VILNIAUS UNIVERSITETAS

INGRIDA VAIČIULYTĖ

MARKOVO GRANDINĖS MONTE-KARLO METODO TYRIMAS IR TAIKYMAS

Daktaro disertacija

Fiziniai mokslai, informatika (09 P)

Vilnius, 2014

Disertacija rengta 2009–2014 metais Vilniaus universiteto Matematikos ir informatikos institute.

Mokslinis vadovas:

prof. habil. dr. Leonidas Sakalauskas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P).

PADĖKA

Nuoširdžiai dėkoju moksliniam vadovui prof. habil. dr. Leonidui Sakalauskui už neįkainojamą pagalbą rašant disertaciją, už suteiktas žinias, nuoseklų vadovavimą, patarimus bei nuolatinį skatinimą dirbti ir tobulėti, VU Matematikos ir informatikos instituto direktoriui prof. habil. dr. Gintautui Dzemydai už visapusišką paramą studijuojant doktorantūroje, taip pat VU MII, o ypač Operacijų tyrimo sektoriaus darbuotojams bei doktorantams už palaikymą ir bendradarbiavimą.

Dėkoju disertacijos recenzentėms doc. dr. Danutei Krapavickaitei, dr. Vaidai Bartkutei-Norkūnienei ir kolegai Valerijonui Dumskiui atidžiai perskaičiusiems disertaciją ir pateikusiems vertingus patarimus, pasiūlymus bei kritines pastabas, padėjusias tobulinti disertaciją.

Nuoširdžiai dėkoju savo šeimai – vyrui Vitalijui, jo tėvams, vaikams Arinai ir Deividui – už jų meilę, moralinį palaikymą, kantrybę ir supratimą.

Ingrida Vaičiulytė

TURINYS

ŽYMĖJIMAI IR SANTRUMPOS

MCMC – Monte-Karlo Markovo grandinė (angl. *Markov chain Monte Carlo*);

MLM – didžiausio tikėtinumo metodas (angl. *maximum likelihood method*);

EB – empirinis Bajeso metodas (angl. *empirical Bayes*);

EM algoritmas – rekurentinis tikėtinumo maksimizavimo algoritmas (angl. e*xpectation maximization*);

CRT – centrinė ribinė teorema;

DSD – didžiųjų skaičių dėsnis;

$ST(\mu, \Omega, \Theta, \alpha, \eta)$ – asimetrinis *t* skirstinys (angl. *a skew t distribution*);

$S_\alpha(\sigma, \beta, \mu)$ – stabilusis dėsnis;

σ – mastelio parametras;

β – asimetrijos parametras;

μ – poslinkio (vidurkio) parametras;

Ω, Θ – kovariacijų matricos;

α – formos parametras;

η – srities parametras;

cf – pasikliautinojo intervalo ilgis;

d – dimensija;

k – iteracijų skaičius (arba Markovo grandžių skaičius);

L – tikėtinumo funkcija;

K – populiacijos imties dydis;

N – Monte-Karlo imties dydis;

M – retų įvykių rūšių skaičius;

H – statistinė hipotezė;

$\Psi_{\delta,p}$ – χ_p^2 skirstinio kvantilis, δ – pasikliovimo lygmuo;

$F_{\nu,p}$ – Fišerio skirstinio kvantilis, ν – pasikliovimo lygmuo;

τ_{γ} – normaliojo skirstinio kvantilis, γ – pasikliovimo lygmuo;

p – laisvės laipsnis;

suic – savižudybės (suicidai);

hom – nužudymai;

Close – akcijų kainos;

Volume – pardavimo apimtis;

BMI – kūno masės indeksas;

Bfat – kūno riebalų masė;

ssf – odos raukšlių skaičius;

LBM – liesa kūno masė.

PAVEIKSLĖLIŲ SĄRAŠAS

LENTELIŲ SĄRAŠAS

ĮVADAS

Tyrimų sritis

Daugeliui verslo, technikos ir gamtos procesų yra būdingas atsitiktinumas. Tokie procesai gali būti modeliuojami bei prognozuojami tikimybiniais statistiniais metodais, pasinaudojus duomenimis apie proceso eigą. Atsitiktiniams procesams aprašyti ir tirti dažnai taikomi įvairūs stochastiniai metodai, susiję su atsitiktinių dydžių sekų generavimu ir apdorojimu: Markovo grandinės Monte-Karlo metodas, Gibso imties išrinkimo ir Metropolio-Hastingso algoritmai, stochastinė aproksimacija bei kt. (Rubinstein ir Kroese, 2007; Spall, 2003). Šitokie metodai padeda tirti atsitiktines sistemas bei procesus kompiuterinio ir matematinio modeliavimo būdu.

Markovo grandinės Monte-Karlo (angl. *Markov Chain Monte Carlo* – MCMC) metodas yra kompiuterinio imitavimo būdas, plačiai taikomas statistikoje, technikoje, fizikoje, bioinformatikoje ir t. t., vertinant matematinių modelių nežinomus parametrus. Markovo grandinės metodai ir juos realizuojantys algoritmai yra aktualūs, kai generuojant atsitiktines sekas, nustatomi atsitiktinių dydžių pasiskirstymo tipai bei parametrai. MCMC metodas dažnai taikomas retų įvykių tikimybėms apskaičiuoti imties išrinkimo metodu (angl. *importance sampling*), duomenų analizėje EM (angl. e*xpectation maximization*) algoritmu, Bajeso metodo praktiniam pritaikymui, modeliuojant aposteriorinius skirstinius bei skaitiniais metodais nustatant jų parametrus, ir pan. Kadangi tikimybinių modelių sritis, kuriems gali būti pritaikyti MCMC metodai, yra labai plati, disertacijoje apsiribojama daugiamačiais skirstiniais, kurie gali būti sukonstruoti hierarchiniu būdu iš elipsinių skirstinių. Tokiu būdu gauti skirstiniai gali būti pritaikyti sprendžiant daugelį praktinių ir teorinių duomenų analizės uždavinių. Svarbu pažymėti, kad taip sukurti modeliai leidžia sukonstruoti asimetrinius skirstinius su didelėmis atsitiktinėmis reikšmėmis, kurie dažnai pasitaiko modeliuojant finansinius procesus ir informacinius srautus kompiuterių tinkluose (Kabašinskas, 2007; Kaklauskas, 2012).

Problemos aktualumas

Skaičiuojamuoju požiūriu MCMC metodas leidžia spręsti lygtis, į kurias įeina sudėtingi daugialypiai integralai, sukonstruojant Monte-Karlo imčių Markovo grandinę. Žinomuose MCMC algoritmuose paprastai sugeneruojamos kelios arba keliolika grandžių, empiriškai nustatant konvergavimą ir užfiksavus visose grandyse pakankamai didelį Monte-Karlo imčių tūrį (Bradley ir Thomas, 2000). Aišku, kad skaičiuojamuoju požiūriu tokios procedūros yra nelabai efektyvios, nes tenka sunaudoti daug kompiuterio laiko grandžių generavimui, o nutraukus grandinės generavimą empiriškai, statistiškai reikšmingas konvergavimas dar gali būti nepasiektas. Taip pat taikant MCMC dažnai kyla problema, kokio dydžio imtys turėtų būti generuojamos atskirose grandyse.

Tokiu būdu, aktualios MCMC skaitmeninimo problemos yra grandinės grandžių skaičiaus nustatymas ir Monte-Karlo imčių tūrio atskirose grandyse reguliavimas. Prie kitų aktualių MCMC skaitmeninimo problemų galima priskirti pradinės grandies parametrų parinkimą, skaičiavimo atlikimą esant tikimybinių modelių singuliarumui, skaičiavimai esant labai didelėms arba labai mažoms tarpinėms reikšmėms. Gana aktuali MCMC panaudojimo problema yra asimetrinių skirstinių su didelėmis atsitiktinėmis reikšmėmis konstravimas ir parametrų vertinimas.

Tyrimų objektas

Disertacijos tyrimų objektas yra adaptuotos Markovo grandinės Monte-Karlo metodo tyrimas, skaitinis realizavimas ir taikymas duomenų analizėje, tikslumo vertinimo, grandžių skaičiaus parinkimo, algoritmo stabdymo ir Monte-Karlo imčių tūrio reguliavimo būdai.

Tyrimų tikslas ir uždaviniai

Darbo tikslas – ištirti Markovo grandinės Monte-Karlo adaptavimo metodus, sudarant efektyvius skaitinius algoritmus, leidžiančius priimti duomenų analizės sprendimus su iš anksto nustatytu patikimumu, bei ištirti šių algoritmų efektyvumą.

Siekiant šio tikslo disertacijoje sudaryti algoritmai bei juos realizuojanti programinė įranga, skirta Monte-Karlo imčių tūrio adaptavimui atskirose grandyse, įvertinių tikslumo vertinimui bei Markovo grandinės proceso stabdymui. Sudaryti metodai ir algoritmai yra pritaikyti duomenų statistiniam vertinimui MCMC metodu, pasinaudojant praktiškai surinktais arba žinomais literatūroje duomenimis. Šių metodų bei algoritmų efektyvumas tiriamas pasinaudojant disertacijoje sudarytu statistinio modeliavimo metodu. MCMC skaitmeninimo problemos, nagrinėjamos disertacijoje, yra ištirtos sprendžiant kelis duomenų analizės uždavinius (asimetrinio t skirstinio, Puasono-Gauso modelio ir stabiliojo dėsnio parametrų vertinimo), pasižyminčius ypatybėmis, kurios yra būdingos daugeliui kitų uždavinių, ir tokiu būdu gauti rezultatai gali būti sėkmingai pritaikomi ir kitiems uždaviniams spręsti.

Tyrimų metodai

Disertacijoje suformuluoti uždaviniai sprendžiami pasinaudojant daugiamatės statistikos ir kompiuterinio modeliavimo metodais: daugiamačių imčių generavimas priėmimo-atmetimo (angl. *acceptance-rejection*) metodu, didžiausio tikėtinumo metodas, EM algoritmas, Bajeso metodas, Monte-Karlo metodas, stochastinio gradientinio nusileidimo (angl. *gradient descent*) metodas, (Čiegis, Būda, 1997; Sakalauskas 2000), statistinių hipotezių tikrinimas.

Mokslinis naujumas

Disertacijoje gauti šie rezultatai:

1) Markovo grandinės Monte-Karlo imčių tūrio adaptavimo taisyklė;
2) Markovo grandinės generavimo stabdymo taisyklė;
3) Markovo grandinės Monte-Karlo algoritmų efektyvumo tyrimo metodas;
4) adaptuotos Markovo grandinės Monte-Karlo metodo pritaikymas uždaviniams, kurių tikimybiniai modeliai konstruojami hierarchiniu būdu iš elipsinių skirstinių, spręsti (asimetrinio t skirstinio, Puasono-Gauso modelio ir daugiamačio α-stabiliojo dėsnio parametrams vertinti).

Praktinė darbo reikšmė

Disertacijoje sudaryti MCMC algoritmai daugiamačio asimetrinio t skirstinio, Puasono-Gauso modelio ir daugiamačio α-stabiliojo dėsnio parametrams vertinti duotu tikslumu bei kompiuterinio modeliavimo būdu ištirtas šių algoritmų efektyvumas. Sudaryti algoritmai gali būti pritaikyti praktikoje iškylantiems uždaviniams spręsti (finansinių sekų prognozavimui, biologinių populiacijų ir draudiminių įvykių populiacijose tyrimams ir pan.). Gauti rezultatai gali būti pritaikyti sprendžiant įvairius statistinio vertinimo uždavinius MCMC metodu: imties išrinkimo metodas, EM algoritmas, didžiausio tikėtinumo metodas ir pan.

Disertacijoje gauti šie praktiniai rezultatai:

1) sudarytas algoritmas asimetrinio t skirstinio parametrams vertinti;
2) sudarytas algoritmas Puasono-Gauso modelio parametrams vertinti;
3) sudarytas algoritmas stabiliojo simetrinio skirstinio parametrams vertinti;
4) sudarytas statistinio modeliavimo metodas MCMC algoritmų efektyvumui tirti.

Ginamieji teiginiai

1. Sudaryti algoritmai bei juos realizuojanti programinė įranga, skirta:
    a) Monte-Karlo imčių tūrio reguliavimui Markovo grandyse;
    b) įvertinių tikslumo vertinimui;
    c) Markovo proceso grandinės stabdymui.
2. Sudaryti algoritmai gali būti pritaikyti duomenų statistiniam vertinimui adaptuotu MCMC metodu sprendžiant praktinius ir testinius uždavinius.
3. Sudaryti algoritmai leidžia spręsti statistinio vertinimo uždavinius MCMC metodu duotu tikslumu sumažinant (maždaug dvigubai) skaičiavimo apimtį palyginus su žinomais algoritmais.

Darbo rezultatų aprobavimas

<u>Skaityti pranešimai respublikinėse konferencijose:</u>

1. Markovo grandinės Monte Karlo metodo taikymas tiriant sociologinius duomenis. *Operacijų tyrimas versle ir inžinerijoje* LOTD-2014, Vilnius, 2014-09-22.
2. Daugiamatis retų įvykių tikimybių vertinimo algoritmas. *Kompiuterininkų dienos – 2013*, Šiauliai, 2013-09-19–2013-09-21.
3. Daugiamačių stabiliųjų skirstinių parametrų vertinimas Monte-Karlo Markovo grandinės metodu. *Operacijų tyrimas ir taikymai* LOTD-2013, Šiauliai, 2013-09-19.
4. Antisimetrinio *t* skirstinio taikymas finansinių duomenų tyrimui. *Verslas, studijos ir aš*, Šiauliai, 2013-03-28.
5. Daugiamatis mažų dažnių vertinimo algoritmas. *Lietuvos matematikų draugijos 53-oji konferencija* LMD-53, Klaipėda, 2012-06-11–2012-06-12.
6. Daugiamačio antisimetrinio *t*-skirstinio parametrinis įvertinimas. *Operacijų tyrimai versle, inžinerijoje ir informacinėse technologijose* LOTD-2011, Kaunas, 2011-09-30.

7. Daugiamačio antisimetrinio *t*-skirstinio parametrų vertinimas Monte-Karlo Markovo grandinių metodu. *Lietuvos matematikų draugijos 51-oji konferencija* LMD-51, Šiauliai, 2010-06-17–2010-06-18.
8. Apie antisimetrinio *t*-skirstinio vertinimą Monte Karlo grandžių metodu. *Lietuvos matematikų draugijos 50-oji konferencija* LMD-50, Vilnius, 2009-06-18–2009-06-19.

Skaityti pranešimai tarptautinėse konferencijose:

9. Application of skew *t*-distribution in the field of investors' preferences visualization. *Stochastic Programming* ICSP-2013, Bergamo, Italy, 2013-07-08–2013-07-12.
10. The assessment of multi-dimensional frequency by Monte-Carlo Markov chain approach. *European Conference on Operational Research* EURO-2013, Roma, Italy, 2013-07-01–2013-07-04.
11. The application of stable and skew *t* distributions in predicting the change in accounting and governance risk ratings. *Electrical and Control Technologies* ECT-2013, Kaunas, 2013-05-02–2013-05-03.
12. Adaptive Monte-Carlo Markov chain for multivariate statistical estimation. *Stochastic Programming for Implementation and Advanced Applications* STOPROG-2012, Neringa, Lithuania, 2012-07-03–2012-07-06.
13. Adaptive Monte-Carlo Markov chain. *Stochastic Modeling Techniques and Data Analysis* SMTDA-2012, Crete, Greece, 2012-06-05–2012-06-08.
14. Maximum likelihood estimation of multivariate skew *t*-distribution. *Operations Research and Enterprise Systems* ICORES-2012, Algarve-Vilamoura, Portugal, 2012-02-04–2012-02-06.
15. Estimation of shape parameter of the multivariate *t*-skew distribution. *Applied Stochastic Models and Data Analysis* ASMDA-2011, Roma, Italy, 2011-06-07–2011-06-10.
16. Estimation of skew *t*-distribution by Monte-Carlo Markov chain approach. *Computer Data Analysis and Modeling* CDAM-2010, Minsk, Belarus, 2010-09-07–2010-09-11.

17. Estimation of skew *t*-distribution by the Monte-Carlo Markov chain approach. *Stochastic Modeling Techniques and Data Analysis* SMTDA-2010, Crete, Greece, 2010-06-08–2010-06-11.
18. Estimation of skew-*t* distribution by Monte-Carlo Markov chain approach. *Applied Stochastic Models and Data Analysis* ASMDA-2009, Vilnius, Lithuania, 2009-06-30–2009-07-03.

