizduosime $G_i(r)$ pasiskirstymo funkciją nuo fiksuotos žymės taškų iki kito artimiausio bet kokio tipo taško. Grafikas pateiktas *priede nr. 4, 40 pav.* Kryžminė $G_i(r)$ funkcija nuo fiksuotos žymės iki kito bet kokio taško.

Matome, jog pasiskirstymo funkcijos yra panašios. Iš visų kryžminių atstumų funkcijų grafikų galime daryti išvadą, jog medžių pavadinimai buvo taškams priskirti nepriklausomai ir atsitiktinai.

Duomenims priglodinsime modelį, nurodydami, jog trendo funkcija yra žymės.

```
> ppm(p, ~marks)
Stationary multitype Poisson process
Possible marks:
Ažuolas Beržas Klevas
Trend formula: ~marks
Intensities:
beta_Ažuolas beta_Beržas beta_Klevas
39.79778 41.36189 40.31915
Estimate S.E. Ztest CI95.lo CI95.hi
```

```
(Intercept) 3.68381121 0.06608186 3.5542931 3.8133293
marksBeržas 0.03854867 0.09256615 -0.1428777 0.2199750
marksKlevas 0.01301537 0.09315126 -0.1695577 0.1955885
```

Matome, jog duomenims buvo priglodintas multi-tipo stacionarus Puasono procesas.

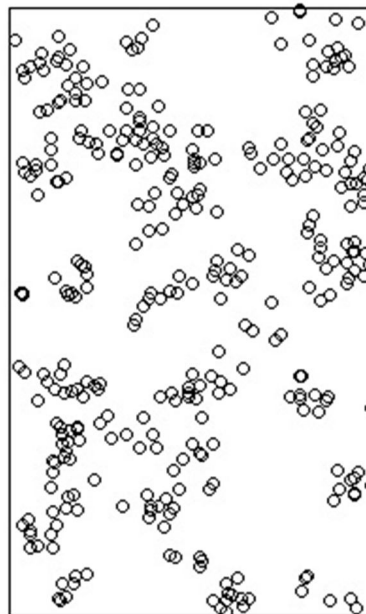
IV.4. Klasterinio ETP analizė

Anksčiau atliktos analizės duomenis, R programa sukūrė atsitiktinai, nenurodant programai jokio proceso, kuris tuos duomenis sugeneruotų. Dažniausiai nenurodant jokio proceso, R programa sukuria duomenis, kurie atitinka homogeninį Puasono procesą. Klasterinius duomenis taip pat galima sugeneruoti R programos pagalba, tačiau jau nurodysime, jog duomenys būtų sugeneruoti pagal *NeymanScott* klasterinį procesą. Tai reiškia, jog analizės pabaigoje glodinti duomenims tam tikro proceso modelio neberekės, nes šie duomenys jau bus sukurti pagal *NeymanScott* procesą.

Įsivaizduokime, jog sukurtas taškinis vaizdas atvaizduos šaknų puvinimo liga sergančius medžius, tik kito miško teritorijoje, pvz. Vėžaičių gyvenvietėje esančiame miške. Šį kartą apibrėšime ilgio vienetus – kilometrus. Taigi, turėsime ETP X , kurio stebiniai bus sergantys medžiai, o sritis – Vėžaičių miško plotas, lygus 15 km^2 .

```
> nclust <- function(x0, y0, radius, n) {return(runifdisc(n, radius, centre=c(x0, y0))) }
> X<-(rNeymanScott(5, 0.1, nclust, radius=0.2,n=5,win=owin(c(0,3),c(0,5))))
> unitname(X)<-c("kilmetres", "kilometres")
> any(duplicated(X))
[1] FALSE
> plot(X, main="Erdvinis taškinis procesas X")
```

Erdvinis taškinis procesas X



41 pav. Erdvinis taškinis procesas X

```
> summary(X)
```

```
Planar point pattern: 343 points
```

```
Average intensity 22.9 points per square kilometres
```

```
Window: rectangle = [0, 3]x[0, 5]kilometres
```

```
Window area = 15 square kilometres
```

```
Unit of length: 1 kilometres
```

Sukūrėme klasterinį procesą X (ETP X duomenų lentelė patalpinta el. laikmenoje, pavadinimu „duomenys_X“, txt formatu). Iš grafinio ETP X vaizdo matome, jog taškų klasteriai aiškiai pastebimi, o vidutinis taškų intensyvumas yra 22.9 taškai kvadratiniam kilometre. Langą parinkome stačiakampio formos, kurio x ašis lygi 3 km, o y ašis lygi 5 km. Intensyvumą ištirsime kvadratų skaičiavimo technika.

```
> Q<- quadratcount(X,nx=3, ny=5)
```

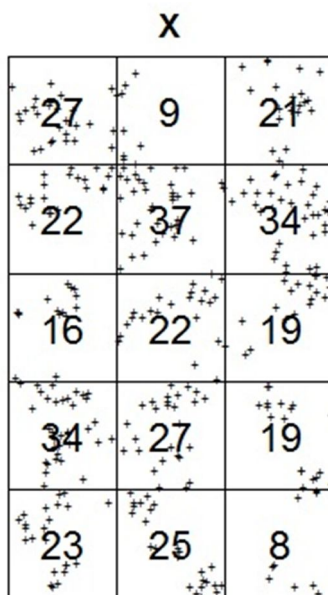
```
> Q
```

```
      x  
y [0,1] (1,2] (2,3]  
(4,5] 27  9  21  
(3,4] 22 37  34  
(2,3] 16 22  19  
(1,2] 34 27  19  
[0,1] 23 25  8
```

```
> plot(Q)
```

```
> plot(X,cex=0.5,pch="+")
```

```
> plot(Q,add=T,cex=1.5)
```



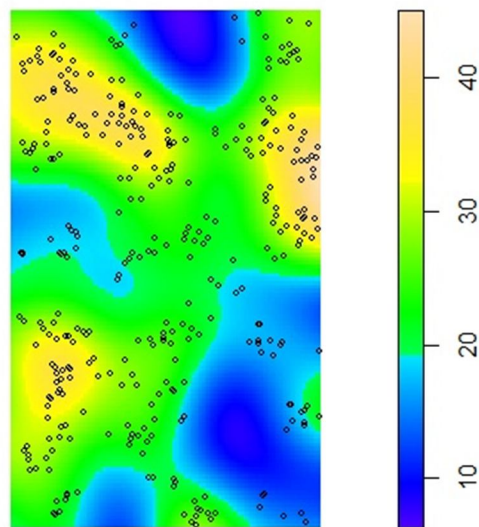
42 pav. Taškų skaičius kvadratuose

```
> den<-density(X)
```

```
> plot(den,main="Kernel'io intensyvumo įvertinys")
```

```
> plot(X,add=T,cex=0.5)
```

Kernel'io intensyvumo įvertinys

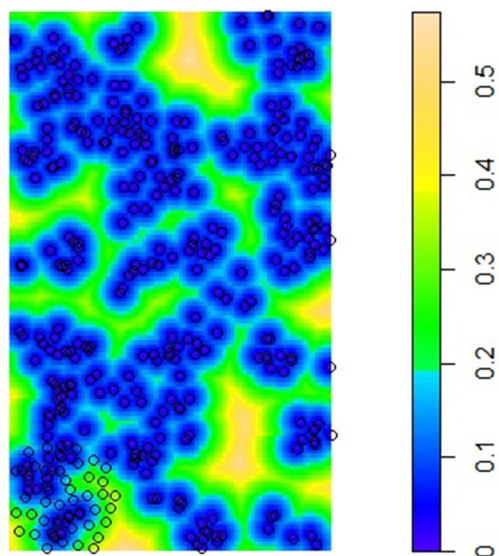


43 pav. Kernel'io intensyvumo įvertinys

Iš kvadratų skaičiavimo matome, jog taškų skaičius kvadratuose skiriasi gana stipriai. Mažiausias taškų skaičius kvadrate yra 8, didžiausias 37. Kernel'io intensyvumo įvertinys, pavaizduotas grafiškai, puikiai parodo mūsų tiramos srities vietas, kur taškų yra labai mažai, o kur susidarę taškų klasteriai. Panagrinėsime atstumus tarp taškų. Visų pirma pavaizduosime tuščius erdvinis atstumus.

```
> plot(distmap(X),main="Tuščių erdvių atstumai")  
> plot(cells,add=T)  
> plot(X,add=T)
```

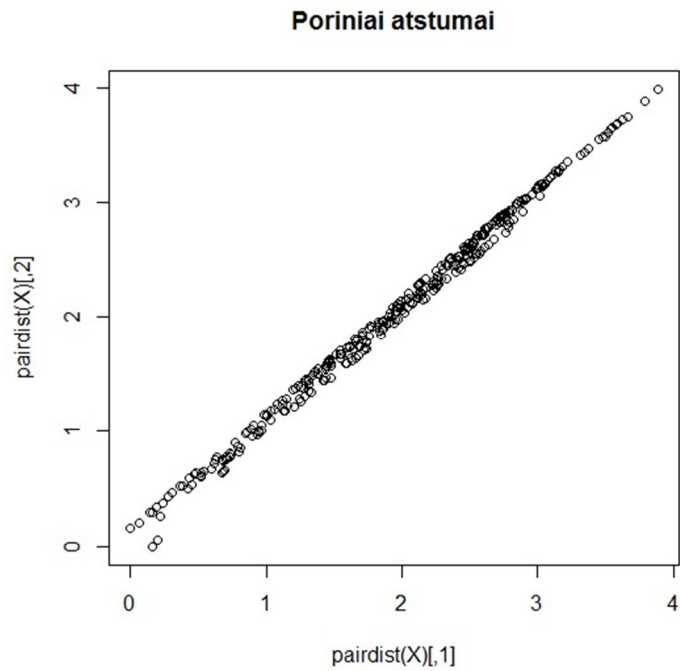
Tuščių erdvių atstumai



44 pav. Tuščių erdvių atstumai

Pavaizduosime atstumus tarp visų taškų porų.

```
> plot(pairdist(X),main="Poriniai atstumai")
```

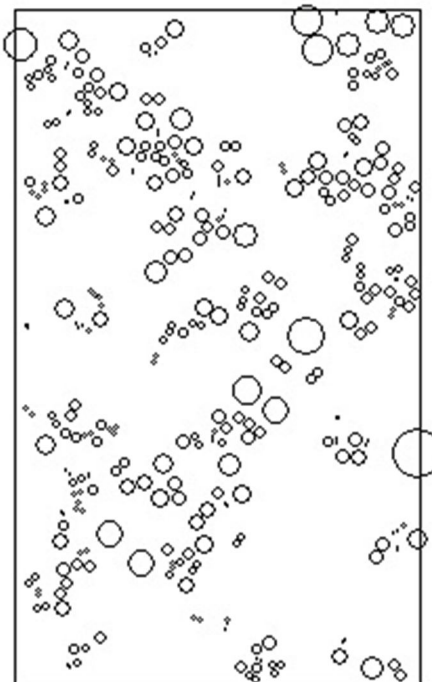


45 pav. Poriniai atstumai

Pavaizduosime artimiausio kaimyno atstumus.

```
> plot(X%mark%(nndist(X)/2),markscale=1,main="Stienen'o diagrama")
```

Stienen'o diagrama



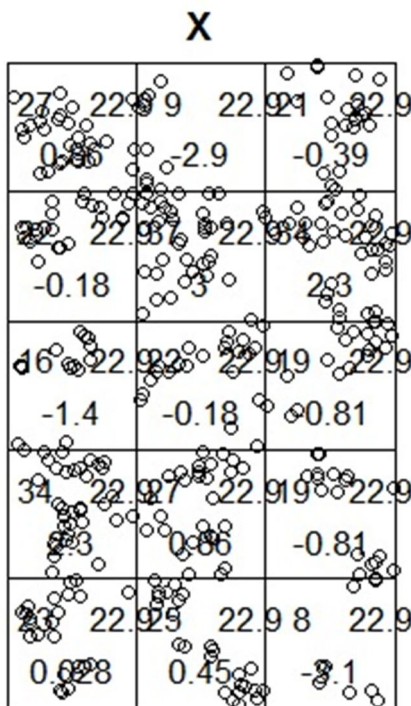
46 pav. Artimiausio kaimyno atstumai

Apskaičiuosime ir pavaizduosime grafiškai visas atstumų funkcijas. Grafikas pateiktas *priede nr. 5, 47 pav.* ETP X atstumų funkcijos.

Iš pateiktų atstumų funkcijų grafikų matome, jog visos funkcijos skiriasi nuo teorinių Puasono funkcijų. Tuščių erdvių atstumų funkcija F ir visos F funkcijos korekcijos eina žemiau Puasono teorinės F funkcijos, o tai reiškia kad duomenims būdingi klasteriai. Artmiausio kaimyno atstumo funkcija G eina virš Puasono teorinės G funkcijos, o tai taip pat parodo, jog duomenims būdingi klasteriai. Porinių atstumų funkcija K taip pat eina virš Puasono teorinės funkcijos. Funkcijos J grafikas, kuris Puasono proceso atveju eitų apie 1, staigiai leidžiasi žemyn, didėjant atstumui tarp taškų. Tai parodo duomenims būdingą klasterizaciją.

Patikrinsime, ar mūsų tiriamas sergančių medžių ETP tenkina nulinę hipotezę.

```
> M<-quadrat.test(X,nx=3,ny=5)
> M
      Chi-squared test of CSR using quadrat counts
data: X
X-squared = 42.9329, df = 14, p-value = 8.781e-05
Quadrats: 3 by 5 grid of tiles
> plot(X,cex=1)
> plot(M,add=T,cex=1.1)
```

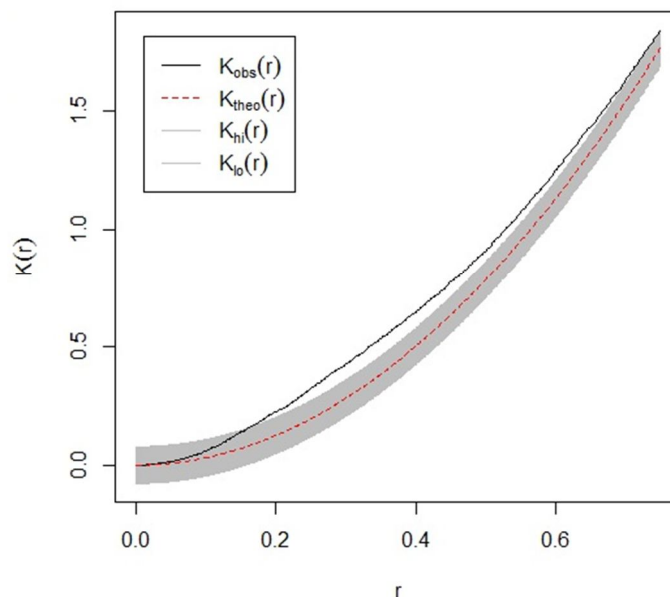


48 pav. χ^2 testas

Programos apskaičiuota Chi kvadrato reikšmė 42.9329 gerokai viršija reikšmę, kurią randame Chi kvadrato testo reikšmių lentelėje 23.68, kur laisvės laipsnių turime 14, o reikšmingumo lygmenį pasirenkame $\alpha = 0.05$. p reikšmė taip pat yra mažesnė už 0.05, todėl galime daryti išvadą, jog Chi kvadrato testas atmeta nulinę hipotezę. Atliksime *Monte Carlo* testą.

```
> KE<- envelope(X, Kest, nsim = 59,rank=1, global=TRUE)
> plot(KE, main="Monte Carlo testas")
```

Monte Carlo testas



49 pav. Monte Carlo testas ETP X

Matome, jog K funkcija kerta kritines ribas ir nesutampa su teorine Puasono K funkcijos reikšme. Todėl galime patvirtinti, jog ETP X atmeta nulinę hipotezę, o tai reiškia, jog taškai nėra nepriklausomai ir vienodai pasiskirstę.

Parinksime geriausiai ETP X tankį aprašančią funkciją. Parenkame keturias tankio funkcijas, iš kurių išrinksime geriausiai tinkančią.

- $\sim y$, trendas pagal y koordinatę;
- $\sim x$, trendas pagal x koordinatę;
- $\sim x+y$, pirmojo laipsnio trendas;
- $\sim x+y+I(x^2)+I(y^2)+I(x*y)$, antrojo laipsnio trendas.

```
> fit1<-ppm(X, ~y)
> fit1
Nonstationary Poisson process
Trend formula: ~y
Fitted coefficients for trend formula:
(Intercept)      y
2.98208210 0.05765747
      Estimate   S.E. Ztest  CI95.lo  CI95.hi
(Intercept) 2.98208210 0.11209904  2.76237202 3.2017922
y      0.05765747 0.03749818 -0.01583762 0.1311526

> fit2<-ppm(X, ~x)
> fit2
Nonstationary Poisson process
Trend formula: ~x
Fitted coefficients for trend formula:
```

```

(Intercept)      x
3.23139372 -0.06899264
      Estimate S.E. Ztest CI95.lo CI95.hi
(Intercept) 3.23139372 0.1053134 3.0249833 3.43780412
x -0.06899264 0.0624250 -0.1913434 0.05335812

> fit3<-ppm(X,~x+y)
> fit3
Nonstationary Poisson process
Trend formula: ~x + y
Fitted coefficients for trend formula:
(Intercept)      x      y
3.08379671 -0.06899879 0.05766056
      Estimate S.E. Ztest CI95.lo CI95.hi
(Intercept) 3.08379671 0.14401537 2.80153177 3.36606165
x -0.06899879 0.06242472 -0.19134898 0.05335141
y 0.05766056 0.03749839 -0.01583493 0.13115605

> fit4<-ppm(X, ~polynom(x,y,2))
> fit4
Nonstationary Poisson process
Trend formula: ~polynom(x, y, 2)
Fitted coefficients for trend formula:
(Intercept) polynom(x, y, 2)[x] polynom(x, y, 2)[y]
3.27460476 -0.55881152 0.19676854
polynom(x, y, 2)[x^2] polynom(x, y, 2)[x.y] polynom(x, y, 2)[y^2]
0.10749528 0.06538521 -0.04579857
      Estimate S.E. Ztest CI95.lo CI95.hi
(Intercept) 3.27460476 0.26654944 2.75217745 3.79703207
polynom(x, y, 2)[x] -0.55881152 0.26102500 * -1.07041112 -0.04721192
polynom(x, y, 2)[y] 0.19676854 0.16574794 -0.12809146 0.52162854
polynom(x, y, 2)[x^2] 0.10749528 0.07910228 -0.04754234 0.26253291
polynom(x, y, 2)[x.y] 0.06538521 0.04385010 -0.02055941 0.15132983
polynom(x, y, 2)[y^2] -0.04579857 0.02998083 -0.10455992 0.01296279