Darbo rezultatų publikavimas

Straipsniai recenzuojamuose Lietuvos ir užsienio leidiniuose:

1. Sakalauskas, L., Kalsyte, Z., Vaiciulyte, I., Kupciunas, I. (2014). The relationship between the transparency in provision of financial data and the change in investors' expectations. *Ekonominė inžinerija – Engineering Economics*. (**ISI Web of Science**, po recenzavimo priimtas spaudai, išduota pažyma Nr. 221) (ISSN 1392-2785)
2. Vaičiulytė, I., Sakalauskas, L. (2014). Markovo grandinės Monte Karlo metodo taikymas tiriant sociologinius duomenis. *Jaunųjų mokslininkų darbai*, 2(42). (**Index Copernicus, CEEOL,** po recenzavimo priimtas spaudai, išduota pažyma Nr. 036-14) (ISSN 1648-8776)
3. Vaičiulytė, I., Sakalauskas, L. (2014). Sub-gausinio vektoriaus skirstinio parametrų vertinimas Monte-Karlo Markovo grandinės metodu. *Jaunųjų mokslininkų darbai*, 1(41), 104–107. (**Index Copernicus, CEEOL**) (ISSN 1648-8776)
4. Vaičiulytė, I. (2014). Antisimetrinio *t* skirstinio taikymas tiriant finansinius duomenis. *Jaunųjų mokslininkų darbai*, 1(41), 147–151. (**Index Copernicus, CEEOL**) (ISSN 1648-8776)
5. Sakalauskas, L., Vaičiulytė I. (2013). Multidimensional rare event probability estimation algorithm. *Computational Science and Techniques*, 1(2), 222–228. (**Google Scholar**, The Directory of Research Journal Indexing (**DRJI**), **Directory of Lithuanian science journals**) (ISSN 2029-9966)

6. Sakalauskas, L., Vaičiulytė, I. (2012). Daugiamatis mažų dažnių vertinimo algoritmas. *Lietuvos matematikos rinkinys, Lietuvos matematikų draugijos darbai*, 53B, 260–263. (ISSN 0132-2818)
7. Vaičiulytė, I., Sakalauskas, L. (2011). Daugiamačio antisimetrinio *t*-skirstinio parametrinis įvertinimas. *Jaunųjų mokslininkų darbai*, 4(33), 157–163. (**Index Copernicus, CEEOL**) (ISSN 1648-8776)
8. Vaičiulytė, I., Sakalauskas, L. (2011). Daugiamačio antisimetrinio *t*-skirstinio parametrų vertinimas Monte-Karlo Markovo grandinių metodu. *Jaunųjų mokslininkų darbai*, 1(30), 137–141. (**Index Copernicus, CEEOL**) (ISSN 1648-8776)

Straipsniai recenzuojamoje užsienio tarptautinių konferencijų medžiagoje:

9. Sakalauskas, L., Vaiciulyte, I. (2012). Maximum likelihood estimation of multivariate skew *t*-distribution. *Proceedings of 1st International Conference on Operations Research and Enterprise Systems*, 200–203, Vilamoura: SciTePress. (**Scopus, INSPEC, DBLP**) (ISBN 978-989-8425-97-3)
10. Sakalauskas, L., Vaiciulyte, I. (2012). Adaptive Monte-Carlo Markov chain. *Proceedings of 2nd International Conference „Stochastic Modeling Techniques and Data Analysis“*, 653–660. (ISBN 978-981-270-968-4)
11. Sakalauskas, L., Vaiciulyte, I. (2011). Estimation of shape parameter of the multivariate *t*-skew distribution. *Proceedings of 14th International Conference „Applied Stochastic Models and Data Analysis”*, 1211–1218. (ISBN 97888467-3045-9)
12. Sakalauskas, L., Vaiciulyte, I. (2010). Estimation of skew *t*-distribution by Monte-Carlo Markov chain approach. *Proceedings of 9th International Conference „Computer Data Analysis and Modeling“*, 207–210. (ISBN 978-985-476-847-2)
13. Sakalauskas, L., Vaiciulyte, I. (2010). Estimation of skew *t*-distribution by the Monte-Carlo Markov chain approach. *Proceedings of 1th International*

Conference „Stochastic Modeling Techniques and Data Analysis“, 747–753. (ISBN 978-981-270-968-4)

Straipsniai recenzuojamoje Lietuvos tarptautinių konferencijų medžiagoje:

14. Sakalauskas, L., Kalsyte, Z., Vaiciulyte, I., Kupciunas, I. (2013). The application of stable and skew *t*-distributions in predicting the change in accounting and governance risk ratings. *Proceedings of the 8th International Conference „Electrical and Control Technologies“*, 53–58. (**Web of Science**) (ISSN 1822-5934)
15. Vaiciulyte, I. (2012). Adaptive Monte-Carlo Markov Chain for Multivariate Statistical Estimation. *Proceedings of International Workshop „Stochastic programming for implementation and advanced applications“*, 119–124. (ISBN 9786099524146)

Disertacijos struktūra

Darbą sudaro įvadas, keturi skyriai, išvados, literatūros apžvalga ir priedai.

Įvade pateikiamas disertacijos tikslas, uždaviniai, metodai, darbo rezultatų aprobavimo ir publikavimo sąrašas.

Pirmame skyriuje aptariamas pasirinktos temos aktualumas ir bendra jos problematika.

Antrame skyriuje sudaromas Markovo grandinės Monte-Karlo algoritmas daugiamačio asimetrinio t skirstinio parametrams vertinti, pateikiamas algoritmo pritaikymas Monte-Karlo didžiausio tikėtinumo įvertinimui.

Trečiame skyriuje sudaromas daugiamatis empirinio Bajeso Puasono-Gauso modelis, aptariami kiti Bajeso skaičiavimo aspektai.

Ketvirtame skyriuje aprašomas Markovo grandinės Monte-Karlo algoritmas daugiamačio α-stabiliojo skirstinio parametrams vertinti.

1 skyrius. MARKOVO GRANDINĖS MONTE-KARLO ALGORITMŲ ANALIZINIS TYRIMAS

Šiame skyriuje pateikiama analizinė Markovo grandinės Monte-Karlo metodo apžvalga bei nagrinėjami statistinės analizės ir statistinio vertinimo uždaviniai, taikomi disertacijoje atliekamam MCMC metodo tyrimui.

1.1. Markovo grandinės Monte-Karlo metodas

MCMC metodas laikomas Hastingso algoritmo (Hastings, 1970), Metropolio metodo (Metropolis, Ulam, 1949) ir Gibso metodo (Geman ir Geman, 1984) apibendrinimas. Šis metodas leidžia generuoti sudėtingų stochastinių modelių priklausomas atsitiktines sekas. Nors yra daug MCMC algoritmų, pritaikytų daugiamatėms sekoms gauti, dažniausiai naudojami pirmieji MCMC algoritmai, t. y. Metropolio-Hastingso algoritmas ir Gibso metodas. Metropolio-Hastingso algoritmas leidžia generuoti imtis iš tikimybinio skirstinio, jei yra žinoma, kaip apskaičiuoti tam tikrą funkciją, kuri yra proporcinga šio skirstinio tankiui. Šis algoritmas generuoja imtį, kurios reikšmės kiekvienoje iteracijoje vis geriau aproksimuojamos norimu skirstiniu (Bradley ir Thomas, 2000). Gibso metodas yra atskiras Metropolio-Hastingso algoritmo atvejis, taikomas tuomet, kai nėra žinomas bendras imties skirstinys, arba jį tiesiog sunku nustatyti, tačiau yra žinomas sąlyginis kiekvieno atskiro kintamojo skirstinys (Casella ir George, 1992; Smith ir Roberts, 1993; Tanner, 1996; Tierney, 1994; Gilks ir kt., 1996). Gibso metodas yra plačiai pritaikytas įvairių modelių Bajeso analizei atlikti (Gelfand ir Smith, 1990; Carlin, 1992; Scollnik, 1996).

Statistikoje MCMC metodas sudaro algoritmų klasę tikimybinių skirstinių imitavimui, konstruojant Markovo grandinę, kuri leidžia gauti reikiamą skirstinį kaip grandinės pusiausvyros būsenos skirstinį. Markovo grandinę sudaro kelios nuosekliai sugeneruotos Monte-Karlo imtys, vadinamos grandimis, bei įverčiai, apskaičiuoti pasinaudojus šiomis imtimis. Kadangi kiekviena generuojama grandis priklauso tik nuo įverčių apskaičiuotų prieš tai buvusioje grandyje ir nepriklauso nuo ankstesnių grandžių imčių bei įverčių, iš šių grandžių sudaryta grandinė pasižymi Markovo savybe. Žinoma, generuojant imtis Markovo grandyse galima pasinaudoti kelių ankstesnių grandžių rezultatais, tačiau tokiu būdu sudaryta grandinė nebebus Markovo. Įverčiai kiekvienoje grandyje apskaičiuojami siekiant Markovo grandinės konvergavimo į stacionarią būseną. Šių įverčių apskaičiavimas priklauso nuo konkretaus sprendžiamo uždavinio, atskiri atvejai nagrinėjami kituose skyriuose.

MCMC metodas generuoja naujas grandis pagal duotus duomenis, pasiskirsčiusius pagal tam tikrą skirstinį. Čia skirstinys, nusakantis grandinės būseną po didelio žingsnių skaičiaus, yra laikomas ieškomu skirstiniu. MCMC metodas dažnai taikomas daugiamačiams integralams, sutinkamiems Bajeso statistikoje, apskaičiuoti (žr. 3 skyrių).

Skaičiuojamuoju požiūriu MCMC metodai yra taikomi spręsti lygtims, į kurias įeina sudėtingi daugialypiai integralai, sukonstruojant Monte-Karlo imčių Markovo grandinę. Šios lygtys dažnai gali būti išvedamos kaip būtinosios kokio nors stochastinio kriterijaus optimalumo sąlygos (Dennis ir Schnabel, 1996). Tokiais kriterijais gali būti tikėtinumo funkcija, dispersija, Bajeso aposteriorinės rizikos funkcija ir pan., o optimizavimo kintamaisiais būna imituojamų statistinių skirstinių parametrai. Gana dažnai galima priimti, kad šiuos kriterijus aprašančios tikslo funkcijos yra tolydžios ir glodžiai diferencijuojamos, tačiau jas apskaičiuoti būna labai sunku, kai jos yra išreiškiamos per daugialypius integralus. Pavyzdžiui, ši prielaida išplaukia iš žinomo rezultato, kad Lipšico sąlygą tenkinančios funkcijos vidurkis yra glodžiai diferencijuojama funkcija, jei neapibrėžtumą aprašanti tikimybinio tankio funkcija taip pat tenkina Lipšico sąlygą (Bartkutė, Sakalauskas, 2007). Todėl tokioms funkcijoms bei jų gradientams apskaičiuoti yra taikomas Monte-Karlo metodas, o imituojamų skirstinių parametrai gaunami stochastinės gradientinės paieškos būdu. Patogumo dėlei, Markovo grandys kartais vadinamos iteracijomis. Dažnai galima parodyti, kad statistikoje plačiai taikomas EM algoritmas yra atskiras stochastinės gradientinės paieškos atvejis (žr. 2.4 skyrelyje). Gautų įvertinių tikslumą taip pat galima įvertinti pasinaudojus (1.14), (1.21) ir (1.22) pasikliautinaisiais intervalais (žr. 1.6 skyrelį). Kadangi disertacijoje kuriami metodai remiasi stochastine gradientine paieška, jie leidžia rasti lokalųjį tikslo funkcijos ekstremumą. MCMC metodo taikymas daugiaekstremaliniams uždaviniams disertacijoje nenagrinėjamas.

Pasaulyje sukurta daug kompiuterio programų, realizuojančių MCMC metodą, pvz. *BUGS, Laplaces Demon, JAGS* ir kt., įtrauktų į *R* programų paketų sąrašą. Šioms programoms sukurti taip pat buvo parašyti algoritmai, pvz., *Hamiltonian Monte Carlo, Metropolis within Gibbs, Griddy-Gibbs, Slice Slamper, t-walk, Robust Adaptive Metropolis, Elliptical Slice Sampler,* ir kt. (http://www.bayesian-inference.com/mcmc#algorithms). MCMC metodo realizavimui reikalingi tam tikri kompiuterio laiko ir atminties ištekliai. Kompiuterių sunaudojamas laikas priklauso nuo naudojamo kompiuterio galingumo ir yra apytikriai proporcingas generuojamų Monte-Karlo bandymų skaičiui, kadangi paprastai tarpiniai skaičiavimai pareikalauja nedaug laiko. Sukurti metodai yra skirti spręsti statistiniams uždaviniams su nedideliu parametrų skaičiumi (iki 10–20), todėl problemų dėl kompiuterio atminties disertacijoje neiškyla.

Žinomuose MCMC algoritmuose paprastai sugeneruojamos kelios arba keliolika grandžių, užfiksavus visose grandyse fiksuotą bei pakankamai didelį Monte-Karlo imčių tūrį, ir nustatant konvergavimą empiriškai (Bradley ir Thomas, 2000). Aišku, kad skaičiuojamuoju požiūriu tokios procedūros yra nelabai efektyvios, nes tenka sunaudoti daug kompiuterio laiko grandžių generavimui, o nutraukus grandinės generavimą empiriškai dar gali būti nepasiektas statistiškai reikšmingas konvergavimas.

Tokiu būdu, naudojant MCMC metodą susiduriama su Markovo grandinės grandžių skaičiaus parinkimo problema. Vienas iš būdų šiai problemai spręsti yra nutraukti grandinės generavimą, jei imtys, gautos gretimose grandyse, statistiškai nesiskiria, pritaikius statistinius metodus hipotezėms apie šių imčių skirtingumą arba sutapimą patikrinti (žr. 1.7 skyrelį;

Brooks ir Gelman, 1998; Sakalauskas, 2000). Kai kurie autoriai bandė įvesti testus dviem gretimoms grandims palyginti, tačiau šie testai yra vienmačiai arba geriausiu atveju leidžia palyginti du vektorius (Brooks ir Gelman, 1998), kai praktiniuose uždaviniuose tikimybiniai skirstiniai dažnai yra aprašomi keliais vektoriais ir keliomis matricomis. Disertacijoje pasiūlyti ir išnagrinėti metodai bei algoritmai statistiniam Markovo grandžių skirtumui įvertinti pasinaudojus standartiniais Hotelingo ir Andersono kriterijais (Anderson, 1958; Krishnaiah, 1984). Hotelingo kriterijaus panaudojimas stochastiniame programavime buvo pasiūlytas ir išnagrinėtas L. Sakalausko darbuose (2000, 2002).

Kita problema, susijusi su skaičiavimo apimties sumažinimu, yra Monte-Karlo imties tūrio atskirose grandyse reguliavimas. Iš tikrųjų, konstruojant pirmąsias Markovo grandinės grandis nėra reikalo generuoti dideles Monte-Karlo imtis, nes ir nedidelio tūrio imčių pakanka modelio parametrų iteraciniam modifikavimui. Didelės Monte-Karlo imtys turėtų būti generuojamos tik grandinės pabaigoje, kai statistinis kriterijus neprieštarauja hipotezei apie paskutiniųjų grandžių tikimybinių modelių sutapimą. Disertacijoje pasiūlyti bei kompiuterinio modeliavimo būdu ištirti Monte-Karlo imties tūrio reguliavimo metodai pasinaudojus statistiniu kriterijumi apie dviejų Markovo grandžių Monte-Karlo imčių skirstinių sutapimą. Imties tūrį galima imti atvirkščiai proporcingu stabdymo statistikos ir stabdymo kriterijaus kvantilio santykiui. Pasinaudojus martingaliniais metodais yra įrodytas tokių metodų konvergavimas, sprendžiant netiesinio stochastinio programavimo uždavinius nuosekliai generuojamomis kintamo tūrio Monte-Karlo imtimis (Sakalauskas, 2000, 2002).

1.2. Monte-Karlo metodas

Kadangi pagrindinis MCMC metodo skaitmeninimo instrumentas yra Monte-Karlo metodas, šiame ir 1.6 skyrelyje aptariamos šio metodo realizacijos, naudojamos disertacijoje, ypatybės.

Monte-Karlo metodas sukurtas 1949 metais, kartu su pirmaisiais elektroniniais kompiuteriais, ir siejamas su dviejų matematikų – Džono fon Neimano (veng. *John fon Neumann*, 1903–1957) ir Stanislavo Ulamo (lenk. *Stanislaw Marcin Ulam*, 1909–1984) – vardais. Monte-Karlo metodas yra skaitmeninis algoritmas, pagrįstas statistiniu imitavimu ir gautų rezultatų apdorojimu statistiniais metodais, ir dažnai laikomas sąvokos „statistinis modeliavimas" sinonimu. Monte-Karlo metodai dažniausiai naudojami fizikinių ir matematinių sistemų modeliavimui, kai neįmanoma gauti tikslių rezultatų naudojant deterministinį algoritmą. Šio metodo esmė – atsitiktinių situacijų modeliavimas tam tikros procedūros pagalba, gaunant atsitiktinį rezultatą. Gauti rezultatai apdorojami matematiniais statistiniais metodais, nustatomi atsitiktinių dydžių pasiskirstymo tipai ir parametrai (pvz., vidurkis, dispersija, vidutinis kvadratinis nuokrypis bei kitos skaitinės charakteristikos).

Statistinio modeliavimo procedūra susideda iš tokių etapų:

1) sugeneruojamos atsitiktinių dydžių, tolygiai pasiskirsčiusių atkarpoje $[0, 1]$, realizacijos;
2) pasinaudojus šiomis realizacijomis, gaunamos atsitiktinių dydžių arba atsitiktinių procesų reikšmės su įvairiais pasiskirstymų dėsniais;
3) pasinaudojant tokiu būdu gautomis realizacijomis, apskaičiuojamos atsitiktinių dydžių, apibūdinančių tiriamą objektą, reikšmės;
4) statistiškai apdorojami gauti rezultatai;
5) pasinaudojant šiais rezultatais tikrinamos hipotezės apie nagrinėjamą objektą, įvertinamos jo charakteristikos, parenkami racionalūs objekto funkcionavimo režimai, teikiamos rekomendacijos sprendimams priimti, ir pan.

Teorinį Monte-Karlo metodo pagrindą sudaro didžiųjų skaičių dėsnis ir centrinė ribinė teorema (žr. 1.4 skyrelį).

1.3. Atsitiktinių dydžių generavimas

Statistinis modeliavimas gali būti taikomas labai plačiam tikimybinių modelių įvairiose verslo, technikos, ekonomikos ir fizikos srityse, ratui kurti bei tyrinėti. Svarbus tikimybinio modelio realizavimo kompiuteriu etapas yra atsitiktinių dydžių, pasiskirsčiusių pagal įvairius tikimybinius skirstinius, generavimas.