```

Apskaičiuosime Akaikės informacijos kriterijų, kad sužinotume, kuri funkcija geriausiai aprašo tankį.

```

> AIC(fit1,fit2,fit3,fit4)
df AIC
fit1 2 -1459.330
fit2 2 -1458.183
fit3 3 -1458.553
fit4 6 -1458.979

```

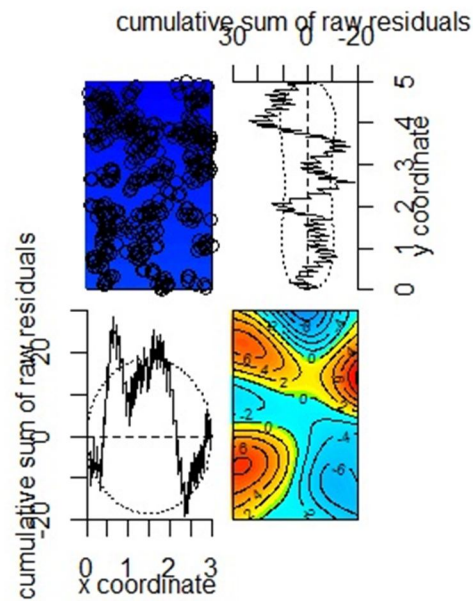
Matome, kad visos reikšėms panašios, tačiau mažiausia reikšmė yra funkcijos, kurios trendas yra y koordinatė. Grafiškai pavaizduosime visų funkcijų keturių tipų paklaidas.

```

> diagnose.ppm(fit1,main="Keturių tipų paklaidos funkcijos fit1")
Model diagnostics (raw residuals)
<...>
sum of raw residuals in entire window = -5.312e-06
range of smoothed field = [ -11.5,10.37 ]

```

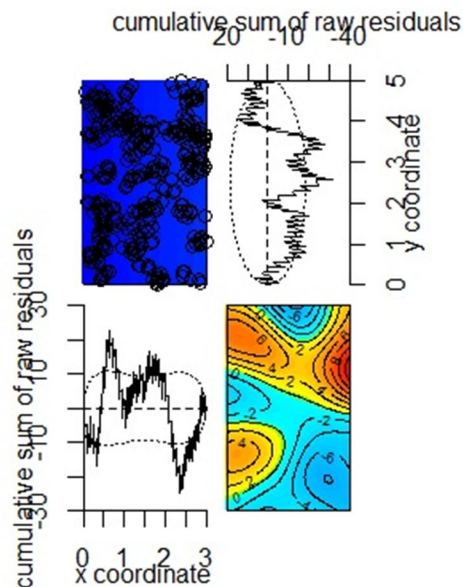
Keturių tipų paklaidos funkcijos fit1



50 pav. Keturių tipų paklaidos tankio funkcijos fit1

```
> diagnose.ppm(fit2,main="Keturių tipų paklaidos funkcijos fit2")
Model diagnostics (raw residuals)
<...>
sum of raw residuals in entire window = -3.161e-06
range of smoothed field = [ -8.417,13.16 ]
```

Keturių tipų paklaidos funkcijos fit2

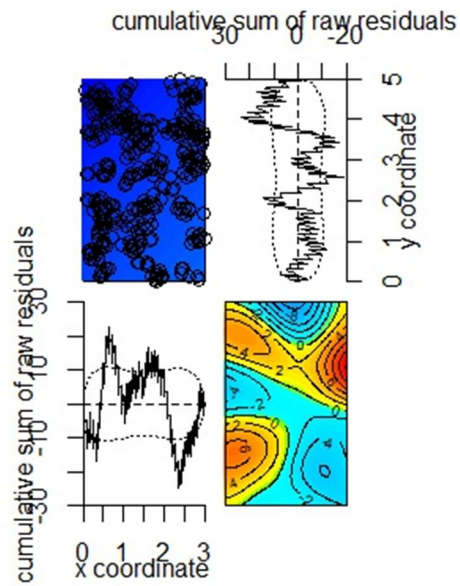


51 pav. Keturių tipų paklaidos tankio funkcijos fit2

```
> diagnose.ppm(fit3,main="Keturių tipų paklaidos funkcijos fit3")
Model diagnostics (raw residuals)
<...>
```

sum of raw residuals in entire window = $-5.77e-06$
 range of smoothed field = [-11.15, 12.05]

Keturių tipų paklaidos funkcijos fit3



52 pav. Keturių tipų paklaidos tankio funkcijos fit3

> diagnose.ppm(fit4, main="Keturių tipų paklaidos funkcijos fit4")

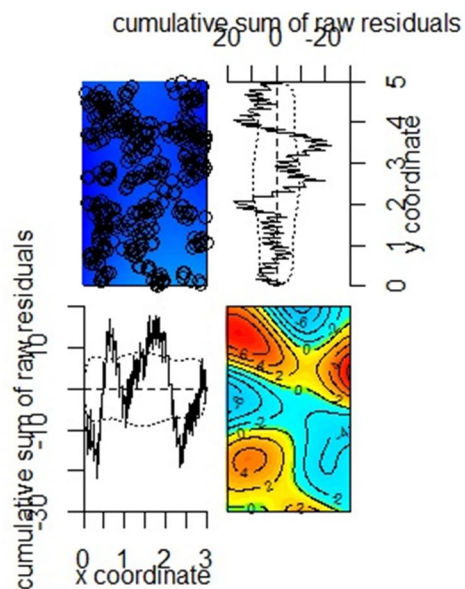
Model diagnostics (raw residuals)

<...>

sum of raw residuals in entire window = $-3.818e-12$

range of smoothed field = [-8.647, 8.542]

Keturių tipų paklaidos funkcijos fit4

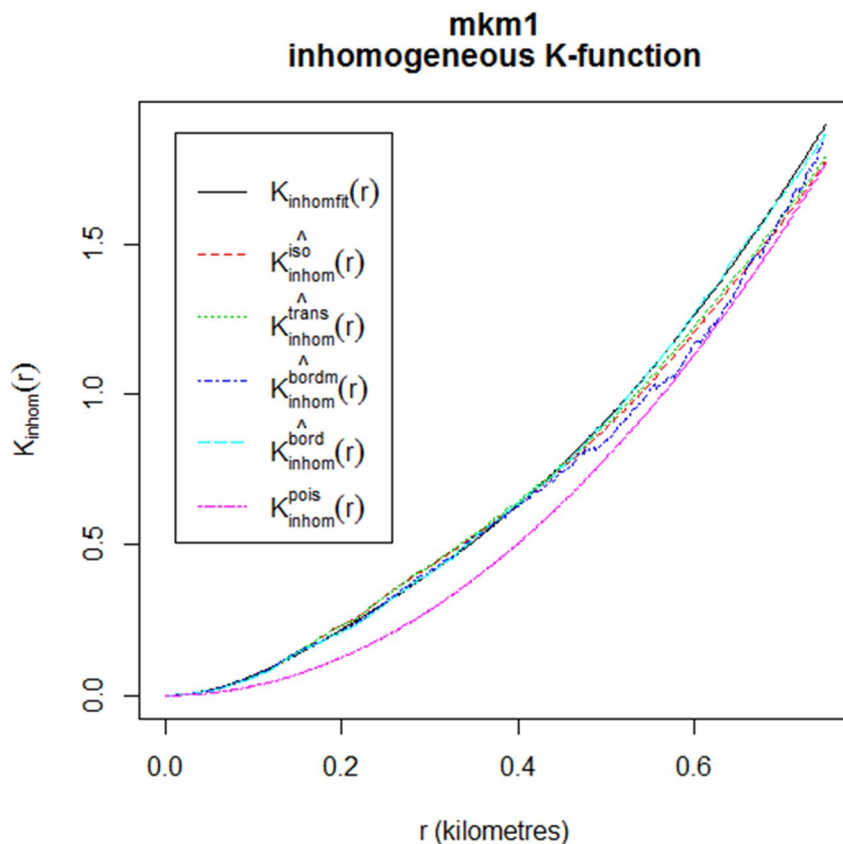


53 pav. Keturių tipų paklaidos tankio funkcijos fit4

Iš programos pateiktų suglodintų paklaidų sumų matome, jog mažiausia suma yra tankio funkcijos *fit4*, t.y. antrojo laipsnio trendo atvejis. Akaikės informacijos kriterijus nurodė geriausiai tinkančią tankio funkciją *fit1*, tačiau skirtumai tarp kriterijų buvo labai maži, o tuo tarpu suglodintų paklaidų sumos skyrėsi gana ženkliai. Todėl, minimalaus kontrasto metodu įvertinsime *K* funkciją naudodami šias abi tankio funkcijas *fit4* ir *fit1*.