Atsitiktinių dydžių generatoriai kuriami pasinaudojant tuo, jog nesunku sukurti algoritmą atsitiktinio dydžio sekoms generuoti, kai yra žinoma šio atsitiktinio dydžio skirstinio išraiška.

Jei yra žinoma skirstinio funkcija, tai atsitiktinio dydžio realizaciją galima gauti apskaičiuojant jo kvantilį (atvirkštinių transformacijų metodas):

$$X = F^{-1}(U),$$

čia X pasiskirstęs pagal skirstinį $F(\cdot)$, jei atsitiktinis dydis U yra tolygiai pasiskirstęs intervale $[0, 1]$.

Taigi, galima generuoti įvairiai pasiskirsčiusių atsitiktinių dydžių sekas, jei žinomas skirstinio pavidalas ir tolygiai pasiskirsčiusių vienetiniame intervale atsitiktinių dydžių gavimo būdas. Reikia pažymėti, kad to paties atsitiktinio dydžio reikšmių sekas galima gauti įvairiais būdais. Labai dažnai efektyvus generavimo algoritmas sukuriamas pasinaudojant specifinėmis nagrinėjamo atsitiktinio dydžio savybėmis.

1.3.1. Priėmimo-atmetimo metodas

Nagrinėjamas priėmimo-atmetimo metodas atsitiktiniam dydžiui X generuoti, kai X pasiskirstęs pagal tolydųjį dėsnį $F(x)$ ir žinoma skirstinio tankio funkcija $f(x)$. Tarkime atsitiktinio dydžio Y, kurio generavimo algoritmas žinomas, tankio funkcija $g(x)$ yra tokia, kad santykis

$f(x)/g(x) < \infty$ visoms galimoms *Y* reikšmėms. Nesunku pastebėti, jog tuomet egzistuoja toks skaičius $1 \le c < \infty$, jog $f(x)/g(x) \le c$. Atsitiktinio dydžio *X* realizacijai gauti generuojamos tolygiai intervale $[0, 1]$ pasiskirsčiusių skaičių *U* ir nepriklausomų atsitiktinių dydžių *Y* poros, kol $U \le \frac{f(Y)}{c \cdot g(Y)}$. Tuomet *X* atsitiktiniam dydžiui priskiriama reikšmė *Y* (Gentle, 2003).

Pažymėkime dviejų skaičių *Y* ir *U* porų skaičių *l*, kurias reikia sugeneruoti, norint gauti vieną *X* reikšmę priėmimo-atmetimo metodu. Nesunku pastebėti, jog šis skaičius yra pasiskirstęs pagal geometrinį dėsnį su parametru $p = \frac{1}{c}$. Pasinaudojant tuo, jog $EX = \frac{1-p}{p}$, galima gauti, kad vidutinis generuojamų porų *U* ir *Y* skaičius, kol gaunama viena *X* reikšmė, yra lygus $c - 1$. Taigi, sudarant priėmimo-atmetimo algoritmą, svarbu parinkti *c*, kiek įmanoma mažesniu, ir atitinkamai atsitiktinio dydžio *Y* generavimo algoritmas turi būti taip pat kuo paprastesnis.

Pavyzdžiui, atsitiktiniams α-stabiliesiems dydžiams su parametrais α, β, σ, μ generuoti galima pritaikyti tokį priėmimo-atmetimo algoritmą (Janicki, Weron, 1993):

$$Y = B \cdot \frac{\sin(\alpha \cdot (V + G))}{\cos^{\frac{1}{\alpha}}(V)} \cdot \left(\frac{\cos(V - \alpha \cdot (V + G))}{W} \right)^{\frac{1-\alpha}{\alpha}},$$

čia *V* tolygiai pasiskirstęs intervale $\left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$, o *W* pasiskirstęs pagal eksponentinį dėsnį su vienetiniu parametru ir

$$B = \left(1 + \beta^2 \cdot tg^2\left(\alpha \cdot \frac{\pi}{2} \right) \right)^{\frac{1}{2\alpha}},$$

$$G = \frac{1}{\alpha} \operatorname{arctg}\left(\beta \cdot tg\left(\alpha \cdot \frac{\pi}{2} \right) \right).$$

Finansiniai duomenys ir informaciniai srautai kompiuterių tinkluose dažnai adekvačiai aprašomi stabiliaisiais dėsniais (žr. 1.4 skyrelį, Kabasinskas ir kt., 2009; Kaklauskas, Sakalauskas, 2009).

1.3.2. Daugiamačių atsitiktinių vektorių generavimas

Sudėtingoms sistemoms aprašyti gali prireikti ne vieno, o daugelio atsitiktinių kintamųjų. Paprasčiausiu atveju galima nagrinėti šiuos kintamuosius atskirai, pritaikius kiekvienam atitinkamą vienmatį tikimybinį modelį. Tačiau toks būdas tinka tik nepriklausomų kintamųjų atveju. Kelių atsitiktinių kintamųjų tikimybinį modelį bendru atveju galima aprašyti daugiamačiu skirstiniu, kurio pasiskirstymo funkcija išreiškiama per daugialypį integralą (Kubilius, 1996)

$$F(x_1, x_2, \ldots, x_d) = \int_{z \le x} p(z) dz = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_d} p(z_1, z_2, \ldots, z_d) dz_1 dz_2 \ldots dz_d \, ,$$

čia $p(z)$ yra daugiamačio skirstinio tankio funkcija.

Daugiamatis normalusis atsitiktinis vektorius su nežinomais skirstinio parametrais – vidurkių vektoriumi μ ir kovariacijų matrica Ω – yra dažnai naudojamas daugiamačių duomenų statistinis modelis. Daugiamačio normaliojo atsitiktinio vektoriaus $X = (X_1, X_2, \ldots, X_d)$ tankio funkcija išreiškiama pavidalu

$$p(x_1, x_2, \ldots, x_d) = \left((2\pi)^{\frac{d}{2}} \cdot \sqrt{\Omega} \right)^{-1} \cdot \exp\left(-\frac{1}{2} (x - \mu)^T \Omega^{-1} (x - \mu) \right),$$

čia $\mu = (\mu_1, \mu_2, \ldots, \mu_d)$, $\Omega = (\sigma_{ij})_1^d$, yra skirstinio parametrai. Nesunku įsitikinti, kad normaliojo skirstinio vidurkis ir kovariacijų matrica yra

$$EX = \mu,$$
$$\left(\operatorname{cov}\left(X_i, X_j\right)\right)_1^d = \Omega.$$

Vienmačiu atveju tankio funkcija lygi

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Dvimačio normaliojo vektoriaus $X = (X_1, X_2)$ tankio funkcija užrašoma taip:

$$p(x_1, x_2) = \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2 \cdot \sqrt{1-\rho^2}} e^{-\frac{1}{2\sqrt{1-\rho^2}}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)},$$

čia $\mu_1 = EX_1$, $\mu_2 = EX_2$, $\sigma_1^2 = DX_1$, $\sigma_2^2 = DX_2$, $\rho = \dfrac{\operatorname{cov}(X_1, X_2)}{\sigma_1\sigma_2}$.

Generuojant didesnio negu 2 matavimo priklausomus atsitiktinius vektorius *X*, pasiskirsčiusius pagal normalųjį skirstinį, kovariacinė matrica faktorizuojama Choleckio metodu (Čiegis, Būda, 1997). Tegu Σ yra matrica, $\Omega = \Sigma^T \cdot \Sigma$. Tuomet atsitiktinis vektorius $X \sim N(\mu, \Omega)$, modeliuojamas vektoriumi

$$X = \Sigma \cdot \eta + \mu,$$

čia $\eta = (\eta_1, \eta_2, \ldots, \eta_d)$ atsitiktinis vektorius, kurio komponentės yra nepriklausomi atsitiktiniai dydžiai, pasiskirstę pagal standartinį normalųjį skirstinį. Matricą Σ patogu parinkti trikampe

$$\Sigma = \begin{vmatrix} q_{11} & 0 & \cdots & 0 \\ q_{21} & q_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ q_{d1} & q_{d2} & \cdots & q_{dd} \end{vmatrix}.$$

Jos koeficientus galima nustatyti Choleckio metodu:

$$q_{11} = \sqrt{\Omega_{11}}\,,\; q_{j1} = \frac{\Omega_{j1}}{q_{11}},\; q_{ii} = \sqrt{\Omega_{ii} - \sum_{j=1}^{i-1} q_{ij}^2}\,,\; q_{ij} = \frac{\Omega_{ij} - \sum_{h=1}^{j-1} q_{ih} q_{jh}}{\sqrt{\Omega_{jj} - \sum_{h=1}^{j-1} q_{jh}^2}}\,,$$

kai $2 \le j < i \le d$, ir $\Omega_{ij} = 0$, kai $j > i$.

1.4. Didžiųjų skaičių dėsnis ir centrinė ribinė teorema

Monte-Karlo metodo teorinį pagrindą sudaro didžiųjų skaičių dėsnis (DSD) ir centrinė ribinė teorema (CRT) (Aksomaitis, 2000; Stepanauskas, 2006; Sakalauskas, 2013).

Tarkime, $X^1, X^2, \ldots, X^N$ yra nepriklausomų vienodai pasiskirsčiusių atsitiktinių dydžių seka. DSD tvirtina, kad kai N didelis, atsitiktinis dydžių aritmetinis vidurkis yra beveik pastovus, t. y. mažai skiriasi nuo atsitiktinių dydžių skirstinio vidurkio (Aksomaitis, 2000; Kubilius, 1996):

$$\lim_{N\to\infty} P\left(\left|\bar{x} - \mu\right| < \varepsilon\right) = 1, \tag{1.1}$$

čia $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} X^i$, $EX^1 = EX^2 = \ldots = EX^N = \mu$.

Žinomos kelios DSD formuluotės. Pavyzdžiui, silpnasis DSD teigia, kad (1.1) konverguoja į 0 pagal tikimybę, stiprusis DSD – kad (1.1) konverguoja į 0 su tikimybe 1. Beje, stiprusis DSD tvirtina, kad beveik visada kiekvienai skaičių porai $\varepsilon > 0$, $\delta > 0$ galima rasti tokį N_0, kad

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N} X^i - \mu\right| > \varepsilon\right) > 1 - \delta\,,$$

kai $N > N_0$ (Kubilius, 1996). Kolmogorov (1956) įrodė, kad ši nelygybė teisinga, jei egzistuoja atsitiktinio dydžio X vidurkis: $|EX| < \infty$. Didžiųjų skaičių dėsnis taip pat galioja nepriklausomų skirtingai pasiskirsčiusių atsitiktinių dydžių atveju, jei tenkinami tam tikri reikalavimai jų dispersijoms (pvz., Lindebergo sąlyga). Nepriklausomų atsitiktinių dydžių klasėje pakankamas DSD sąlygas yra apibūdinęs P. Čebyšovas (Kubilius, 1996).

Praktikoje pasitaikančių atsitiktinių dydžių pasiskirstymai labai dažnai būna normalieji arba mažai nuo jų skiriasi, tad normalusis skirstinys užima centrinę vietą tarp kitų skirstinių. Nors atskirų dėmenų pasiskirstymas paprastai gali būti nežinomas, dėmenų sumos pasiskirstymas dažnai yra artimas normaliajam (Kubilius, 1996). CRT leidžia įvertinti skirtumą tarp aritmetinio vidurkio ir atsitiktinio dydžio vidurkio. Tarkime atsitiktinių dydžių X^i, $i = 1, 2, \ldots, N$ sumos $S_N = X^1 + X^2 + \ldots + X^N$ vidurkis ES_N, o sumos dispersija DS_N. CRT tvirtina, kad, jei šių atsitiktinių dydžių aukštesnės eilės centriniai momentai tenkina gana bendras sąlygas (žr. Kubilius, 1996), tai dėmenų skaičiui didėjant jų sumos skirstinys artėja į normalųjį skirstinį, t. y.

$$P\left(\frac{S_N - ES_N}{\sqrt{DS_N}} < x \right) \Rightarrow \Phi(x), \quad \text{kai } N \to \infty, \qquad (1.2)$$

čia $x \in [-\infty, \infty]$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$ yra standartinio normaliojo atsitiktinio dydžio pasiskirstymo funkcija. Iš čia išplaukia, kad šių atsitiktinių dydžių aritmetinio vidurkio skirstinys taip pat konverguoja su tikimybe 1 į normalųjį skirstinį dėmenų skaičiui be galo didėjant.

Atskiru atveju, jei atsitiktiniai dydžiai yra vienodai pasiskirstę su vidurkiais μ ir dispersijomis σ^2, tai

$$\lim_{N \to \infty} P\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} < x \right) = \Phi(x), \quad x \in [-\infty, \infty]. \tag{1.3}$$

(1.2) ir (1.3) aproksimacijų tikslumas priklauso nuo sumuojamų atsitiktinių dydžių skaičiaus ir pasiskirstymo funkcijos argumento *x*.

Konvergavimo greitį CRT nustato Beri-Eseno (angl. *Berry-Esseen*) nelygybė

$$\sup_x \left| P\left(\frac{\bar{x}_N - \mu}{\frac{\sigma}{\sqrt{K}}} < x \right) - \Phi(x) \right| \leq \frac{C \cdot \rho_3}{\sqrt{K}}, \tag{1.4}$$

čia $\rho_3 = E|X - EX|^3$, $0.4097 < C < 0.7056$ – tam tikra konstanta (Kubilius, 1996). Neseniai patikslintas šios konstantos viršutinio rėžio įvertis dėl vienodai pasiskirsčiusių atsitiktinių dydžių – $C < 0.4784$ (Korolev ir Shevtsova, 2012) ir dėl skirtingai pasiskirsčiusių atsitiktinių dydžių – $C < 0.5600$ (Shevtsova, 2010).

(1.2) ir (1.3) aproksimacijų santykinis tikslumas mažėja, pasiskirstymo funkcijos argumentui *x* didėjant absoliutiniu didumu. Tai reiškia, kad ši aproksimacija netaikytina labai mažoms arba labai didelėms argumento *x* reikšmėms. Rekomenduojamas intervalas, kuriame atsitiktinių dydžių vidurkio skirstiniui galima taikyti normaliąją aproksimaciją, apytiksliai lygus $|x| < A \cdot K^{\alpha}$, čia $0 < \alpha < \frac{1}{2}$, $A > 0$ – tam tikri skaičiai, priklausantys nuo sumuojamų atsitiktinių dydžių.

Konvergavimo greitį ir normaliosios aproksimacijos tikslumą galima įvertinti, apskaičiavus sumuojamų atsitiktinių dydžių trečiuosius momentus, pasinaudojant Beri-Eseno teorema.

Reikia pažymėti, kad atitinkamas aproksimacijas taip pat galima taikyti statistikų (pvz., Hotelingo, χ^2 statistika) skirstiniams, kurie apskaičiuojami, pasinaudojant normaliuoju dėsniu ir aproksimuojant tuos skirstinius χ^2 arba Fišerio skirstiniais.

DSD ir CRT taip pat galioja daugiamačiu atveju. Reikia pažymėti, kad daugiamačiu atveju dar nėra gauti įvertiniai, analogiški Beri-Eseno teoremai (1.4), taip pat nėra žinomi analogiški Beri-Eseno teoremos pavidalo rezultatai Fišerio skirstinio atveju, tačiau yra nustatytos Beri-Eseno įvertiniai χ^2 skirstinio atvejui (Bentkus ir kt., 1996).

CRT plačiai taikoma tikimybiniams-statistiniams uždaviniams spręsti technikoje, biologijoje, sociologijoje, susijusiems su didelio skaičiaus mažų veiksnių įvertinimu. Tačiau reikia pažymėti, kad statistiniai finansinių rinkų bei informacinių srautų telekomunikacijų ir kompiuterių tinkluose tyrimai rodo, kad tokiems srautams gali būti netenkinamos CRT sąlygos, iš kurių išplaukia normalioji aproksimacija. Mat tokių modelių atsitiktiniai dydžiai gali neturėti dispersijos (Kabasinskas ir kt., 2009; Kaklauskas, Sakalauskas, 2009). Pastaruoju atveju taikoma apibendrinta CRT, kurioje atsitiktinių dydžių vidurkis konverguoja į α-stabilųjį dėsnį, $\alpha \in (0, 2]$, pasižymintį Pareto savybe, t. y. jo nuokrypių tikimybės asimptotiškai yra aprašomos laipsniniu dėsniu (Sakalauskas, 2013):

$$\Pr(X \geq x) \approx \frac{\sigma^{\alpha}(1+\beta)\sin\left(\pi \frac{\alpha}{2}\right) \cdot \Gamma(\alpha)}{\pi \cdot x^{\alpha}}.$$

Šia savybe pasinaudojama modeliuojant fraktalinius procesus, dažnai pasitaikančius finansuose, versle, informacijos tinkluose ir pan. (Rachev, Xin, 1993).

1.5. Statistinių modelių parametrų vertinimas

Markovo grandinės Monte-Karlo metodas dažnai taikomas tikimybinių statistinių modelių parametrų įvertiniams gauti. MCMC metodas taip pat dažnai pritaikomas konstruojant EM algoritmus, taikomus įvertiniams apskaičiuoti didžiausio tikėtinumo (žr. 1.5.1 skyrelį) arba Bajeso metodais (žr. 1.5.2 skyrelį). EM algoritmas atliekamas šiais žingsniais (Archambeau ir kt., 2003; Boyles, 1983; Wu, 1983):

E-žingsnis: logaritminės tikėtinumo funkcijos sąlyginio vidurkio apskaičiavimas

$$L(\theta^i) = E\left[\log L\left(X^i \middle| \theta\right)\right],$$

M-žingsnis: tikėtinumo funkcijos sąlyginio vidurkio maksimizavimas

$$\theta^{i+1} \longrightarrow \max_{\theta} L(\theta^i).$$

Iš kitų stochastinių optimizavimo algoritmų EM algoritmai išsiskiria santykiniu stabilumu ir mažesniu jautrumu pradinių reikšmių parinkimui. Šie algoritmai yra pritaikomi 2, 3 ir 4 skyriuose Monte-Karlo Markovo grandinėms konstruoti.