Kadangi turime klasterinį ETP, *K* funkciją vertinsime naudodami *Thomas* metodą, kuris skirtas įvertinti būtent klasterinių duomenų funkciją *K*.

```
> mkm1 <- kppm(X, ~polynom(x,y,2), "Thomas")
> mkm1
Inhomogeneous cluster point process model
Fitted to point pattern dataset 'X'
Fitted using the inhomogeneous K-function
Trend formula: ~polynom(x, y, 2)
Fitted coefficients for trend formula:
      (Intercept) polynom(x, y, 2)[x] polynom(x, y, 2)[y]
      3.27460476   -0.55881152      0.19676854
polynom(x, y, 2)[x^2] polynom(x, y, 2)[x.y] polynom(x, y, 2)[y^2]
      0.10749528      0.06538521     -0.04579857
Cluster model: Thomas process
Fitted cluster parameters:
      kappa sigma2
7.774191339 0.007730034
Mean cluster size: [pixel image]
> plot(mkm1)
```

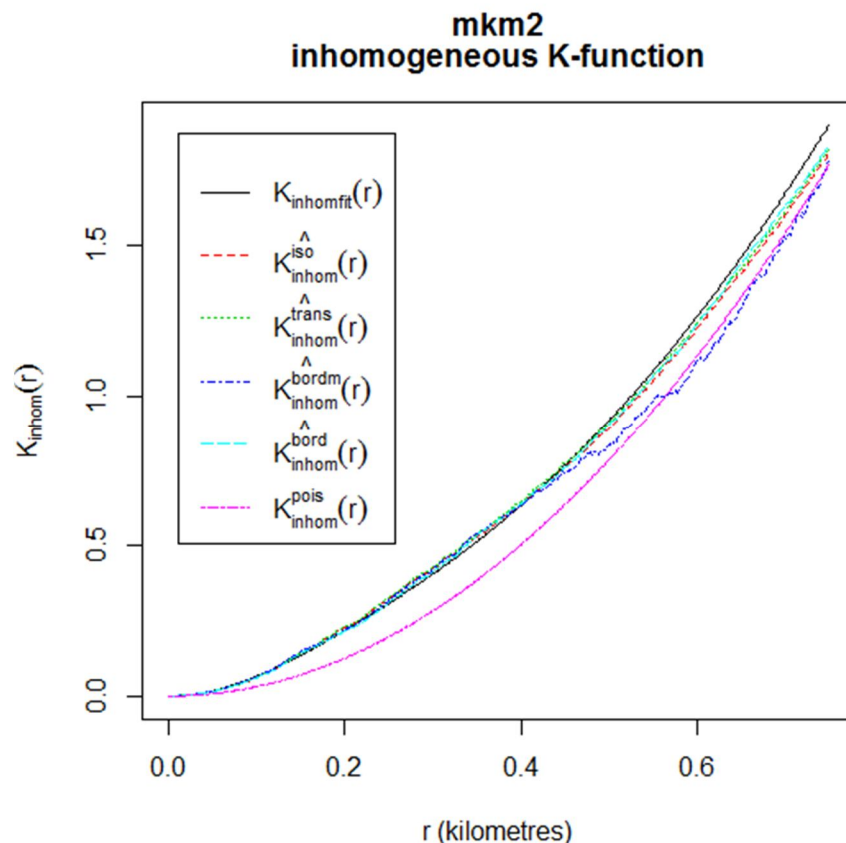


54 pav. *K* funkcija įvertinta mažiausių kvadratų metodu su tankio funkcija *fit4*

```

> mkm2 <- kppm(X, ~y, "Thomas")
> mkm2
Inhomogeneous cluster point process model
Fitted to point pattern dataset 'X'
Fitted using the inhomogeneous K-function
Trend formula: ~y
Fitted coefficients for trend formula:
(Intercept)      y
2.98208210 0.05765747
Cluster model: Thomas process
Fitted cluster parameters:
      kappa  sigma2
7.580630520 0.008252812
Mean cluster size: [pixel image]
> plot(mkm2)

```



55 pav. K funkcija įvertinta mažiausių kvadratų metodu su tankio funkcija *fit4*

Iš grafikų matome, jog abi įvertintos K funkcijos ir visos jų korekcijos eina virš Puasono teorinės nehomogeninės K funkcijos, ką matėme ir ankstesnėje analizėje. Tankio funkcijos *fit4* atveju, viena iš K funkcijos korekcijų greičiau priartėja prie teorinės K funkcijos. Galima sakyti, jog abi tankio funkcijos panašiai aprašo ETP X tankį, tačiau tankio funkcija *fit4* yra arčiau teorinės Puasono K funkcijos, todėl sakome, kad *fit4* yra geriau tinkama tankio funkcija.

ETP X taškams sukursime kiekybines žymes ir atliksime žvalgomąją analizę. Pridėsime taškams papildomos informacijos – medžių aukščius. Intervalą imame nuo 15 iki 30 metrų.

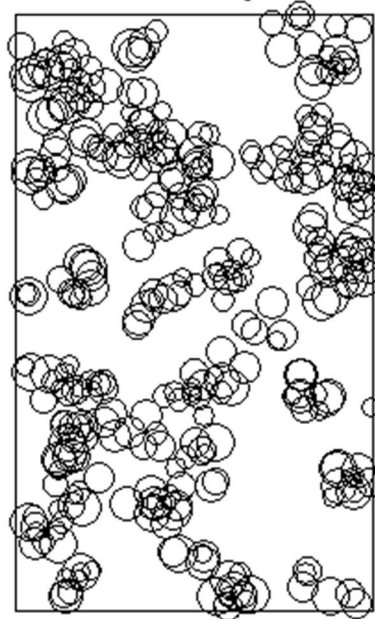
```

> zym2 <- sample(c(15:30), 343, replace=TRUE)

```

```
> marks(X)<-zym2
> plot(X, main="Žymėtas taškinis procesas X")
      15      20      25      30
0.0825889 0.1101185 0.1376482 0.1651778
```

Žymėtas taškinis procesas X



56 pav. Žymėtas ETP X

```
> as.data.frame(X)
      x      y marks
1 0.27893219 1.98509354 29
2 0.12230053 2.00094389 29
3 0.24142518 1.76668992 21
4 0.08885868 2.03846178 16
5 0.31013783 1.92430896 18
...
340 2.56120315 2.59218192 29
341 0.41126941 0.13419276 21
342 0.47660366 0.15584982 20
343 0.42832871 0.12251374 27
> summary(x)
```

*Marked planar point pattern: 343 points
Average intensity 22.9 points per square kilometres
marks are numeric, of type 'integer'*

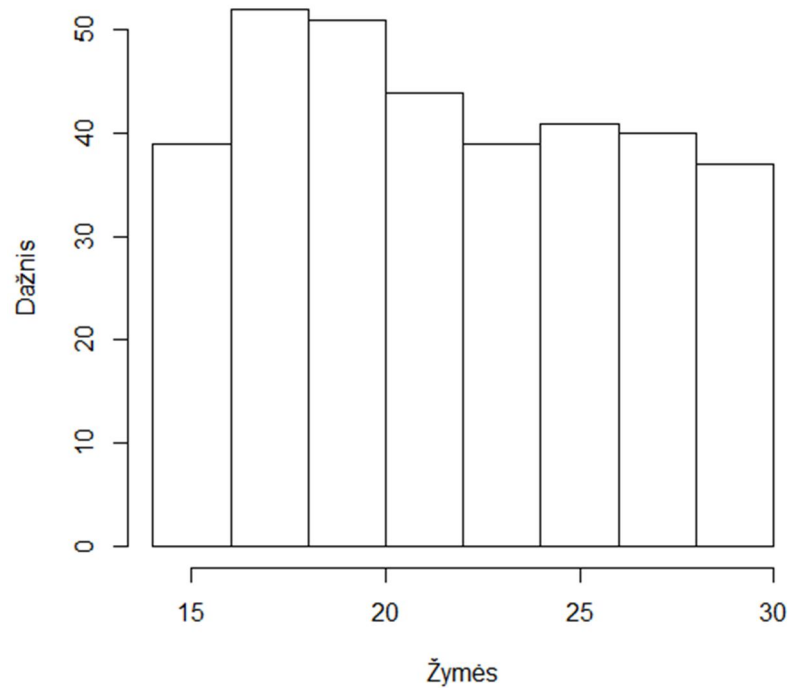
Summary:

```
  Min. 1st Qu. Median Mean 3rd Qu. Max.
 15.00 18.00 22.00 22.21 26.00 30.00
Window: rectangle = [0, 3]x[0, 5]kilometres
Window area = 15 square kilometres
Unit of length: 1 kilometres
```

Matome, jog programa atpažįsta žymes kaip sveikuosius skaičius, t.y. „integer“ tipo. Aukščio vidurkis yra 22.21 metro. Kad geriau pamatytume žymių pasiskirstymą, braižome histogramą.

```
> hist(marks(X), main="Žymių histograma", xlab="Žymės", ylab="Dažnis")
```

Žymių histograma

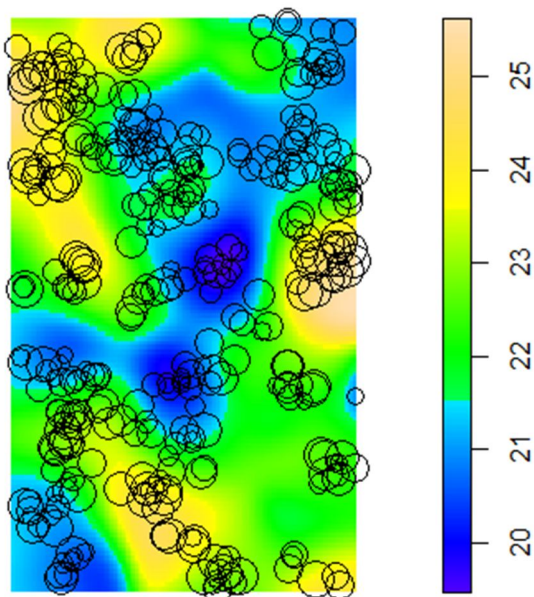


57 pav. Žymėto ETP X žymių pasiskirstymas

Matant histogramą galima sakyti, jog daugiausiai buvo sukurta 16-20 metrų aukščio žymių, taip pat žymių. Atvaizduosime erdvinį žymių trendą, kitaip sakant tankio funkciją erdvėje.

```
>plot(smooth.ppp(X),main="ETP X žymių tankis")  
>plot(X,add=T)
```

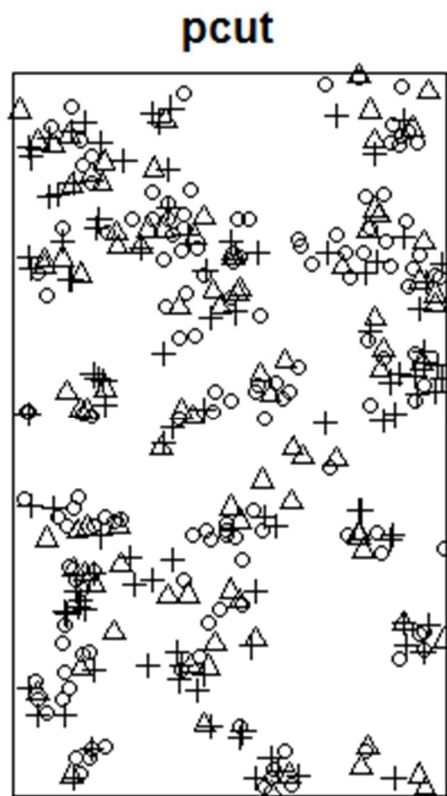
ETP X žymių tankis



58 pav. ETP X erdvinis žymių trendas

Žymės suskirstysime į tris intervalus pagal medžių aukštį. Medžiai, kurių aukštis bus iki 20 metrų bus priskirti „Žemieji“ grupei, kurių aukštis nuo 20 iki 25 metrų – „Vidutiniai“ grupei ir likę, patys aukščiausi bus priskirti į grupę „Aukštieji“. Tokiu būdu taškų kiekybinės žymės virs kokybinėmis.

```
> pcut <- cut(X, breaks = c(14, 20, 25, 30), labels = c("Žemieji", "Vidutiniai", "Aukštieji"))
> pcut
marked planar point pattern: 343 points
multitype, with levels = Žemieji Vidutiniai Aukštieji
window: rectangle = [0, 3] x [0, 5] units
> plot(pcut)
Žemieji Vidutiniai Aukštieji
  1      2      3
```



59 pav. Multi-tipo taškinis procesas X

Vektoriuje nurodytas pirmasis intervalo skaičius 14 buvo pasirinktas specialiai, kadangi nurodžius intervalo pradžią 15 į intervalą nepakliūva šis skaičius ir programa automatiškai duoda pranešimą, kad kai kurie taškai įgijo reikšmę NA. Kaip matėme anksčiau, dar nesuskirsčius aukščio į grupes, programa taškinį procesą X vertino kaip žymėtą taškinį procesą, kurio žymės buvo sveikieji skaičiai. Dabar turime žymes kaip faktorius, todėl modelis aprašomas kaip multi-tipo ir analizei taikomos technikos, kurios tinka multi-tipo procesams (analogiškai kaip su žymėmis-medžių pavadinimais).

```
> is.multitype(pcut)
[1] TRUE
> summary(pcut)
```

Marked planar point pattern: 343 points
 Average intensity 22.9 points per square kilometres
 Multitype:
 frequency proportion intensity
 Žemieji 142 0.414 9.47
 Vidutiniai 94 0.274 6.27
 Aukštieji 107 0.312 7.13
 Window: rectangle = [0, 3]x[0, 5]kilometres

Iš pateiktų statistikų matome, jog daugiausiai buvo sukurta žymių „Žemieji“, o mažiausiai „Vidutiniai“ tipo. Apžvelgsime artimiausių taškų kaimynus.

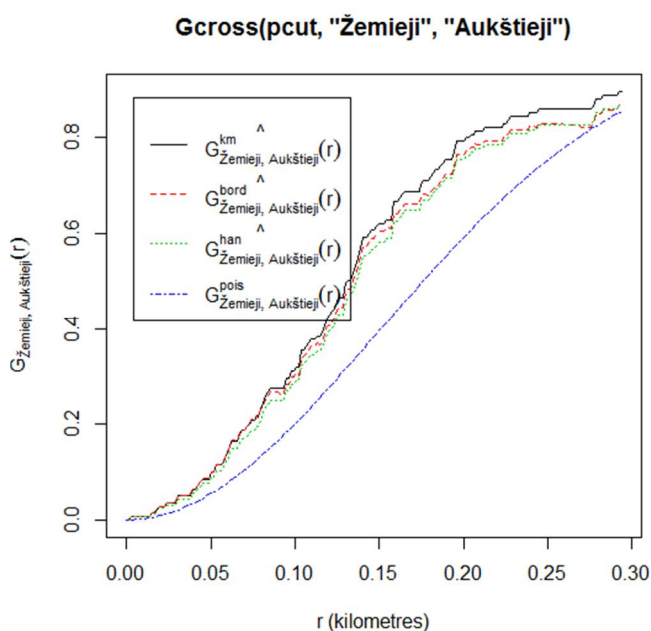
```
> M <- marktable(pcut, R = 0.2)
> M[1:15, ]
mark
```

```
point Žemieji Vidutiniai Aukštieji
```

1	5	0	1
2	1	0	1
3	2	0	0
4	0	0	2
5	2	2	1
6	4	1	2
7	2	1	2
8	2	2	4
9	3	0	2
10	4	1	2
11	5	2	2
12	6	0	1
13	4	3	1
14	6	1	0
15	4	1	1

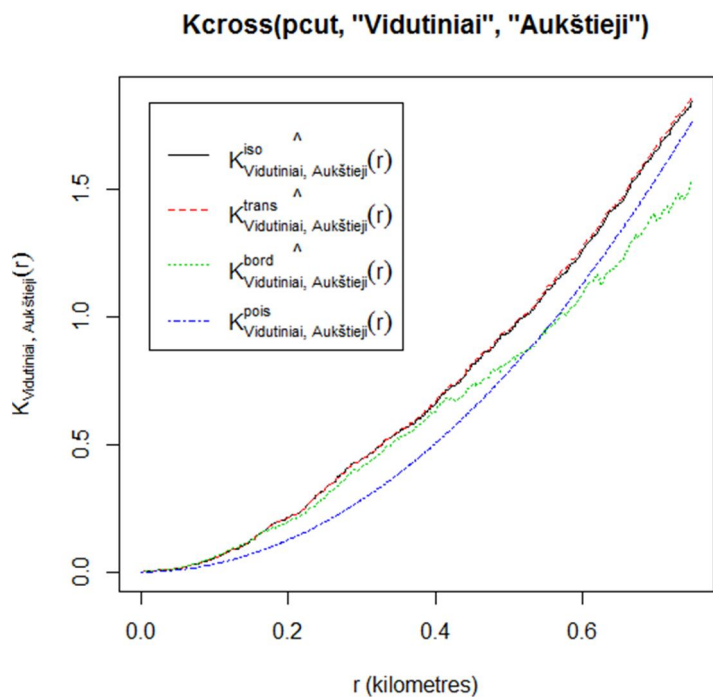
Matome, jog mažiausiai artimiausių kaimynų pirmiesiems penkiolikai taškų yra iš grupės „Vidutiniai“, o daugiausiai grupės „Žemieji“. Apskaičiuosime kryžmines atstumų funkcijas.

```
> plot(Gcross(pcut, "Žemieji", "Aukštieji"))
```



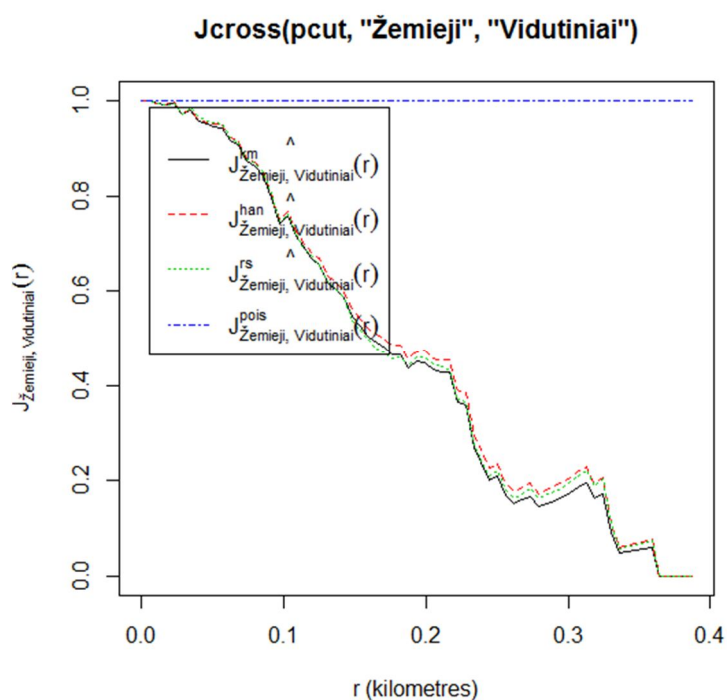
60 pav. Kryžminė G_{cross} funkcija porai „Žemieji“-„Aukštieji“