1.5.1. Didžiausio tikėtinumo metodas

Statistikoje didžiausio tikėtinumo metodas (angl. *maximum likelihood method* – MLM) – tai metodas, skirtas statistinio modelio parametrams vertinti. Fiksuotai nepriklausomų vienodai pasiskirsčiusių duomenų imčiai MLM leidžia gauti modelio parametrų rinkinių reikšmes, maksimizuojančias tikėtinumo funkciją. Šiuo metodu surandamos tokios parametrų reikšmės, su kuriomis gaunamieji rezultatai tampa labiausiai tikėtini duotajam modeliui (Li, Stephens, 2003).

Kuo imtis didesnė, tuo labiau tikėtina, kad didžiausio tikėtinumo metodu gauti parametrų įvertiniai mažai skirsis nuo tikrųjų parametro reikšmių. Esant gana bendroms sąlygoms (Durrett, 2010; Owen, 2002) gauti įvertiniai yra:

- ✓ pagrįsti (konverguoja pagal tikimybę į nežinomo ir vertinamo parametro reikšmę),
- ✓ asimptotiškai normalieji,
- ✓ efektyvūs (asimptotiškai turi mažiausią dispersiją tarp visų galimų nežinomo parametro įvertinių).

Monte-Karlo metodo požiūriu ypač svarbios pirmosios dvi savybės, kadangi iš pagrįstumo savybės išplaukia, kad šie įvertiniai, imties tūriui didėjant, konverguoja į vertinamo parametro reikšmę, o asimptotinio normalumo savybė leidžia sukonstruoti nesudėtingas taisykles gauto įvertinio pasikliautiniesiems intervalams įvertinti, analogiškus (1.14), (1.21) ir (1.22) (žr. 1.6 skyrelį).

Tarkime, atsitiktinis vektorius $X \in \Re^d$ yra aprašomas tikimybiniu tankiu $p(x|\theta)$, čia $\theta \in \Re^d$ yra nežinomų parametrų vektorius. Duota X nepriklausomi vienodai pasiskirstę atsitiktiniai dydžiai

$$X^1, X^2, \ldots, X^K, \tag{1.5}$$

su tankio funkcija $p(x|\theta)$. Atsitiktinio dydžio tankio funkcijų sandauga vadinama tikėtinumo funkcija

$$L(\theta) = \prod_{i=1}^{K} p\left(X^i \middle| \theta\right).$$

Didžiausio tikėtinumo įverčiais parenkamos tokios parametrų θ reikšmės, su kuriomis funkcija $L(\theta)$ įgyja didžiausią reikšmę, o pats parametrų θ įvertinys vadinamas didžiausio tikėtinumo įvertiniu (Owen, 2002). Skaičiavimams palengvinti funkcija $L(\theta)$ logaritmuojama.

Jei tikėtinumo funkcija diferencijuojama, tai funkcijos $-\ln L(\theta)$ minimumas ieškomas taip:

1) surandamos dalinės išvestinės $\frac{dL(\theta)}{d\theta_j}$, $j = 1, 2, \ldots, n$;
2) išvestinės prilyginamos nuliui ir sprendžiama lygčių sistema

$$\frac{dL(\theta)}{d\theta_j} = 0, \quad j = 1, 2, \ldots, n, \tag{1.6}$$

   su n lygčių ir n nežinomųjų, kuri gana dažnai turi vienintelį sprendinį;
3) pasinaudojus (1.5) atsitiktinių dydžių skirstinio parametrų didžiausio tikėtinumo įvertinių asimptotiniu normališkumu apskaičiuojami gautų didžiausio tikėtinumo įvertinių pasikliautinieji intervalai.

Parametrų didžiausio tikėtinumo įvertinių apskaičiavimas, sprendžiant skaitmeniniais metodais (1.6), yra iteracinis procesas, kuriam reikia parinkti pradines įvertinių reikšmes, ir atlikti tam tikrą iteracijų kiekį, kol gretimų iteracijų įverčiai pradės skirtis pakankamai mažai. Didžiausio tikėtinumo metodo realizavimas Markovo grandinės Monte-Karlo metodu yra tiriamas kituose skyriuose.

1.5.2. Bajeso metodas

Tarkime, nežinomų parametrų vektorius θ yra pasiskirstęs pagal tam tikrą apriorinį skirstinį $p(\theta)$. θ parametrams įvertinti dažniausiai naudojama mažiausia vidutinė kvadratinė paklaida (angl. *minimum mean square error*). Tokiu atveju nežinomo parametro Bajeso įvertinys apskaičiuojamas taip (Berger, 1985):

$$\hat{\theta}(x) = E(\theta|x) = \int \theta \cdot p(\theta|x) d\theta. \tag{1.7}$$

Norint įvertinti θ parametrus pagal (1.7) formulę, reikia žinoti jų aposteriorinį skirstinį arba jo tankio funkciją $p(\theta|x)$. Pagal Bajeso formulę šis aposteriorinis tankis apskaičiuojamas taip:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}, \tag{1.8}$$

čia $p(\theta)$ – žinomas skirstinys, $p(x|\theta)$ – imties tikėtinumo funkcija, $p(x) = \int p(x|\theta)p(\theta)d\theta$ – apriorinis tankis (angl. *prior predictive distribution*).

Įvertinių ir sprendimų teorijoje Bajeso metodas leidžia gauti įvertinį arba sprendimo taisyklę, minimizuojančią sąlyginę nuostolio (rizikos) funkcijos $L(\theta, \hat{\theta})$ reikšmę (arba maksimizuojančią sąlyginę naudingumo funkcijos reikšmę).

Nesunku pastebėti, kad jei

$$L(\theta) = \begin{cases} \infty, & \text{jei } \theta = \hat{\theta}, \\ 0, & \text{jei } \theta \neq \hat{\theta}, \end{cases}$$

tai Bajeso įvertiniai sutampa su didžiausio tikėtinumo įvertiniais (žr. 1.5.1 skyrelį).

Bajeso metodas leidžia tirti hipotezes, apskaičiuoti įvykių tikimybes. Kadangi apie retai pasirodančius įvykius surinkta nedaug statistinių duomenų, tai disertacijoje plėtojami tikimybinio vertinimo metodai retų įvykių tikimybėms vertinti (žr. 3 skyrių).

1.6. Apytikslis integralų skaičiavimas Monte-Karlo metodu

Tikimybių teorijoje ir statistikoje dažnai tenka apskaičiuoti įvairių atsitiktinių dydžių bei vektorių vidurkius, išreiškiamus sudėtingais daugialypiais integralais. Panaudojant DSD ir CRT daugialypius integralus galima apskaičiuoti Monte-Karlo metodais. Sakykim, nagrinėjamas toks daugialypio integralo apskaičiavimo uždavinys:

$$\mathrm{I} = \int_{z \in A} u(z) p(z) dz \,, \tag{1.9}$$

čia $A \subset \Re^d$ yra integravimo sritis (Borelio aibė), $u(z)$ – integruojama funkcija, $p(z)$ – tankio funkcija, $z = (z_1, z_2, \ldots, z_d)$.

Jei integravimo sritis sutampa su kintamųjų erdve $\Re^d$, skaičiuojamas integralas

$$\mathrm{I} = \int_{\Re^d} u(z) \cdot p(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u(z_1, z_2, \ldots, z_d) \cdot p(z_1, z_2, \ldots, z_d) dz_1 dz_2 \ldots dz_d \,. \tag{1.10}$$

Pažymėsime, kad neneigiamą funkciją $p(z_1, z_2, \ldots, z_d)$ galima laikyti daugiamačio atsitiktinio vektoriaus tankio funkcija, jei tenkinama normavimo sąlyga (Durrett, 2010):

$$\int_{\Re^d} p(z) dz \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(z_1, z_2, \ldots, z_d) dz_1 dz_2 \ldots dz_d = 1 \,. \tag{1.11}$$

Tokiu būdu, galima nustatyti sąryšį tarp daugialypio integralo (1.10) ir funkcijos u vidurkio nuo atsitiktinio vektoriaus $X = (X_1, X_2, \ldots, X_d)$, jei tenkinama (1.11) sąlyga,

$$\mathrm{I} = Eu(X) = Eu(X_1, X_2, \ldots, X_d). \tag{1.12}$$

(1.10) integralą arba (1.12) vidurkį galima įvertinti statistinio modeliavimo būdu. Tegul žinomas būdas atsitiktiniams vektoriams, pasiskirsčiusiems su tankio funkcija $p(z_1, z_2, \ldots, z_d)$, generuoti, ir yra sugeneruota tokių vektorių N tūrio imtis X^i, $i = 1, 2, \ldots, N$ bei apskaičiuotos pointegrinės funkcijos $u(\cdot)$ reikšmės $u\left(X^i\right)=u\left(X_1^i, X_2^i, \ldots, X_d^i\right),\ i=1, 2, \ldots, N$. Tuomet, pasinaudojus DSD ir CRT (žr. 1.4 skyrelį), (1.10) daugialypį integralą apytiksliai galima apskaičiuoti taip:

$$\mathrm{I}_N = \frac{1}{N}\sum_{i=1}^{N} u\left(X^i\right) = \frac{1}{N}\sum_{i=1}^{N} u\left(X_1^i, X_2^i, \ldots, X_d^i\right). \tag{1.13}$$

Apskaičiavus imties dispersiją

$$D_N = \frac{1}{N-1}\sum_{i=1}^{N}\left(u\left(X^i\right) - \mathrm{I}_N\right)^2$$

ir padarius prielaidą apie I_K normalųjį skirstinį, galima įvertinti daugialypio integralo $1-2\beta$ lygmens pasikliautinojo intervalo įvertį

$$\left[\mathrm{I}_N - \eta_\beta \cdot \sqrt{\frac{D_N}{N}}, \mathrm{I}_N + \eta_\beta \cdot \sqrt{\frac{D_N}{N}}\right], \tag{1.14}$$

čia η_β yra normaliojo skirstinio (arba t skirstinio, jei norima gauti tikslesnį įvertį) β lygmens kvantilis, $\beta \in (0, 1)$. Pasikliautinąją tikimybę β galima patikslinti pasinaudojus Beri-Eseno nelygybe (1.4).

Gana dažnai tenka apskaičiuoti įvertinius, kurie priklauso netiesiškai nuo Monte-Karlo įvertinių (1.13), pvz.:

$$\hat{L} = -\sum_{i=1}^{K}\ln\left(P_i\right), \tag{1.15}$$

čia $P_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij}$, u_{ij}– nepriklausomi vienodai pasiskirstę atsitiktiniai dydžiai, $Eu_{ij} = \mu_i$, N – Monte-Karlo imties dydis, K – koks nors skaičius. Norint įvertinti šios funkcijos pasikliautinąjį intervalą, reikia žinoti jos imties dispersiją. Iš (1.15) išplaukia lygybė

$$\hat{L} = -\sum_{i=1}^{K} \ln(\mu_i) - \sum_{i=1}^{K} \ln\left(1 + \frac{P_i - \mu_i}{\mu_i}\right). \tag{1.16}$$

Pasinaudojus gerai žinoma nelygybe

$$\left|\ln(1+t) - t\right| \le \frac{t^2}{2},$$

ir (1.16) lygybe, galima gauti

$$\left|\hat{L} + \sum_{i=1}^{K} \ln(\mu_i) + \sum_{i=1}^{K} \frac{P_i - \mu_i}{\mu_i}\right| \le \frac{1}{2} \sum_{i=1}^{K} \left(\frac{P_i - \mu_i}{\mu_i}\right)^2. \tag{1.17}$$

Jei imtis sudaryta iš nepriklausomų atsitiktinių dydžių ir $EP_i = \mu_i$, tai nesunku įsitikinti, kad dydis dešiniojoje šios nelygybės pusėje kinta kaip $O\left(\frac{1}{N}\right)$, kai imties tūris N didėja. Iš čia išplaukia, kad (1.15) yra pagrįstas išraiškos

$$L = -\sum_{i=1}^{K} \ln(\mu_i)$$

įvertinys (angl. *consistent estimator*), nes:

$$E\hat{L} = -\sum_{i=1}^{K} \ln(\mu_i) + O\left(\frac{1}{N}\right). \tag{1.18}$$

Panašiai iš (1.17) ir (1.18) gaunama, kad (1.15) įvertinio dispersija yra lygi

$$D\hat{L} = \sum_{i=1}^{K} \frac{E(P_i - \mu_i)^2}{\mu_i^{\,2}} + O\left(\frac{1}{N}\right). \tag{1.19}$$

Iš čia ir DSD gaunama, kad tikėtinumo funkcijos Monte-Karlo (1.15) įvertinio dispersija aproksimuojama taip:

$$D\hat{L} \approx \sum_{i=1}^{K} \left(\frac{P1_i}{P_i^2} - 1 \right), \tag{1.20}$$

čia $P1_i = \dfrac{1}{N} \sum_{j=1}^{N} u_{ij}^{\;2}$.

Tad, atlikus nesudėtingus pertvarkymus ir pasinaudojus asimptotiniu Monte-Karlo įvertinių normališkumu, (1.15) įvertinio 95% pasikliautinieji rėžiai aproksimuojami taip:

$$\left[\hat{L} - \frac{2}{\sqrt{N}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P1_i}{P_i^2} - 1 \right)}, \hat{L} + \frac{2}{\sqrt{N}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P1_i}{P_i^2} - 1 \right)} \right]. \tag{1.21}$$

(1.9) integralo apskaičiavimą galimą suvesti į (1.12) integralo vertinimą aprašytuoju būdu įvedus pointegrinę funkciją

$$U(z_1, z_2, \ldots, z_d) = H(z_1, z_2, \ldots, z_d) \cdot u(z_1, z_2, \ldots, z_d),$$

čia $H(z_1, z_2, ..., z_d)$ yra indikatorinė integravimo aibės funkcija:

$$H(z_1, z_2, \ldots, z_d) = \begin{cases} 1, & z_1, z_2, \ldots, z_d \in A, \\ 0, & z_1, z_2, \ldots, z_d \notin A. \end{cases}$$

Jeigu $u(z_1, z_2, \ldots, z_d) \equiv 1$, tai (1.9) integralo apskaičiavimas leidžia įvertinti atsitiktinio vektoriaus, pasiskirsčiusio su tankio funkcija $p(z_1, z_2, \ldots, z_d)$, patekimo į sritį A tikimybę:

$$\Pr(x \in A) = \int_{z \in A} p(z)dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} H(z_1, z_2, \dots, z_d) \cdot p(z_1, z_2, \dots, z_d) dz_1 dz_2 \dots dz_d .$$

Vertinant šią tikimybę statistinio modeliavimo būdu sugeneruojama atsitiktinių vektorių, pasiskirsčiusių su tankio funkcija $p(z_1, z_2, \dots, z_d)$, tūrio N imtis ir nustatomas atsitiktinių vektorių, patekusių į sritį $A \subset \Re^d$, skaičius m. Patikrinus (1.4) Beri-Eseno sąlygą, kuri nagrinėjamu atveju atrodo taip (Durrett, 2010):

$$m\left(1 - \frac{m}{N}\right) > 6,$$

gaunamas tikimybės $\Pr(x \in A)$ įvertinys

$$\hat{p} = \frac{m}{N},$$

ir pasikliautinasis intervalas

$$\left[\hat{p} - \eta_\beta \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N}},\ \hat{p} + \eta_\beta \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N}}\right], \qquad (1.22)$$

čia $1 - 2\beta$ pasikliovimo lygmuo.

Statistinio modeliavimo įvertinių (1.15) ir (1.20) pasikliautinųjų intervalų tyrimas rodo, kad šių įvertinių pasikliovimo intervalo ilgis yra atvirkščiai proporcingas $\sqrt{N}$, čia N – imties tūris. Taigi, norint sumažinti pasikliovimo intervalą du kartus, t. y. padidinti tikslumą du kartus, tenka generuoti keturis kartus didesnę imtį ir t. t.

Statistinių įvertinių pasikliautinųjų intervalų vertinimas yra ne mažiau svarbus uždavinys negu pačių įvertinių apskaičiavimas. Labai dažnai statistinio modeliavimo įvertiniai turi būti gauti tam tikru tikslumu, pavyzdžiui, pasikliautinieji intervalai turi būti reikiamo ilgio. Pasikliautinojo intervalo ilgis, kuris yra svarbiausias įvertinio tikslumo matas, gali būti reguliuojamas, atitinkamai parenkant Monte-Karlo imties tūrį N. Tegu intervalo (1.14), (1.21) arba (1.22) ilgis neturi viršyti ε, $\varepsilon > 0$. Iš (1.22) išplaukia išraiška imties tūriui, kuris būtų pakankamas norimo ilgio pasikliautinajam intervalui gauti,

$$N \geq \frac{\eta_\beta^2 \cdot D_N}{4 \cdot \varepsilon^2},$$

čia D_N yra vertinamo atsitiktinio dydžio dispersija. Tačiau pritaikyti šią formulę iš anksto, reikalingam imties tūriui apskaičiuoti ne visada įmanoma, nes į ją įeina nežinoma imties dispersija D_N.Taigi, generuojant atsitiktinę imtį, galima rekurentiškai skaičiuoti sumas, įeinančias į imties vidurkį bei dispersiją, ir generavimą nutraukti, pasiekiamus reikiamą Monte-Karlo įvertinių tikslumą.