> plot(Kcross(pcut, "Vidutiniai", "Aukštieji"))



61 pav. Kryžminė $Kcross$ funkcija porai „Vidutiniai“-„Aukštieji“

> plot(Jcross(pcut, "Žemieji", "Vidutiniai"))



62 pav. Kryžminė $Jcross$ funkcija porai „Žemieji“-„Vidutiniai“

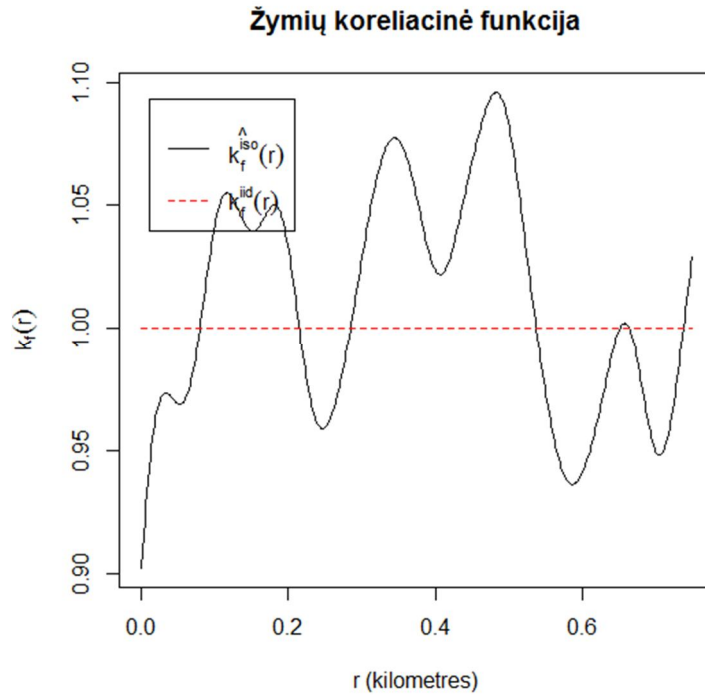
Iš funkcijų grafikų galime daryti išvadą, kad aukštis taškams taip pat priskirtas nepriklausomai, kaip ir medžių pavadinimai.

Apskaičiuosime žymių koreliaciją.

```

> eqfun<-function(m1,m2){m1==m2}
> Mc<-markcorr(pcut,eqfun,correction="isotropic", method="density",kernel="gaussian")
> plot(Mc)
      lty col key          label
iso  1  1  iso     $\hat{k}_f^{iso}(r)$ 
theo 2  2  theo {  $k_f^{iid}(r)$  }
              meaning
iso  Ripley isotropic correction estimate of  $k_f(r)$ 
theo theoretical value (independent marks) for  $k_f(r)$ 

```



63 pav. Žymių koreliacinė funkcija

Grafiką galime pakomentuoti taip: stipriausias ryšys tarp žymių yra atstumu ~ 0.5 , ten, kur funkcija įgyja didžiausią reikšmę, tuo tarpu kai funkcija yra lygi 1 sakome, jog ryšio visai nėra.

IŠVADOS

- Atliekant užsienio literatūros apžvalgą buvo pastebėta, jog naujesni literatūros šaltiniai erdvinius taškinius procesus aprašo ne tik teoriškai, bet šalia parodo ir praktinį jų realizavimą tam tikrų statistinių programų pagalba. Daugelis mokslininkų pastebi, jog ETP teorija nėra dar išgryninta, ypatingai ETP modeliavimas, todėl daugelis analizavimo technikų yra praktinio pobūdžio.
- Rašant šį darbą, atvirojo kodo programa R buvo kuriama be galo daug atsitiktinių erdvinių taškinių vaizdų. Pastebėta, jog programa visus kartus sukūrė erdvinį taškinį vaizdą, kuris buvo realizacija homogeninio Puasono proceso. Galime sakyti, jog R programa „pagal nutylėjimą“ sukuria Puasono procesą, jeigu nėra nurodomas kitas ETP modelis.
- Savo darbuose mokslininkai vis labiau akcentuoja klasterinius ETP, kurie yra vis dažniau sutinkami realybėje, o tuo tarpu klasikinis Puasono proceso modelis naudojamas kaip teorinis pagrindas, sulyginimo objektas, atliekant įvairius statistinius testus.
- Erdvinių taškinių procesų analizėje, pagrindinis tyrimo objektas yra nulinės hipotezės priėmimas arba atmetimas. Šis žingsnis plačiausiai analizuojamas, nes hipotezei patvirtinti reikia ištirti ETP intensyvumą bei pilną erdvinį atsitiktinumą.
- Praktiškai analizuojant ETP buvo pastebėta, jog taškų klasteriai gali būti pastebėti jau pavaizdavus erdvinį taškinį vaizdą grafiškai, tačiau vien grafinio taškinio vaizdo vertinimo nepakanka. Tam turi būti naudojamos įvairios technikos (testai, atstumų tarp taškų nustatymai ir kt.).
- ETP modeliavimo etape svarbu parinkti tankio funkciją tokią, kuri geriausiai aprašytų ETP tankį. Dažniausiai yra parenkamos kelios funkcijos ir iš jų išrenkama geriausiai tinkanti. Jei viena kažkuri funkcija neišsiskiria kaip geriausiai tinkanti atliekant funkcijų vertinimą, patariama modelio glodinimą atlikti su geriausiai tinkančiomis keliomis tankio funkcijomis ir iš priglodintų modelių išrinkti geriausiai tinkantį.
- Darbe buvo analizuojami žymėti taškiniai procesai su kokybinėmis ir kiekybinėmis žymėmis. Užsienio literatūroje labiausiai analizuojami ir daugiausiai praktinių pavyzdžių pateikiama multi-tipo ETP, kurių žymės yra kokybinės. Analizuojant žymėtus ETP svarbu ištirti žymių pasiskirstymą ir priklausomybę, nes tai gali būti priežastis tam tikro taškų išsidėstymo.