1.7. Statistinių hipotezių tikrinimas

Tegu yra generuojama Markovo grandinė. Sugeneravus vieną šios grandinės grandį Monte-Karlo metodu yra gaunama atsitiktinė imtis, kuri po to yra apdorojama tikimybiniais statistiniais metodais. Vienas iš uždavinių, susijusių su tokiu būdu gautomis imtimis, yra statistinės hipotezės apie kelių Markovo grandžių skirtumą tikrinimas.

1.7.1. Statistinių hipotezių tikrinimo principai

Statistinių hipotezių tikrinimas – yra metodas, leidžiantis patikrinti bei pateikti tam tikras išvadas, pasiremiant imties duomenimis. Statistiniai kriterijai, pirmąkart pasiūlyti D. Bernulio, o vėliau ir K. Pirsono bei R. Fišerio, vaidina lemiamą vaidmenį statistinių duomenų analizėje. Šiame skyrelyje aptariami pagrindiniai hipotezių tikrinimo žingsniai.

1. Duomenys.

Tarkime, turime atsitiktinę imtį $X^1, X^2, \ldots, X^K$, kurios skirstinys priklauso nuo parametro θ.

2. Statistinių hipotezių formulavimas.

Hipotezių tikrinimas pradedamas iškeliant prielaidas apie atsitiktinių dydžių tikimybinį skirstinį, priklausantį nuo nežinomų parametrų. Viena iš galimų hipotezių yra pagrindinė – nulinė hipotezė H_0, o visos kitos yra alternatyvios hipotezės H_1. Sprendimo tikslas yra nustatyti, ar hipotezė H_0 gali būti teisinga.

3. Statistinio kriterijaus sudarymas.

Iškėlus šias prielaidas, sudaromas statistinis kriterijus, pagal kurį nustatoma ar stebėjimo duomenys neprieštarauja nulinei hipotezei. Tarkime, reikia patikrinti nulinę hipotezę H_0: $\theta = \theta_0$ su alternatyva H_1: $\theta = \theta_1$. Sudaroma imties duomenų funkcija – statistika H^K, – kurios skirstinys žinomas, kai H_0 teisinga. Statistiniai kriterijai dažniausiai konstruojami taip, kad jų statistikos turėtų normalųjį, χ^2, Stjudento arba Fišerio pasiskirstymus.

4. Hipotezės priėmimo/atmetimo taisyklė.

Statistikos H^K galimų reikšmių aibę Q suskaidome į dvi sritis Q_1 (kritinė sritis) ir Q/Q_1. Jei statistikos reikšmė patenka į sritį Q_1, tai hipotezė H_0 prieštarauja imties duomenims ir yra atmestina, o jei į sritį Q/Q_1, hipotezė H_0 yra priimtina. Kad statistika nepatektų į kritinę sritį, kai H_0 teisinga (pirmosios rūšies klaida), parenkama maža sąlyginė tikimybė (Čekanavičius, Murauskas, 2000):

$$P\{H^K \in Q_1 | H_0\} = \gamma,$$

čia γ yra reikšmingumo lygmuo (klaidos tikimybė). Paprastai γ parenkama 0,1; 0,05; 0,01.

5. Išvada. Jei statistikos H^K reikšmė priklauso atmetimo sričiai, tuomet hipotezė H_0 nesuderinama su imties duomenimis. Tai reiškia, kad naudojant parinktą statistinį kriterijų, kai $\gamma = 0{,}05$, vidutiniškai penkiais atvejais iš šimto klaidingai atmetama nulinė hipotezė. Tokiu atveju priimama alternatyva, priešingu atveju nulinei hipotezei neprieštaraujama.

1.7.2. Hipotezės apie vidurkių vektorių ir kovariacijų matricų sutapimą tikrinimas

Šiame skyrelyje aptariamas hipotezės apie vidurkių vektoriaus ir kovariacijų matricos sutapimą tikrinimas, kuris toliau bus pritaikomas dviejų Markovo grandžių skirtumui nustatyti. Sakykim, duota atsitiktinė imtis $X^1, X^2, \ldots, X^K$, kurios elementai yra pasiskirstę pagal *d*-matį normalųjį dėsnį su vidurkių vektoriumi μ ir kovariacijų matrica Ω. Jei kovariacijų matrica yra žinoma, priimama nulinė hipotezė H_0: $\mu = \mu_0$, kuri yra ekvivalenti kanoninei formai $H_0^*: \mu = 0$. Alternatyvi hipotezė H_1: $\mu \neq \mu_0$. Šiai hipotezei tikrinti yra taikomas Hotelingo kriterijus (Anderson, 1958), kuris yra pasiūlytas **stochastiniame programavime** statistinei hipotezei apie dviejų Monte-Karlo imčių vidurkių sutapimą tikrinti (Sakalauskas, 2000, 2002).

Jei imties vidurkių vektorius ir kovariacijų matrica yra nežinomi, priimama nulinė hipotezė

$$H_0 : \begin{cases} \mu = \mu_0, \\ \Omega = \Omega_0, \end{cases} \tag{1.23}$$

čia μ_0 ir Ω_0 yra žinomos reikšmės.

Nulinė hipotezė yra ekvivalenti kanoninei formai H_0^*:

$$\begin{cases} \mu = 0, \\ \Omega = E, \end{cases} \tag{1.24}$$

nes galima pakeisti $x_1, \ldots, x_N \to y_1, \ldots, y_N$ pagal $y_j = G(x_j - \mu_0)$, čia G yra tokia matrica, kad $G\Omega_0 G^T = E$. Alternatyvi hipotezė tvirtina, kad bent viena (1.24) lygybių yra neteisinga:

$$H_1 : \begin{cases} \mu \neq \mu_0, \\ \Omega \neq \Omega_0. \end{cases}$$

Taigi, (1.23) nulinė hipotezė yra ekvivalenti hipotezei, kurioje d-matės sekos kiekvieno nepriklausomo atsitiktinio vektoriaus vidurkis lygus nuliui, o kovariacijų matrica yra vienetinė (1.24).

Nulinei hipotezei patikrinti yra žinoma statistika (Anderson, 1958):

$$\begin{aligned} & -2\log\lambda = \\ & = N\log|\Omega_0| + N \cdot \mathrm{tr}(\Omega_0^{-1}) \cdot \left[S + (\bar{x} - \mu_0)(\bar{x} - \mu_0)^T\right] - N\log|S| - N \cdot d, \end{aligned} \tag{1.25}$$

čia $N \cdot S = \sum_{j=1}^{N} (x_j - \bar{x})(x_j - \bar{x})^T$, $N \cdot \bar{x} = \sum_{j=1}^{N} x_j$, kai statistika

$$\chi^2 = \frac{-2\log\lambda}{(1 - C_1)^2}, \ C_1 = \frac{2d^2 + 9d + 11}{6N(d+3)} \tag{1.26}$$

yra aproksimuojama $\chi^2_{p_1}$ skirstiniu su $p_1 = \frac{1}{2} d(d+3)$ laisvės laipsniais $F > F_{p_1, p_2}(\gamma)$ (Krishnaiah, 1984).

Tikslesnė yra F aproksimacija, kai statistika

$$F = \frac{-2\log\lambda}{b} \tag{1.27}$$

yra aproksimuojama Fišerio skirstiniu su p_1 ir p_2 laisvės laipsniais, čia

$$p_2 = \frac{p_1 + 2}{C_2 - C_1^2}, \; C_2 = \frac{d^3 + 6d^2 + 11d + 4}{6N^2(d+3)}, \; b = \frac{p_1}{1 - C_1 - \frac{p_1}{p_2}}.$$

(1.26) ir (1.27) aproksimacijose nulinė hipotezė yra atmetama, jeigu gauta reikšmė yra per didelė, t. y. $\chi^2 > \chi^2_{p_1}(\gamma)$, $F > F_{p_1,p_2}(\gamma)$ (Krishnaiah, 1984).

1.8. Skyriaus išvados

Toliau pateikiama MCMC metodo analizinės apžvalgos santrauka ir suformuluojamos prielaidos uždaviniams spręsti, sukuriant adaptuotą MCMC metodą.

Statistikoje MCMC metodas sudaro algoritmų klasę tikimybinių skirstinių imitavimui, konstruojant Markovo grandinę, leidžiančią gauti reikiamą skirstinį kaip grandinės pusiausvyros būsenos skirstinį. Teorinį Monte-Karlo metodo pagrindą sudaro DSD, CRT ir Beri-Eseno teorema. Nors suvidurkinami atsitiktiniai dydžiai gali būti pasiskirstę labai įvairiai, jų vidurkių arba su vidurkiu susijusių statistikų skirstiniai gali būti pakankamai gerai aproksimuojami vienmačiu arba daugiamačiu normaliuoju skirstiniu arba su juo susijusiais skirstiniais. Disertacijoje pasiūlytas metodas gauti išraiškų, į kurias Monte-Karlo būdu gauti įvertiniai įeina netiesiškai, asimptotiškai nepaslinktuosius įvertinius (1.15) bei pasikliautinuosius intervalus (1.21).

Skaičiuojamuoju požiūriu MCMC metodas leidžia spręsti lygtis, į kurias įeina sudėtingi daugialypiai integralai, konstruojant Monte-Karlo imčių Markovo grandinę. Šios lygtys dažnai gali būti išvedamos kaip būtinosios kokio nors kriterijaus optimalumo sąlygos. Disertacijoje priimama, kad šiuos kriterijus aprašančios tikslo funkcijos yra tolydžios ir glodžiai diferencijuojamos, o MCMC metodą galima interpretuoti kaip tokios tikslo funkcijos gradientinio optimizavimo metodą.

Naudojant MCMC metodą susiduriama su grandinės grandžių skaičiaus parinkimo ir Monte-Karlo imties tūrio atskirose grandyse reguliavimo problemomis. Sprendžiant grandžių skaičiaus parinkimo problemą disertacijoje pasiūlyta nutraukti grandinės generavimą, jei imtys, gautos gretimose grandyse, statistiškai nesiskiria, pritaikius tam statistinių hipotezių apie šių imčių skirtingumą tikrinimo metodus. Disertacijoje pasiūlyta reguliuoti Monte-Karlo imties tūrį pasinaudojus statistiniu kriterijumi apie dviejų Markovo grandžių Monte-Karlo imčių skirstinių sutapimą.

Kadangi skirstiniai dažnai aprašomi parametrų vektoriais bei matricomis, tai hipotezės apie grandžių sutapimą arba skirtumą gali būti tikrinamos pasinaudojus statistiniais testais apie vidurkių ir kovariacijų matricų sutapimą.

2 skyrius. DAUGIAMAČIO ASIMETRINIO *t* SKIRSTINIO PARAMETRŲ VERTINIMAS

2.1. Įvadas

Nors MCMC metodas yra taikomas labai plačiai, galima išskirti pagrindines šio metodo ypatybes. Toliau disertacijoje sudaryti algoritmai asimetrinio *t* skirstinio, Puasono-Gauso modelio ir daugiamačio α-stabiliojo dėsnio parametrams vertinti adaptuotu MCMC metodu. Statistikos uždaviniai, išspręsti adaptuotu MCMC metodu, leidžia atskleisti šias ypatybes, pritaikant disertacijoje sudarytą metodą kitiems statistikos uždaviniams spręsti adaptuotu MCMC metodu. Adaptuotas MCMC metodas pasižymi Monte-Karlo imties tūrio parinkimo taisykle bei modeliavimo paklaidų įvertinimu statistiniu būdu atskirose grandyse, atitinkamai reguliuojant Markovo grandžių Monte-Karlo imčių tūrį bei generuojamą grandžių skaičių.

MCMC metodas dažnai taikomas hierarchinių statistinių modelių tyrimui (Bradley ir Thomas, 2000). Šiame skyriuje nagrinėjamas adaptuoto MCMC metodo taikymas asimetrinio (daugiamačio) *t* skirstinio (angl. *a skew t distribution*) parametrams vertinti didžiausio tikėtinumo metodu. Asimetrinis *t* skirstinys apibrėžiamas hierarchiniu statistiniu modeliu per daugiamatį normalųjį skirstinį, kurio vidurkio vektorius yra taip pat pasiskirstęs pagal normalųjį skirstinį, kai abiejų šių skirstinių kovariacijų matricos priklauso nuo parametro pasiskirsčiusio pagal gama dėsnį (Azzalini, Genton, 2008). Paprastai tariant, tai yra normaliojo ir gama skirstinių mišinys. Pasinaudojant šiuo apibrėžimu asimetrinio *t* skirstinio tankį galima išreikšti daugiamačiu integralu. Kadangi tokiu atveju tikėtinumo funkcija irgi yra išreiškiama per daugiamačius integralus, disertacijoje sukurtas MCMC algoritmas didžiausio tikėtinumo įvertiniams gauti, parenkant Monte-Karlo imčių tūrį atskirose grandyse taip, kad sumažėtų bendras Monte-Karlo bandymų skaičius ir nutraukiant grandžių generavimą, kai imtys dviejose gretimose grandyse statistiškai nesiskiria. Gautų įvertinių pasikliautinieji intervalai nustatomi pasinaudojant Monte-Karlo įvertinių asimptotiniu normaliuoju aproksimavimu, aptartu pirmame skyriuje. Stabdant MCMC algoritmą (žr. 2.4 skyrelį) yra tikrinama statistinė hipotezė apie nykstamai mažą skirtumą tarp dviejų gretimų Markovo grandžių įverčių (žr. 1.7 skyrelį).

Daugiamatis asimetrinis t skirstinys yra dažniausiai taikomas analizuojant skirstinių, kurie išsiskiria asimetriniais pavidalais, klases (Azzalini, Genton, 2008). Kita asimetrinio skirstinio savybė yra galimybė modeliuoti įvykių asimptotines tikimybes, aprašomas laipsniniais tikimybiniais dėsniais, t. y. didelėmis atsitiktinėmis tikimybėmis, parenkant gama skirstinio parametrą, vadinamą formos parametru. Statistinėje literatūroje yra pavyzdžių, kur asimetrinis t skirstinys taikomas biologiniuose tyrimuose, aptinkant infekcinių ligų sukėlėjus, ar faktinių finansinių rinkų statistinių savybių prognozavimui bei kitose srityse (Azzalini, Capitanio, 2003; Azzalini, Genton, 2008; Cabral ir kt., 2008; Kim, Mallick, 2003; Panagiotelis, Smith, 2008; Wang, Chen, 2009). Šio skyriaus rezultatai buvo publikuoti moksliniuose darbuose: Sakalauskas ir kt. (2013, 2014); Sakalauskas ir Vaiciulyte (2010a, 2010b, 2011, 2012a, 2012b); Vaiciulyte (2012); Vaičiulytė (2014); Vaičiulytė ir Sakalauskas (2011a, 2011b).

2.2. Daugiamačio asimetrinio *t* skirstinio apibrėžimas

Tegu $X = (X_1, X_2, ..., X_d)$ – daugiamatis atsitiktinis vektorius, pasiskirstęs pagal normalųjį skirstinį $X \sim N(z, \Omega)$ su tankiu

$$f(x|z,t,\Omega) = (t/\pi)^{\frac{d}{2}} \cdot |\Omega|^{-\frac{1}{2}} \cdot e^{-t\cdot(x-z)^T\cdot\Omega^{-1}\cdot(x-z)},$$

čia vidurkio vektorius $z = (z_1, z_2, ..., z_d)$ savo ruožtu pasiskirstęs pagal daugiamatį normalųjį skirstinį $z \sim N\left(\mu, \frac{\Theta}{2t}\right)$ puserdvėje $W = \{\eta \cdot (z-\mu) \geq 0, \eta \subset \Re^d\}$, Ω, Θ – teigiamai apibrėžtos kovariacijų matricos, o atsitiktinis dydis t pasiskirstęs pagal gama dėsnį su parametru α (Azzalini, Capitanio, 2003), t. y. su tankiu

$$f_1(t|\alpha) = \frac{t^{\frac{\alpha}{2}-1}}{\Gamma(\alpha/2)} \cdot e^{-t}.$$

Taigi, atsitiktinio asimetrinio t skirstinio tankio funkciją galima išreikšti pavidalu:

$$\begin{aligned} p(x|\mu,\Omega,\Theta,\alpha,\eta) &= 2 \cdot \int_0^{\infty} \int_{\eta\cdot(z-\mu)\geq 0} f(x|z,t,\Omega) \cdot f(z|\mu,t,\Theta) \cdot f_1(t|\alpha) dzdt = \\ &= \int_0^{\infty} \int_{\eta\cdot(z-\mu)\geq 0} \frac{2}{\pi^d \cdot |\Omega|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot \Gamma\left(\frac{\alpha}{2}\right)} \cdot t^{\frac{\alpha}{2}+d-1} \times \\ &\times e^{-t\cdot\left((x-z)^T \cdot \Omega^{-1} \cdot (x-z) + (z-\mu)^T \cdot \Theta^{-1} \cdot (z-\mu) + 1\right)} dzdt. \end{aligned} \tag{2.1}$$

Daugiamatį asimetrinį t skirstinį $ST(\mu, \Omega, \Theta, \alpha, \eta)$ apibūdinantys parametrai yra

- μ – vidurkio vektorius,
- Ω, Θ – kovariacijų matricos,
- α – formos parametras,
- η – integravimo sritį W nusakantis parametras.

Šį modelį galima interpretuoti statistiškai tokiu būdu. Pavyzdžiui, tegu sritis W aprašo arba biologinio užkrato šaltinį, arba taršos šaltinį, arba investuotojų lūkesčių sritį, ir pan., su kuriais yra susijusi jų tolesnė atsitiktinė sklaida, pasiskirsčiusi pagal normalųjį dėsnį arba kokį nors kitą elipsinį skirstinį (Sakalauskas ir kt., 2013, 2014; Vaičiulytė, 2014). Paprastumo dėlei disertacijoje nagrinėjama, kad ši sritis aprašoma tik vienu tiesiniu ribojimu, tačiau, bendru atveju, galima įvesti sritį, apribotą keliais tiesiniais ribojimais.

2.3. Didžiausio tikėtinumo funkcijos įvertiniai

Tegu $X = \left(X^1, X^2, ..., X^K\right)$ yra stebinių matrica, čia X^i, $i = \overline{1, K}$, yra nepriklausomi atsitiktiniai vektoriai, pasiskirstę pagal daugiamatį asimetrinį *t* skirstinį

$$X \sim ST(\mu, \Omega, \Theta, \alpha, \eta),$$

kurio parametrai vertinami didžiausio tikėtinumo metodu (1.5.1 skyrelis). Pasinaudojant (2.1) logaritminė tikėtinumo funkcija užrašoma taip:

$$L(\mu, \Omega, \Theta, \alpha, \eta) = -\sum_{i=1}^{K} \ln\left(p\left(X^i \middle| \mu, \Omega, \Theta, \alpha, \eta\right)\right) = -\sum_{i=1}^{K} \ln\left(E\left(f\left(X^i \middle| Z, G, \Omega\right)\right)\right), \quad (2.2)$$

čia vidurkiai skaičiuojami pagal atsitiktinius dydžius *Z* ir *G* (žr. (2.17)), kurių bendras tankis:

$$f\left(z, t \middle| \mu, \Omega, \alpha\right) = \begin{cases} 2 \cdot f\left(z \middle| \mu, t, \Theta\right) \cdot f_1\left(t \middle| \alpha\right), & \text{jei } \eta \cdot (z - \mu) \geq 0, \\ 0, & \text{jei } \eta \cdot (z - \mu) < 0. \end{cases}$$

čia $\eta \subset \Re^d$.

Skirstinio parametrų įvertiniai turi minimizuoti logaritminę tikėtinumo funkciją:

$$L(\mu, \Omega, \Theta, \alpha, \eta) \to \min_{\mu, \Omega, \Theta, \alpha, \eta}. \quad (2.3)$$

Daugiamačio asimetrinio *t* skirstinio (2.1) parametrų tikėtinumo įvertiniai $\hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}$ gaunami prilyginus nuliui atitinkamas tikėtinumo funkcijos išvestines ir išsprendus gautą lygčių sistemą (žr. 1.7.1 skyrelį), t. y.

$$
\begin{cases}
\dfrac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \mu} = -\sum_{i=1}^{K} \dfrac{\partial p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \mu} \cdot \dfrac{1}{p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)} = 0, \\
\dfrac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \Omega} = -\sum_{i=1}^{K} \dfrac{\partial p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \Omega} \cdot \dfrac{1}{p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)} = 0, \\
\dfrac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \Theta} = -\sum_{i=1}^{K} \dfrac{\partial p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \Theta} \cdot \dfrac{1}{p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)} = 0, \\
\dfrac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \alpha} = -\sum_{i=1}^{K} \dfrac{\partial p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \alpha} \cdot \dfrac{1}{p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)} = 0, \\
\dfrac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \eta} = -\sum_{i=1}^{K} \dfrac{\partial p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \eta} \cdot \dfrac{1}{p\left(X^i \middle| \mu,\Omega,\Theta,\alpha,\eta\right)} = 0.
\end{cases}
\tag{2.4}
$$

Atsižvelgiant į (2.1) tankio funkcijos pavidalą, (2.4) sistemos pirmos trys lygtys užrašomos taip (2 priedas; Sakalauskas ir Vaiciulyte, 2010; Vaičiulytė ir Sakalauskas, 2011a):

$$
\frac{\partial p\left(x \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \mu} = 4 \cdot \int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} t \cdot \Omega^{-1} \cdot (x-z) \cdot f\left(x \middle| z,t,\Omega\right) \cdot f\left(z \middle| \mu,t,\Theta\right) \cdot f_1\left(t \middle| \alpha\right) dz dt,
$$

$$
\begin{aligned}
\frac{\partial p\left(x \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \Omega} = 2 \cdot \int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} & \left(-\frac{1}{2}\Omega^{-1} + t \cdot \Omega^{-1} \cdot (x-z) \cdot (x-z)^T \cdot \Omega^{-1} \right) \times \\
& \times f\left(x \middle| z,t,\Omega\right) \cdot f\left(z \middle| \mu,t,\Theta\right) \cdot f_1\left(t \middle| \alpha\right) dz dt,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial p\left(x \middle| \mu,\Omega,\Theta,\alpha,\eta\right)}{\partial \Theta} = 2 \cdot \int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} & \left(-\frac{1}{2}\Theta^{-1} + t \cdot \Theta^{-1} \cdot (z-\mu) \cdot (z-\mu)^T \cdot \Theta^{-1} \right) \times \\
& \times f\left(x \middle| z,t,\Omega\right) \cdot f\left(z \middle| \mu,t,\Theta\right) \cdot f_1\left(t \middle| \alpha\right) dz dt,
\end{aligned}
$$

Kadangi diferencijuoti išraiškas su gama funkcija nėra patogu, tai pasinaudojant Oilerio formule, (2.1) asimetrinio t skirstinio tankį galima išreikšti integralu (1 priedas):

$$
p\left(x \middle| \mu,\Omega,\Theta,\alpha,\eta\right) =
$$

$$
= \int_{\eta\cdot(z-\mu)\geq 0} \frac{2 \cdot \prod_{i=0}^{d-1} \left(\frac{\alpha}{2} + i \right)}{\pi^d \cdot |\Omega|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot \left[(x-z)^T \cdot \Omega^{-1} \cdot (x-z) + (z-\mu)^T \cdot \Theta^{-1} \cdot (z-\mu) + 1 \right]^{\frac{\alpha}{2}+d}} dz. \tag{2.5}
$$

Trumpumo dėlei vieną iš dauginamųjų vardiklyje pažymėjus

$$A = (x - z)^T \cdot \Omega^{-1} \cdot (x - z) + (z - \mu)^T \cdot \Theta^{-1} \cdot (z - \mu) + 1\,,$$

ir išdiferencijavus (2.5) išraišką formos parametro α atžvilgiu, gaunama (Sakalauskas ir Vaiciulyte, 2011):

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial \alpha} = \int\limits_{\eta\cdot(z-\mu)\geq 0} \frac{2\cdot\prod\limits_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\cdot\left(-\frac{1}{2}\ln(A)+\sum\limits_{i=0}^{d-1}\frac{1}{\alpha+2\cdot i}\right)}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot A^{\frac{\alpha}{2}+d}}dz\ .$$

Paprastumo dėlei tariama, kad integravimo srityje esančio parametro $\eta_d = 1$. Tuomet pasinaudojant integravimo kintamojo keitiniu $z'_d = z_d - \sum\limits_{r=1}^{d-1}\eta_r\cdot(z_r - \mu_r)$, (2.5) integralas suvedamas į pavidalą, kuriame parametrų η_r integravimo srityje jau nebėra. Išdiferencijavus gautą išraišką pagal parametrus η_r, gaunama:

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial \eta} = \int\limits_{\eta\cdot(z-\mu)\geq 0} \frac{\prod\limits_{i=0}^{d}\left(\frac{\alpha}{2}+i\right)\cdot\Lambda}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot A^{\frac{\alpha}{2}+d+1}}dz\ ,$$

čia $\Lambda_r = 2\cdot(z-\mu)_r\cdot\left(\Omega^{-1}\cdot(x-z)-\Theta^{-1}\cdot(z-\mu)\right)_d$, $r = 1, 2, \ldots, d-1$, $\Lambda_d = 0$.

Vėl pasinaudojant Oilerio formule, gautos skirstinio formos ir srities parametrų išvestinės, sugrįžtant prie pradinio tankio pavidalo, užrašomos šitaip (Vaičiulytė ir Sakalauskas, 2011b; Sakalauskas ir Vaiciulyte, 2012b):

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial \alpha}=$$

$$=\int_0^{\infty}\int_{\eta\cdot(z-\mu)\geq 0}\left(-\ln(A)+\sum_{i=0}^{d-1}\frac{1}{\frac{\alpha}{2}+i}\right)\cdot f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)dzdt, \tag{2.6}$$

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial \eta}=\int_0^{\infty}\int_{\eta\cdot(z-\mu)\geq 0} d\cdot t\cdot\Lambda\cdot f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)dzdt.$$

Trumpumo dėlei įvedus sąlyginį tankį

$$\bar{f}(z,t|x,\mu,\Omega,\Theta,\alpha,\eta)=\frac{2\cdot f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)}{p(x|\mu,\Omega,\Theta,\alpha,\eta)}$$

tikėtinumo funkcijos išvestinės užrašomos taip:

$$\frac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \mu}=-2\cdot\sum_{i=1}^{K}E\left(t\cdot\Omega^{-1}\cdot(X^i-z)\middle|X^i,\mu,\Omega,\Theta,\alpha,\eta\right), \tag{2.7}$$

$$\frac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \Omega}=$$

$$=-\sum_{i=1}^{K}E\left(-\frac{1}{2}\Omega^{-1}+t\cdot\Omega^{-1}\cdot(X^i-z)\cdot(X^i-z)^T\cdot\Omega^{-1}\middle|X^i,\mu,\Omega,\Theta,\alpha,\eta\right), \tag{2.8}$$

$$\frac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \Theta}=$$

$$=-\sum_{i=1}^{K}E\left(-\frac{1}{2}\Theta^{-1}+t\cdot\Theta^{-1}\cdot(z-\mu)\cdot(z-\mu)^T\cdot\Theta^{-1}\middle|X^i,\mu,\Omega,\Theta,\alpha,\eta\right), \tag{2.9}$$

$$\frac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \alpha}=-\sum_{i=1}^{K}E\left(-\frac{1}{2}\ln(A)+\sum_{i=0}^{d-1}\frac{1}{\alpha+2\cdot i}\middle|X^i,\mu,\Omega,\Theta,\alpha,\eta\right), \tag{2.10}$$

$$\frac{\partial L(\mu,\Omega,\Theta,\alpha,\eta)}{\partial \eta}=-\sum_{i=1}^{K}E\left(t\cdot\Lambda\middle|X^i,\mu,\Omega,\Theta,\alpha,\eta\right). \tag{2.11}$$

Pagal fiksuoto taško metodą (angl. *fixed-point iteration)* parametrų įverčiai gaunami iš lygčių (Sakalauskas ir Vaiciulyte, 2012b):

$$\frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot\left(X^i - z\right)\middle| X^i, \hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}\right) = 0\,, \tag{2.12}$$

$$\hat{\Omega} = \frac{2}{K}\sum_{i=1}^{K} E\left(t\cdot\left(X^i - z\right)\cdot\left(X^i - z\right)^T \middle| X^i, \hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}\right), \tag{2.13}$$

$$\hat{\Theta} = \frac{2}{K}\sum_{i=1}^{K} E\left(t\cdot\left(z - \hat{\mu}\right)\cdot\left(z - \hat{\mu}\right)^T \middle| X^i, \hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}\right), \tag{2.14}$$

$$\hat{\alpha} = \frac{K\cdot\sum_{i=0}^{d-1}\dfrac{1}{\dfrac{2\cdot i}{\hat{\alpha}}+1}}{\sum_{i=1}^{K} E\left(\dfrac{1}{2}\ln\left(\hat{A}\right)\middle| X^i, \hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}\right)}\,, \tag{2.15}$$

$$\hat{\eta} = \frac{1}{K}\sum_{i=1}^{K} E\left(t\cdot\hat{\Lambda}\middle| X^i, \hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}\right), \tag{2.16}$$

čia

$$\hat{A} = \left(X^i - z\right)^T\cdot\hat{\Omega}^{-1}\cdot\left(X^i - z\right) + \left(z - \hat{\mu}\right)^T\cdot\hat{\Theta}^{-1}\cdot\left(z - \hat{\mu}\right) + 1\,,$$
$$\hat{\Lambda}_r = \left(z - \hat{\mu}\right)_r\cdot\left(\hat{\Omega}^{-1}\cdot\left(X^i - z\right) - \hat{\Theta}^{-1}\cdot\left(z - \hat{\mu}\right)\right)_d, \quad r = 1, 2, \ldots, d-1,\ \hat{\Lambda}_d = 0\,.$$

Pasinaudojant (2.7)–(2.11) išvestinėmis gaunamas gradientinio nusileidimo metodas tikėtinumo funkcijos optimizavimui:

$$\begin{aligned}
\mu^{k+1} &= \mu^k - \zeta_\mu\cdot\frac{\partial L\left(\mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)}{\partial\mu},\\
\Omega^{k+1} &= \Omega^k - 2\cdot\Omega^k\cdot\frac{\partial L\left(\mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)}{\partial\Omega}\cdot\Omega^k,\\
\Theta^{k+1} &= \Theta^k - 2\cdot\Theta^k\cdot\frac{\partial L\left(\mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)}{\partial\Theta}\cdot\Theta^k,
\end{aligned} \tag{2.17}$$

$$\alpha^{k+1} = \alpha^k - \zeta_\alpha \cdot \frac{\partial L\left(\mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)}{\partial \alpha},$$

$$\eta^{k+1} = \eta^k - \zeta_\eta \cdot \frac{\partial L\left(\mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)}{\partial \eta},$$

čia

$$\zeta_\mu = \frac{1}{\sum_{i=1}^{K} E\left(t \middle| X^i, \mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)},$$

$$\zeta_\alpha = \frac{\alpha^k}{\sum_{i=1}^{K} E\left(\frac{\ln\left(A^k\right)}{2} \middle| X^i, \mu^k, \Omega^k, \Theta^k, \alpha^k, \eta^k\right)},$$

$$\zeta_\eta = 1,$$

$$A^k = \left(X^i - z\right)^T \cdot \left(\Omega^k\right)^{-1} \cdot \left(X^i - z\right) + \left(z - \mu^k\right)^T \cdot \left(\Theta^k\right)^{-1} \cdot \left(z - \mu^k\right) + 1,$$

iš kurio išplaukia EM algoritmas, kurio realizavimas MCMC metodu pateiktas 2.4 skyrelyje.

2.4. Markovo grandinės Monte-Karlo algoritmas

Parametrų įvertinius galima rasti iteraciniu būdu, pritaikius EM algoritmą (žr. 1.5 skyrelį), kai parinktos tam tikros pradinės reikšmės, ir apskaičiuojant Monte-Karlo metodu integralus, įeinančius į (2.12)–(2.16) lygtis.

Tegul yra duotas tam tikras vertinamų parametrų pradinis artinys $\mu^0, \Omega^0, \Theta^0, \alpha^0, \eta^0$, kuriame yra sugeneruojama pradinė Markovo Monte-Karlo grandis. Dabar tarkime, kad yra sugeneruota k Markovo Monte-Karlo grandžių. Tegu k-tąją imtį sudaro N^k nepriklausomų atsitiktinių dydžių, pasiskirsčiusių pagal gamą ir normalųjį dėsnius:

$$G_j \sim \mathrm{Gama}\left(\frac{\alpha^k}{2}\right), \qquad \varphi_j \sim \mathrm{N}\left(0, \Theta^k\right), \tag{2.18}$$

$$Z_j = \begin{cases} \mu^k + \varphi_j, \text{ jeigu } \eta \cdot \varphi_j \geq 0, \\ \mu^k - \varphi_j, \text{ jeigu } \eta \cdot \varphi_j < 0, \end{cases}$$

čia $j = 1, 2, \ldots, N^k$, $k = 0, 1, 2, \ldots$.