LITERATŪRA

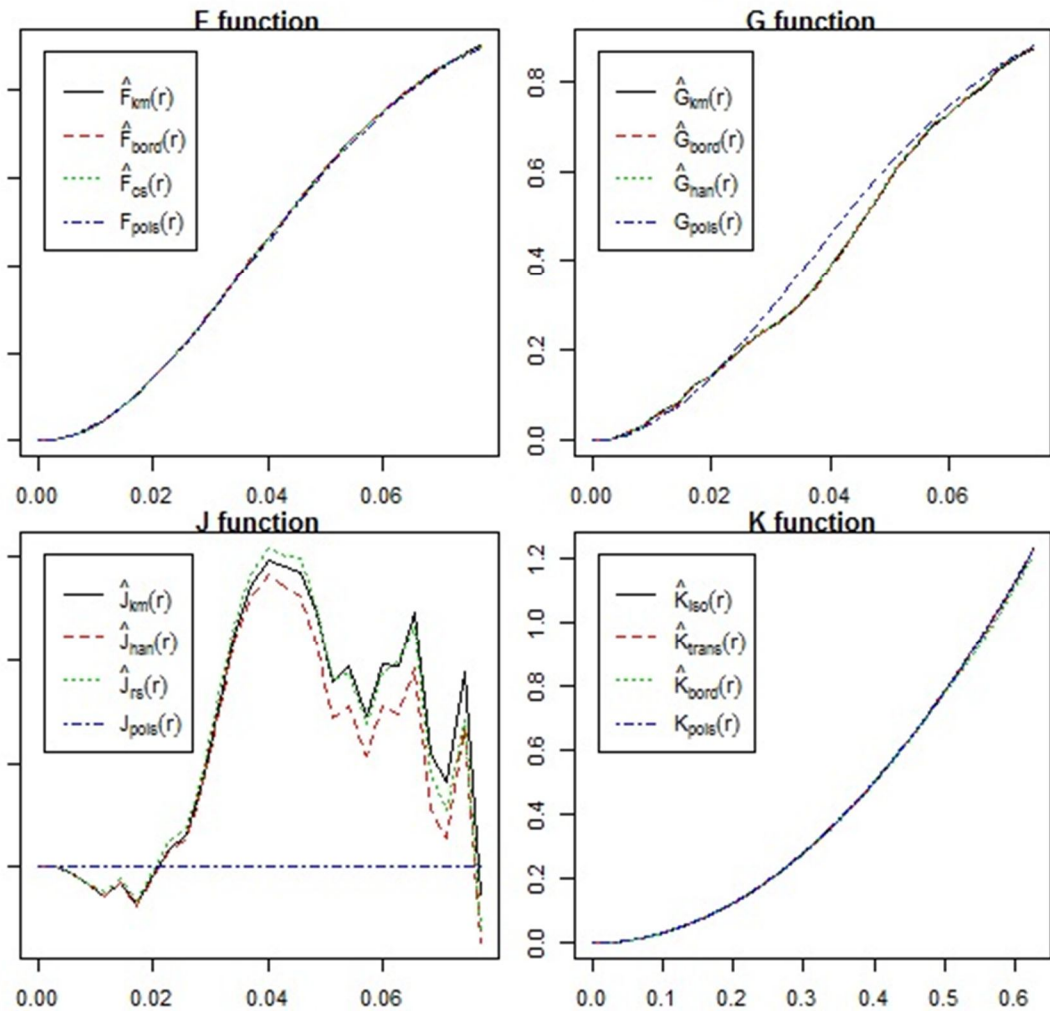
1. Baddeley A., Turner R., 2010. Package „spatstat“: Spatial Point Pattern analysis, model-fitting, simulation, tests, Version 1.18-1.
2. Baddeley A., CSIRO, University of Western Australia, 2008. Analysing Spatial Point Patterns in R, Workshop notes, Version 3. Australia: CSIRO
3. Baddeley A., CSIRO, University of Western Australia, 2008. ASC Workshop Computing with R
4. Berry D.A., B.W. Lindgren, 1995. Statistics, Theory and Methods. Duxbury Press, ISBN 0-534-50479-5.
5. Burden F.F., 2003. Point Pattern Analysis. GIS Resource Document 03-41
6. Cliff A.D., J.K. Ord, 1981. Spatial processes: models and applications. Taylor & Francis, 266.
7. Clifford Cohen A., Marshall G.C., 1965. Estimation in the negative binomial distribution. NASA TECHNICAL MEMORANDUM X-53372.
8. Dixon P.M., 2001. Nearest Neighbor Methods. Prieiga per internetą: <http://www.stat.iastate.edu/preprint/articles/2001-19.pdf>
9. Dučinskas K., Šaltytė-Benth J., 2003. Erdvinė statistika. Klaipėda: Klaipėdos universiteto leidykla.
10. GUAN T.G., 2003. Nonparametric Methods of Assessing Spatial Isotropy. Dissertation. Texas A&M University
11. Jain S.K., College of Agricultural and Environmental Sciences, 1958. Fitting the negative binomial distribution to some data on asynaptic behavior of chromosomes.
12. Kreyszing E., 2006. Advanced engineering mathematics, 9th edition. New Jersey: John Wiley & Sons, Inc.
13. Lieshout M.-C., 2010. Spatial Point Process Theory. Taylor and Francis Group, LLC
14. Merton A.A., J.A. Hoeting, C.T. Webb, 2008. Distribution-free Comparison of Multiple Spatial Point Patterns. Prieiga per internetą: http://www.stat.colostate.edu/research/Technical%20Reports/2008/2008_20.pdf
15. Møller J., Toftaker H., 2012. Geometric anisotropic spatial point pattern analysis and Cox processes. Department of Mathematical Sciences, Aalborg University
16. Møller J., Waagepetersen R.P., 2007. Modern statistics for spatial point processes. Department of Mathematical Sciences, Aalborg University
17. Rypley B.D., 1979. Tests of “Randomness” for Spatial Point Patterns. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 41, Nr.3, psl. 368-374

18. Waagepetersen R., Guan Y., 2009. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society*, 71, 3 part 3, pp. 685-702.
19. Zhang Y., Zhirui Ye, Dominique Lord, Texas A&M University, 2006. Estimating the Dispersion Parameter of the Negative Binomial Distribution for Analyzing Crash Data Using a Bootstrapped Maximum Likelihood Method.
20. Modulio „Atsitiktiniai taškiniai procesai“ paskaitų medžiaga. Modulio kodas P160M010
21. Modulio „Geostatistika ir GIS“ paskaitų medžiaga. Modulio kodas P160M005.
22. Modulio „Statistiniai klasifikavimo metodai“ paskaitų medžiaga. Modulio kodas P160M006