(2.12)–(2.16) lygtims spręsti pritaikomas Monte-Karlo metodas, pasinaudojus EM algoritmu. Tuomet iš (2.17) išplaukia formulės (Vaičiulytė ir Sakalauskas, 2011b)

$$\mu^{k+1} = \mu^k + \frac{1}{K \cdot \zeta^k} \sum_{i=1}^{K} \frac{M_i^k}{P_i^k}, \tag{2.19}$$

$$\Omega^{k+1} = \frac{2}{K} \sum_{i=1}^{K} \frac{S_i^k}{P_i^k}, \tag{2.20}$$

$$\Theta^{k+1} = \frac{2}{K} \sum_{i=1}^{K} \frac{T_i^k}{P_i^k}, \tag{2.21}$$

$$\alpha^{k+1} = \frac{1}{h^{k+1}} \sum_{i=0}^{d-1} \frac{1}{\frac{2 \cdot i}{\alpha^k} + 1}, \tag{2.22}$$

$$\eta^{k+1} = \eta^k - \frac{1}{K} \sum_{i=1}^{K} \frac{Q_i^k}{P_i^k}, \tag{2.23}$$

čia pasinaudojama Monte-Karlo įverčiais, apskaičiuotais k-toje Markovo grandyje, pasinaudojant (2.18) atsitiktine imtimi:

$$P_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \tag{2.24}$$

$$P1_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(f\left(X^i \middle| Z_j, G_j, \Omega^k\right)\right)^2, \tag{2.25}$$

$$M_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(X^i - Z_j\right) \cdot G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \tag{2.26}$$

$$S_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(X^i - Z_j\right)\cdot\left(X^i - Z_j\right)^T \cdot G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \qquad (2.27)$$

$$T_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(Z_j - \mu^k\right)\cdot\left(Z_j - \mu^k\right)^T \cdot G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \qquad (2.28)$$

$$B_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\frac{1}{2}\ln\left(A^k\right)\cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \qquad (2.29)$$

$$B1_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(\frac{1}{2}\ln\left(A^k\right)\cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right)\right)^2, \qquad (2.30)$$

$$Q_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k} d\cdot\Lambda^k \cdot G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \qquad (2.31)$$

$$Q1_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(d\cdot\Lambda^k \cdot G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right)\right)^2, \qquad (2.32)$$

$$Dz_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k} G_j \cdot f\left(X^i \middle| Z_j, G_j, \Omega^k\right), \qquad (2.33)$$

čia

$$A^k = \left(X^i - Z_j\right)^T \cdot \left(\Omega^k\right)^{-1}\cdot\left(X^i - Z_j\right) + \left(Z_j - \mu^k\right)^T \cdot \left(\Theta^k\right)^{-1}\cdot\left(Z_j - \mu^k\right) + 1,$$
$$\Lambda^k = \left(\Lambda_1^k, \Lambda_2^k, \ldots, \Lambda_d^k\right),$$
$$\Lambda_r^k = (Z_r - \mu^k)_r \cdot\left(\left(\Omega^k\right)^{-1}\cdot\left(X^i - Z_r\right) - \left(\Theta^k\right)^{-1}\cdot\left(Z_r - \mu^k\right)\right)_d, \quad r = 1, 2, \ldots, d-1,$$
$$\Lambda_d^k = 0,\ 1 \le i, k \le K.$$

Pažymima

$$h^k = \frac{1}{K}\sum_{i=1}^{K}\frac{B_i^k}{P_i^k}, \qquad q^k = \frac{1}{K}\sum_{i=1}^{K}\frac{Q_i^k}{P_i^k}, \qquad g^k = \frac{1}{K}\sum_{i=1}^{K}\frac{P1_i^k}{\left(P_i^k\right)^2},$$
$$b^k = \frac{N^k}{K}\sum_{i=1}^{K}\frac{B1_i^k}{\left(P_i^k\right)^2}, \quad n^k = \frac{N^k}{K}\sum_{i=1}^{K}\frac{Q1_i^k}{\left(P_i^k\right)^2}, \quad \zeta^k = \frac{1}{K}\sum_{i=1}^{K}\frac{Dz_i^k}{P_i^k}.$$

Pasinaudojant (2.2) ir (2.24) formulėmis gaunamas logaritminės tikėtinumo funkcijos įvertinys:

$$L^k = -\sum_{i=1}^{K} \ln\left(P_i^k\right). \tag{2.34}$$

(2.34) yra pagrįstas tikėtinumo funkcijos (2.2) įvertinys (žr. (1.19)). (2.34) įvertinio 95% pasikliautinasis intervalas yra (žr. (1.21)):

$$\left[L^k - \frac{2}{\sqrt{N^k}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P1_i^k}{P_i^{k^2}} - 1 \right)}, L^k + \frac{2}{\sqrt{N^k}} \cdot \sqrt{\sum_{i=1}^{K} \left(\frac{P1_i^k}{P_i^{k^2}} - 1 \right)} \right]. \tag{2.35}$$

Taigi, Monte-Karlo Markovo grandinė generuojama pagal (2.19)–(2.23) formules tol, kol (2.35) pasikliautinojo intervalo ilgis tampa mažesnis už pasirinktą reikšmę ε, $\varepsilon > 0$, o statistinė hipotezė apie dviejų gretimų grandžių parametrų įverčių nykstamai mažą skirtumą yra neatmetama:

$$\mu^{k+1} = \mu^k, \quad \Omega^{k+1} = \Omega^k, \quad \Theta^{k+1} = \Theta^k, \quad \alpha^{k+1} = \alpha^k, \quad \eta^{k+1} = \eta^k. \tag{2.36}$$

Alternatyvi hipotezė tvirtina, kad bent viena iš (2.36) lygybių yra neteisinga.

Vietoj hipotezės $\alpha^{k+1} = \alpha^k$ yra paprasčiau atitinkamai patikrinti hipotezę $h^{k+1} = h^k$, patikrinant hipotezę apie imties vidurkio lygybę nuliui (žr. 1.7.2 skyrelį). Hipotezė $\eta^{k+1} = \eta^k$ yra suvedama į hipotezės apie $(d-1)$-mačio atsitiktinio vektoriaus lygybę nuliui.

Taigi, (2.36) statistinė hipotezė atmetama, jeigu

$$\begin{aligned} H^k = \frac{K}{g^k} \cdot \Bigg[& -\ln\left(\frac{\left|\Omega^{k+1}\right|}{\left|\Omega^k\right|}\right) - \ln\left(\frac{\left|\Theta^{k+1}\right|}{\left|\Theta^k\right|}\right) + \left(\mu^{k+1} - \mu^k\right)^T \cdot \left(\Omega^k\right)^{-1} \cdot \left(\mu^{k+1} - \mu^k\right) + \mathrm{tr}\left(\Omega^{k+1} \cdot \left(\Omega^k\right)^{-1}\right) + \\ & + \mathrm{tr}\left(\Theta^{k+1} \cdot \left(\Theta^k\right)^{-1}\right) - 2 \cdot d \Bigg] + N \cdot K \cdot \frac{\left(h^{k+1} - h^k\right)^2}{b^k - \left(h^{k+1}\right)^2} + N \cdot K \cdot q^k \cdot \left(n^k\right)^{-1} \cdot \left(q^k\right)^T > \Psi_{\delta,p} \end{aligned} \tag{2.37}$$

čia $\Psi_{\delta,p}$ yra χ_p^2 skirstinio kvantilis, $p = d(d+3)$ laisvės laipsniais, δ yra reikšmingumo lygmuo. Priešingu atveju (2.36) hipotezė neatmetama. Sudarant šį kriterijų atsižvelgta į tai, kad dar yra tikrinama hipotezė apie vieno vidurkio vektoriaus ir dviejų kovariacijų matricų lygybę žinomam vektoriui ir kitoms dviem kovariacijų matricoms. Jei ši hipotezė neatmetama, tai Markovo grandžių generavimą galima nutraukti ir priimti paskutinėje iteracijoje gautus įverčius.

Generuojant pirmąsias Markovo grandis nėra svarbu apskaičiuoti Monte-Karlo įvertinius labai tiksliai, nes pakankamą tikslumą reikia pasiekti tik paskutinėse grandyse. Todėl pasinaudojant tuo, jog (2.19)–(2.23) Monte-Karlo įvertinių išraiškose pasiskirstymą galima asimptotiškai aproksimuoti daugiamačiu normaliuoju dėsniu, Monte-Karlo imčių tūriui reguliuoti yra įvedama taisyklė, analogiška taisyklei, taikomai stochastiniame programavime (Sakalauskas, 2000):

$$N^{k+1} \geq \Psi_{\nu,p} \cdot \frac{N^k}{H^k}, \qquad (2.38)$$

čia ν yra reikšmingumo lygmuo. Naudojantis martingaliniais metodais (Polyak, 1987) galima parodyti, kad atitinkamai parinkus ν galima užtikrinti (2.19)–(2.23) ir (2.38) algoritmų konvergavimą į (2.3) uždavinio sprendimą (Sakalauskas, 2000). Siekiant išvengti labai mažo arba labai didelio Monte-Karlo imties tūrio, apskaičiuojamo pagal (2.38) taisyklę, jis yra ribojamas iš apačios reikšme $N_{\min}$ ir iš viršaus reikšme $N_{\max}$. Siekiant sumažinti laiką, sunaudojamą Monte-Karlo imčių generavimui, imties generavimą galima nutraukti, kai (2.35) pasikliautinojo intervalo ilgis tampa mažesnis už pasirinktą reikšmę, dar nepasiekus (2.38) reikšmės.

Nagrinėjamo uždavinio sprendimo algoritmas pateiktas 2.1 lentelėje.

2.1 lentelė. Asimetrinio *t* skirstinio MCMC metodo tyrimo statistinio modeliavimo algoritmas

Asimetrinio *t* skirstinio parametrų vertinimo algoritmas
Tikslas: asimetrinio *t* skirstinio parametrų įvertinimas MCMC metodu. *Įėjimo parametrai*: vidurkio vektorius μ^0, kovariacijų matricos Ω^0 ir Θ^0, formos parametras α^0, integravimo sritį nusakantis parametras η^0, stebinių skaičius *K*, pradinis minimalus ir maksimalus Markovo grandies Monte-Karlo imties tūriai N^0, $N_{\min}$, $N_{\max}$, maksimalus grandžių skaičius *k*. *Išėjimo parametrai*: logaritminės tikėtinumo funkcijos *L*, skirstinio parametrų $\mu, \Omega, \Theta, \alpha, \eta$ Monte-Karlo įverčių sekos, kiekvienos grandies pasikliautinojo intervalo ilgis *cf*, stabdymo statistika *H* ir imties tūris *N*. *1 žingsnis.* Priskiriamos pradinės parametrų reikšmės *2 žingsnis.* Apibrėžiama skirstinio tankio funkcija per vienlypį integralą (2.5) *3 žingsnis.* Sugeneruojami *K* atsitiktiniai dydžiai, pasiskirstę pagal gama dėsnį su parametru α *4 žingsnis.* Sugeneruojamas atsitiktinis vidurkio vektorius *z*, pasiskirstęs pagal daugiamatį normalųjį skirstinį *5 žingsnis.* Sugeneruojamas atsitiktinis stebinių vektorius *X*, pasiskirstęs pagal normalųjį skirstinį, kurio vidurkis yra *z* *6 žingsnis.* Sprendžiamas minimizavimo uždavinys (2.3) *7 žingsnis.* Gaunami didžiausio tikėtinumo parametrų įverčiai $\hat{\mu}, \hat{\Omega}, \hat{\Theta}, \hat{\alpha}, \hat{\eta}$ (2.12)–(2.16) ir apskaičiuojama minimali tikėtinumo funkcijos reikšmė $\hat{L}$ *8 žingsnis.* *for* iteracijų skaičius *iter* kinta nuo 1 iki *k* *9 žingsnis.* Ciklo kintamajam *n* priskiriamas 1 *10 žingsnis.* Pasikliautinojo intervalo ilgiui *cf* priskiriamas 1 *11 žingsnis.* *while* $n \le N_{\min}$ (arba $n \le N_{\min}$ ir $cf > \varepsilon$) *12 žingsnis.* *for i* kinta nuo 1 iki *K*

Asimetrinio t skirstinio parametrų vertinimo algoritmas
13 žingsnis. Sugeneruojami atsitiktiniai vektoriai pasiskirstę pagal normalųjį ir gamą dėsnius (2.18)
14 žingsnis. Sumuojami integralų įvertiniai (2.24)–(2.33)
15 žingsnis. Nustatomos tikėtinumo funkcijos pasikliautinojo intervalo (2.35) ilgis
16 žingsnis. Kintamajam n priskiriama n+1 reikšmė ir ciklas *while* vėl kartojamas (*11 žingsnis*)
17 žingsnis. Gaunami Monte-Karlo įverčiai (2.19)–(2.23)
18 žingsnis. Tikrinama stabdymo statistika (2.37)
19 žingsnis. Įvedamas imties tūrio reguliavimas (2.38)
20 žingsnis. Pradiniam artiniui priskiriami gauti MCMC įverčiai
21 žingsnis. Sudaromas masyvas duomenims išvesti ir toliau tęsiamas iteracijų ciklas (*8 žingsnis*)

2.5. Kompiuterinis modeliavimas

2.5.1. Skaitinis eksperimentas

Sudarytam algoritmui tirti atliktas eksperimentas. Pradinius duomenis sudarė $K = 100$ dvimačių atsitiktinių vektorių, sugeneruotų pagal asimetrinį t skirstinį $ST(\mu, \Omega, \Theta, \alpha, \eta)$ (žr. 3 priedą) su šiais parametrais:

$$d = 2, \quad \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1{,}61 & 0{,}27 \\ 0{,}27 & 2{,}9 \end{pmatrix}, \quad \Theta = \begin{pmatrix} 3{,}67 & 0{,}86 \\ 0{,}86 & 2{,}55 \end{pmatrix}, \quad \alpha = 5{,}2, \quad \eta = 0{,}451.$$

Palyginimui su programos MathCad paprograme *Minimize*() buvo apskaičiuoti didžiausio tikėtinumo įverčiai:

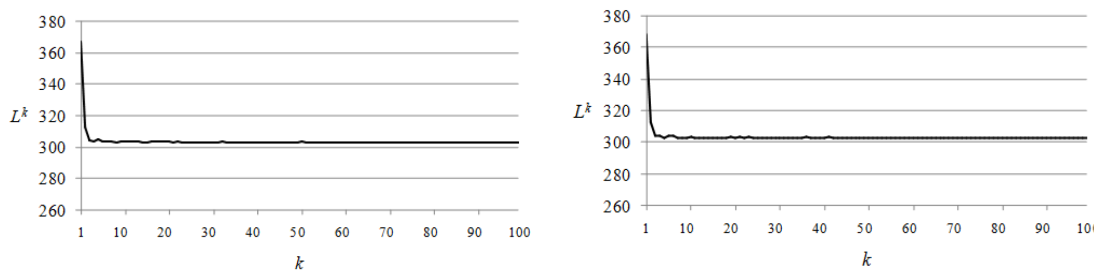
$$\hat{\mu} = \begin{pmatrix} 0{,}98 \\ 2{,}26 \end{pmatrix}, \quad \hat{\Omega} = \begin{pmatrix} 0{,}96 & 0{,}23 \\ 0{,}23 & 2{,}78 \end{pmatrix}, \quad \hat{\Theta} = \begin{pmatrix} 2{,}68 & 0{,}67 \\ 0{,}67 & 1{,}05 \end{pmatrix}, \quad \hat{\alpha} = 3{,}984, \quad \hat{\eta} = 0{,}596. \tag{2.39}$$

Sugeneruota $k = 100$ Markovo grandžių pagal (2.19)–(2.23), (2.38) išraiškas paėmus tokias pradines reikšmes:

$$\mu^0 = 1{,}5 \cdot \hat{\mu}, \quad \Omega^0 = 1{,}5 \cdot \hat{\Omega}, \quad \Theta^0 = 1{,}5 \cdot \hat{\Theta}, \quad \alpha^0 = \hat{\alpha}, \quad \eta^0 = \hat{\eta}.$$

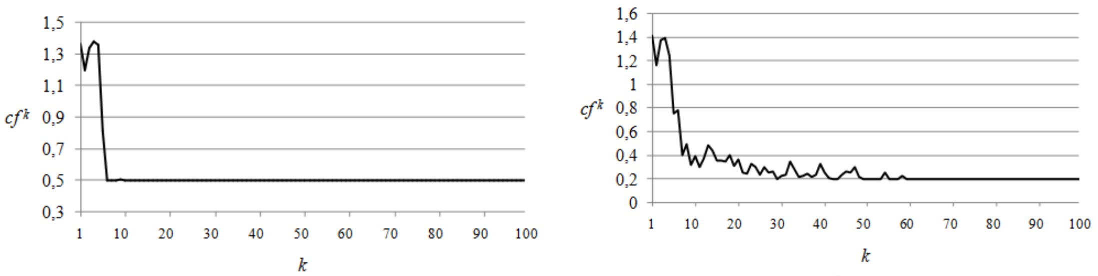
Siekiant išvengti mažų imties tūrio reikšmių, jis buvo ribojamas: $N^k \geq 500$. Imčių generavimas buvo nutraukiamas, nepasiekus imties tūrio reikšmės, apskaičiuotos pagal (2.38) taisyklę, jei (2.35) pasikliautinojo intervalo ilgis tapdavo mažesnis už reikšmę $\varepsilon = 0{,}5$ arba $\varepsilon = 0{,}2$.

2.1–2.4 paveiksluose pavaizduotos asimetrinio t skirstinio vertinimo nagrinėjamu MCMC algoritmu charakteristikų priklausomybės nuo sugeneruotų Markovo grandžių skaičiaus.