PRIEDAI

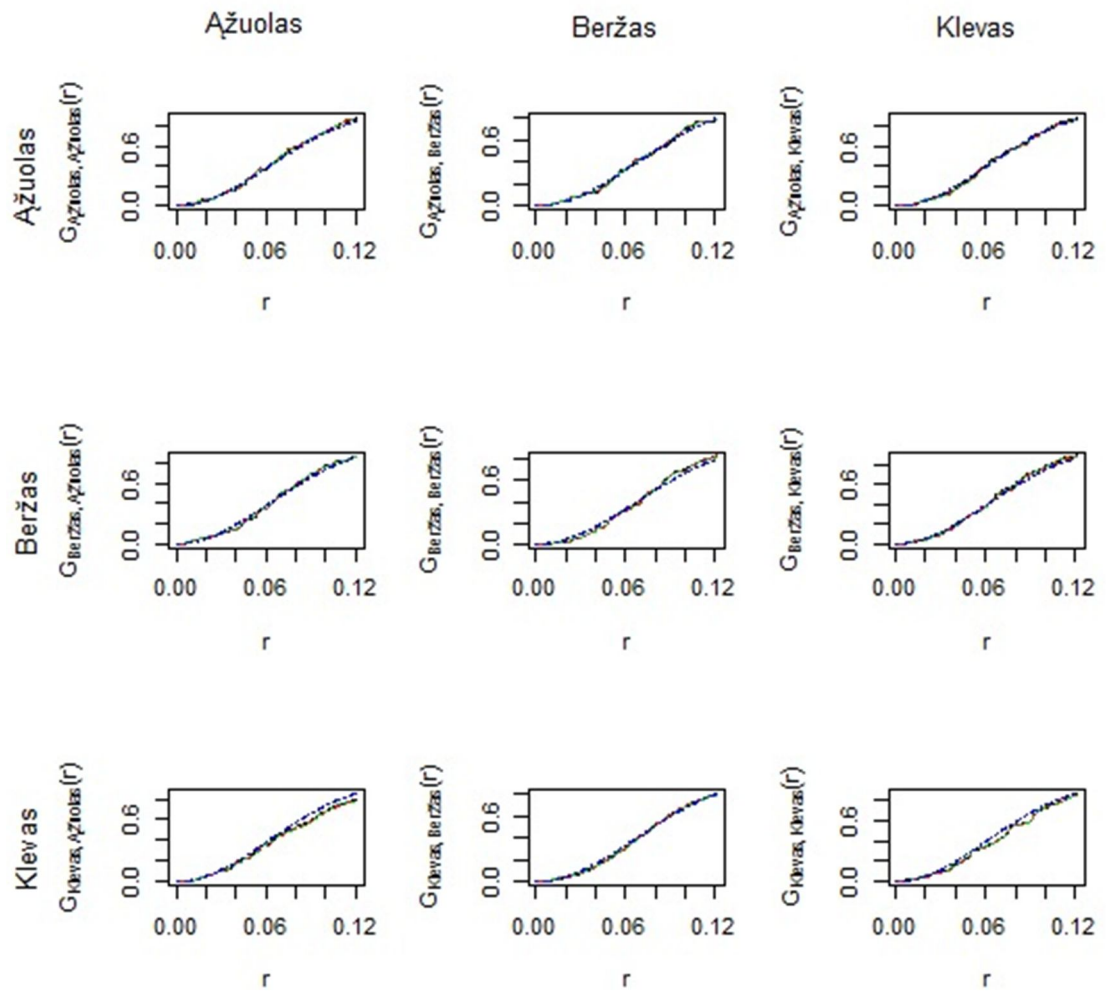
Priedas nr.1

Visos atstumų funkcijos



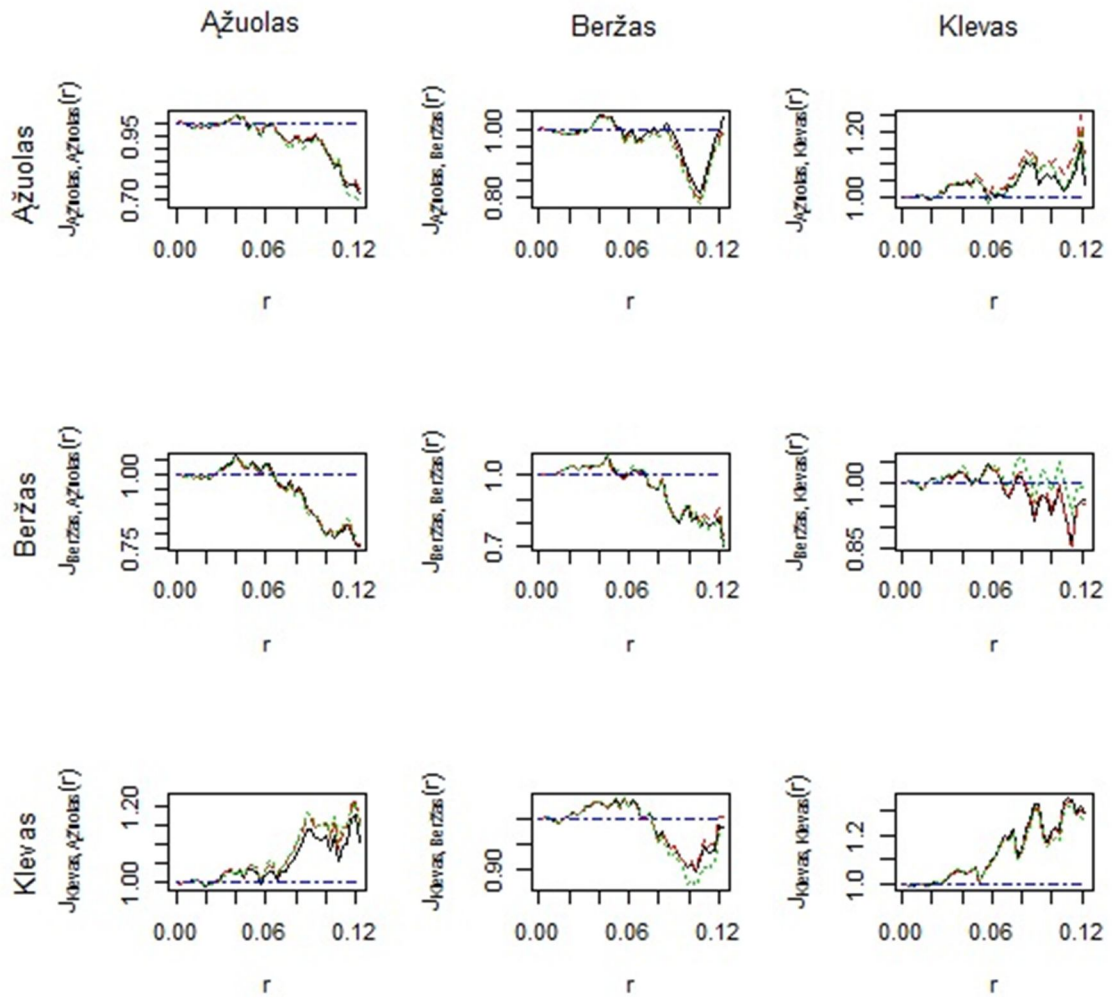
23 pav. Atstumų funkcijos F , G , J , K

array of G functions for p.



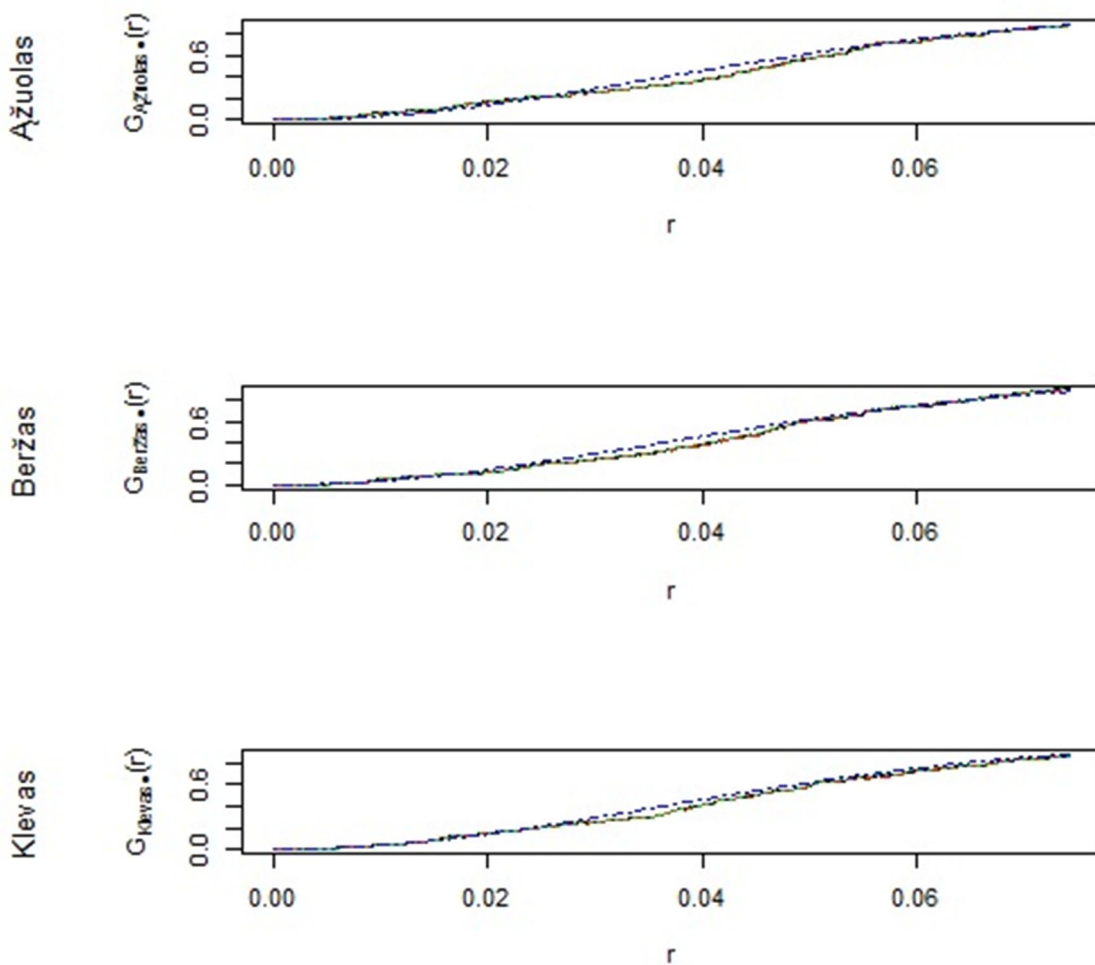
38 pav. Kryžminė G_{ij} funkcija visoms žymių poroms

array of J functions for p.



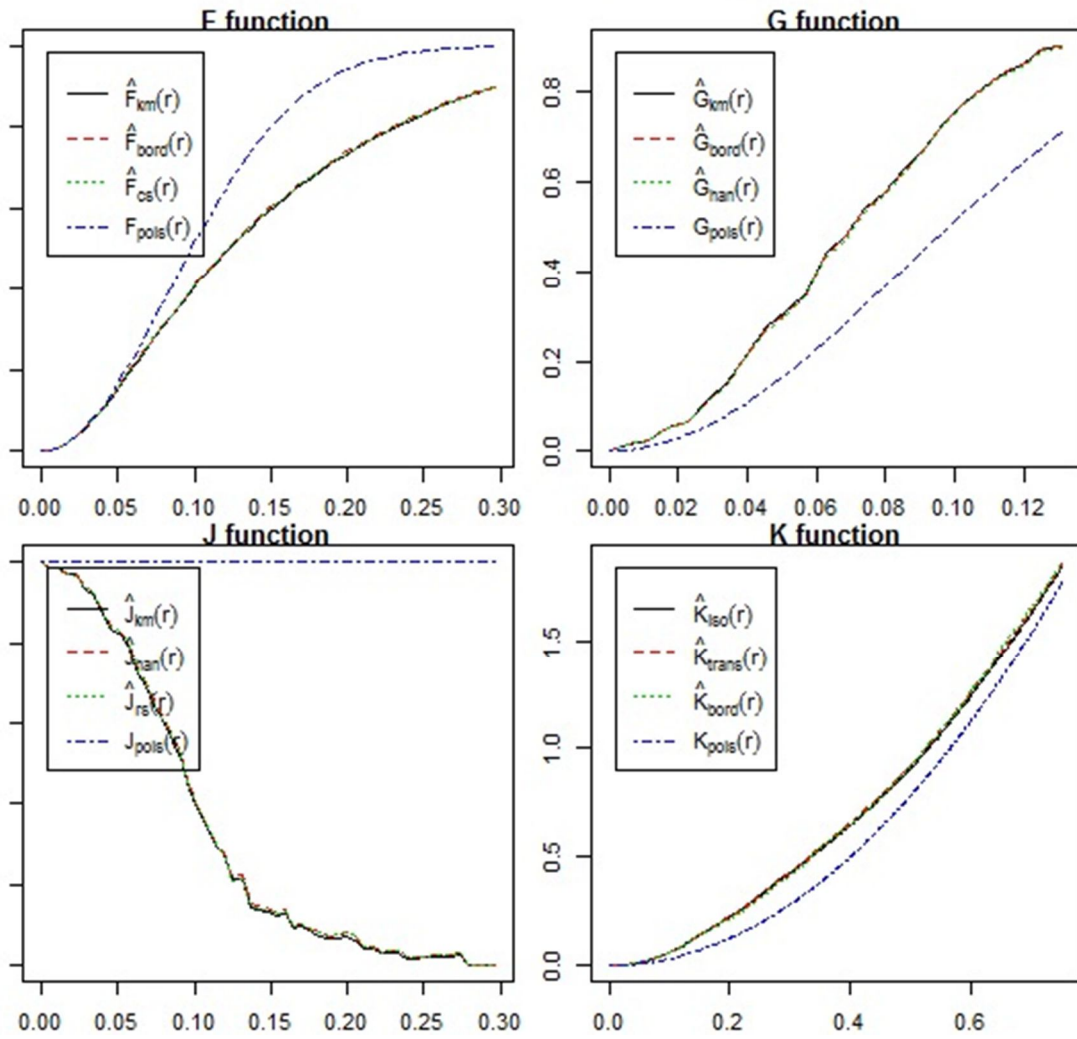
9 pav. Kryžminė J_{ij} funkcija visoms žymių poroms

array of Gdot functions for p.



40 pav. Kryžminė $G_i(r)$ funkcija nuo fiksuotos žymės iki kito bet kokio taško

Visos atstumų funkcijos



47 pav. ETP X atstumų funkcijos