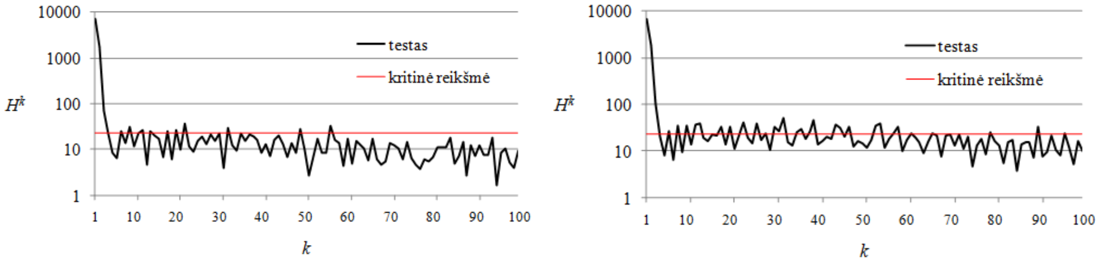
(*a*) (*b*)

2.1 pav. Tikėtinumo funkcija L^k



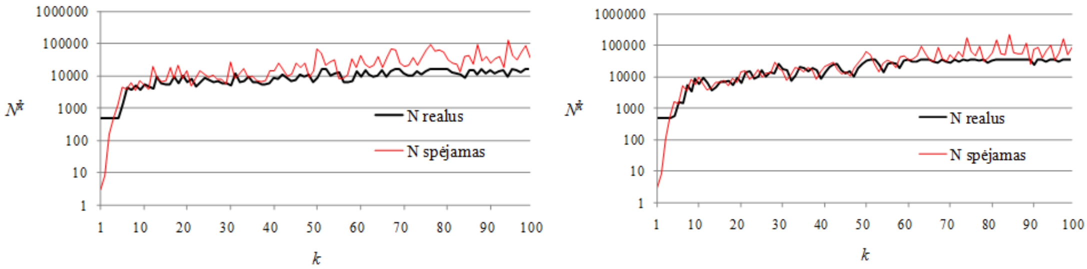
(*a*) (*b*)

2.2 pav. Pasikliautinojo intervalo ilgis cf^k



(*a*) (*b*)

2.3 pav. Stabdymo statistika H^k



(*a*) (*b*)

2.4 pav. Imties tūris N^k

Grafikuose (*a*) pavaizduotos priklausomybės, kai reikiamas pasikliautinojo intervalo ilgis $\varepsilon = 0{,}5$, o grafikuose (*b*) – kai pasikliautinojo intervalo ilgis $\varepsilon = 0{,}2$. 2.1 pav. pastebima, kad tikėtinumo funkcija mažėja tol, kol pasiekiama didžiausio tikėtinumo uždavinio (2.3) sprendinių zona. Tam reikia sugeneruoti maždaug 5–10 Markovo grandis. 2.2 pav. pavaizduotas pasikliautinojo intervalo ilgis grandinės pradžioje yra didelis, o po to sumažėja iki reikiamos reikšmės ε. Atitinkamai nutraukimo kriterijus (2.3 pav. *testas*) mažėja tol, kol pasiekiama nutraukimo kritinė reikšmė. Šiame grafike pavaizduota *kritinė reikšmė* yra χ_p^2 skirstinio su $p = 10$ laisvės laipsniais 0,99-kvantilio reikšmė. 2.4 pav. pateikta imties tūrio, apskaičiuoto pagal (2.38) taisyklę (pažymėtas *N spėjamas*), priklausomybė. Pavaizduotas realus imties tūris (pažymėtas *N realus*) yra gautas nutraukus imties generavimą taip, kad (2.35) pasikliautinojo intervalo ilgis neviršytų pasirinktos kritinės reikšmės ε.

Algoritmo nutraukimo sąlygos (*a*) atveju pradėjo galioti po $k = 8$ iteracijų. Gautos šios parametrų reikšmės:

$$\mu = \begin{pmatrix} 1{,}13 \\ 2{,}34 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1{,}54 & 0{,}64 \\ 0{,}64 & 3{,}8 \end{pmatrix}, \quad \Theta = \begin{pmatrix} 2{,}52 & 0{,}16 \\ 0{,}16 & 0{,}97 \end{pmatrix}, \quad \alpha = 4{,}79, \quad \eta = 0{,}58. \qquad (2.40)$$

(*b*) atveju, kai pasikliautinasis intervalas mažesnis, nutraukimo sąlygos patenkintos po $k = 43$ iteracijų.

Galima matyti, kad stochastiniu algoritmu gauti (2.40) įverčiai yra artimi (2.39) įverčiams, gautiems analiziniu būdu.

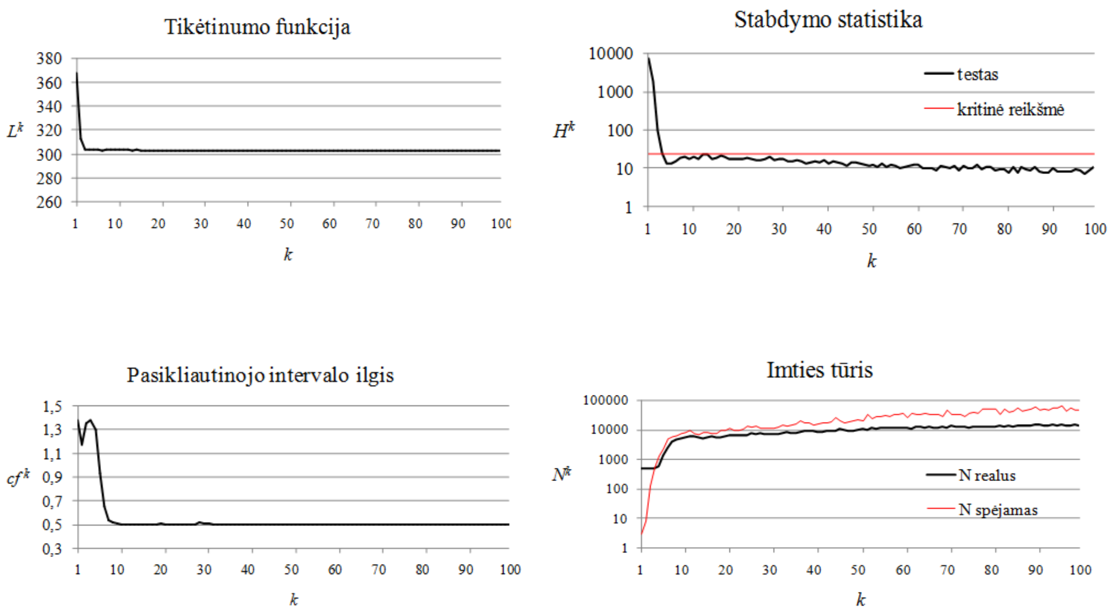
Jeigu Monte-Karlo imties tūris nebūtų reguliuojamas pagal (2.38) taisyklę ir pasirinktas lygus $N = 10\,000$, kai $\varepsilon = 0{,}5$, tai konvergavimas būtų pasiektas po $k = 8$ iteracijų.

Bendras Monte-Karlo imčių tūris, reikalingas sustojimo sąlygoms patenkinti, (*a*) atveju yra 11 245, kai paskutiniojoje grandyje buvo sugeneruota 4 782, (*b*) atveju tam reikia 501 545, kai paskutiniojoje grandyje buvo sugeneruota 29 434, dydžio imtis. Taikant šiam uždaviniui standartinį MCMC metodą su fiksuota imtimi, tai viso būtų reikėję $7 \cdot 4\,000 = 28\,000$ (kitu atveju $42 \cdot 30\,000 = 1\,260\,000$) Monte-Karlo bandymų uždaviniui išspręsti arba būtų tekę nutraukti skaičiavimus anksčiau dar nepasiekus sprendinių reikiamu tikslumu. Tokiu būdu adaptuotas MCMC algoritmas leido 2,5 karto sutrumpinti skaičiavimus.

2.2 lentelė. Asimetrinio *t* skirstinio standartinio ir adaptuoto MCMC algoritmų palyginimas

ε	Imčių tūris	k	Paskutiniosios iteracijos N^k	Bendras N
0,5	reguliuojamas	8	4 782	11 245
	fiksuotas	7	4 000	28 000
0,2	reguliuojamas	43	29 434	501 545
	fiksuotas	42	30 000	1 260 000

2.5 paveiksluose pavaizduotos suvidurkintos uždavinio charakteristikų priklausomybės priklausomai nuo sugeneruotų Markovo grandžių skaičiaus, kai $\varepsilon = 0{,}5$. Vidurkinimas buvo atliekamas pagal 20 algoritmo realizacijų. Sustojimo sąlygos buvo patenkintos visose realizacijose.



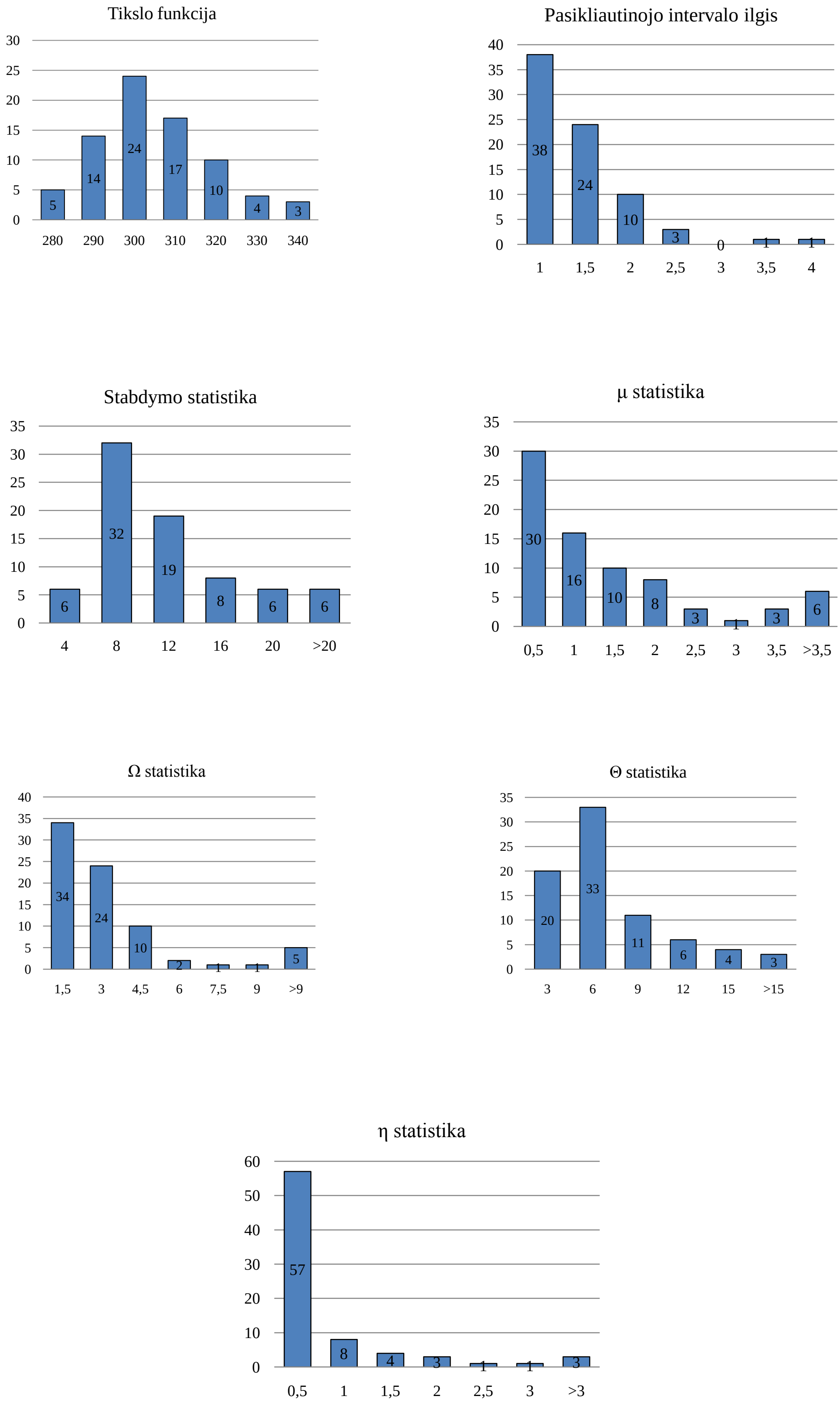
2.5 pav. Suvidurkintos asimetrinio *t* skirstinio vertinimo MCMC algoritmu charakteristikos

Bendru atveju, nagrinėjamos statistinio modelio parametrų vertinimą MCMC metodu, galima nagrinėti kaip stochastinio netiesinio tikėtinumo funkcijos (2.2) optimizavimo uždavinį (2.3). Kadangi sukurtas MCMC algoritmas stochastinio optimizavimo požiūriu yra stochastinio gradientinio nusileidimo metodas (2.17), galima pasinaudoti teoriniais rezultatais (Sakalauskas, 2000), iš kurių seka, kad šis algoritmas konverguoja su tikimybe 1 į stochastinio optimizavimo uždavinio sprendinį, o imties tūris auga apytiksliai kaip geometrinė progresija.

Tokiu būdu, priklausomybės 2.1–2.5 patvirtina šias teorines išvadas apie sukurto metodo konvergavimą ir racionalų Monte-Karlo bandymų skaičiaus parinkimą, vertinant nagrinėjamo statistinio modelio parametrus, nes imties tūris grandinės pradžioje būna nedidelis, ir tik galimų sprendinių zonoje jis padidėja iki imties tūrio, pakankamo įvertinti tikėtinumo funkcijos pasikliautinąjį intervalą reikiamu tikslumu. Iš pateiktų priklausomybių matyti, jog grandinės pradžioje šis intervalas yra didelis, o po to sumažėja iki reikiamos reikšmės (0,5 arba 0,2).

Prielaidai apie adaptuotu MCMC metodu gautų statistikų asimptotinį pasiskirstymą optimaliame taške pagal vienmatį arba daugiamatį normalųjį dėsnį patikrinti buvo atliktas skaičiavimo eksperimentas. Šiam eksperimentui buvo sugeneruota $M = 100$ imčių, kurių kiekviena sudarė $K = 100$ asimetrinio t skirstinio reikšmių, ir pasinaudojus *MathCad* integralų apskaičiavimo programa, buvo analitiškai rasti asimetrinio t skirstinio parametrai kiekvienai imčiai. Pašalinus imtis su singuliariais kovariacijų matricų Ω, Θ įverčiais, gautos 77 imtys, kurių kiekvienai buvo sugeneruota tūrio $N = 2000$ nepriklausomų Monte-Karlo imčių (2.18), ir apskaičiuota (2.37) stabdymo statistika bei kitos adaptuoto MCMC metodo statistikos. Gautų imčių histogramos pavaizduotos 2.6 pav.

Statistiniai χ^2 ir Kolmogorovo-Smirnovo kriterijai neprieštaravo su reikšmingumu 0,05 padarytai prielaidai apie asimptotinį (2.37) stabdymo statistikos pasiskirstymą pagal χ_p^2 dėsnį su $p = 10$ laisvės laipsniais. Kartu su (2.37) stabdymo statistika buvo apskaičiuotos statistikos, skirtos atskiroms (2.36) hipotezėms patikrinti, kurių histogramos taip pat pavaizduotos 2.6 pav. Statistiniai χ^2 ir Kolmogorovo-Smirnovo kriterijai neprieštaravo prielaidai apie šių statistikų asimptotinį pasiskirstymą pagal χ_p^2 dėsnį su atitinkamu laisvės laipsniu ir reikšmingumu 0,001–0,05.



2.6 pav. Stabdymo statistikų tyrimas

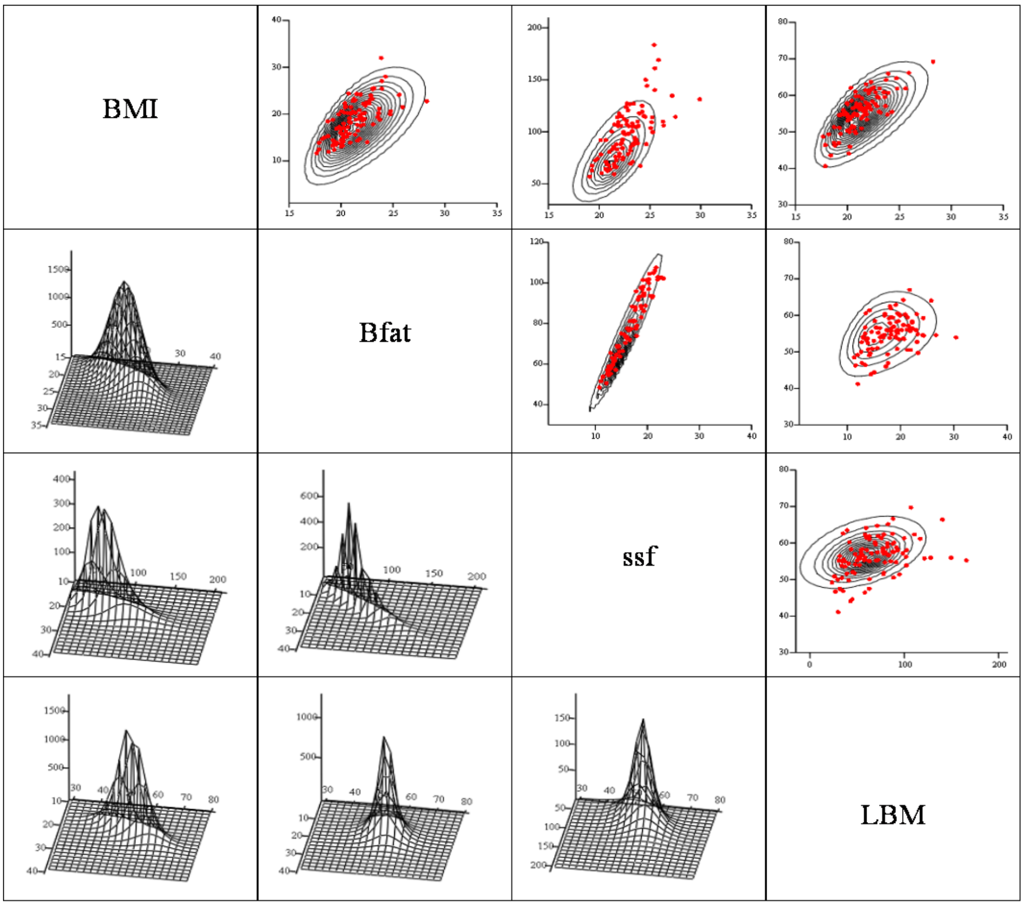
2.5.2. Australijos sporto instituto duomenys

Algoritmas testuojamas pasinaudojus tarptautine Australijos sportininkų duomenų baze, kuri yra dažnai naudojama statistiniams skaičiavimams atlikti (Smyth, GK, 2011; Cook ir Weisberg, 1994). Kiekvienai duomenų – kūno masės indeksas (*BMI*), kūno riebalų masė (*Bfat*), odos raukšlių skaičius (*ssf*) ir liesa kūno masė (*LBM)* – porai (4 priedas) įvertintas dvimatis asimetrinio *t* skirstinio modelis (Sakalauskas ir Vaiciulyte, 2012b; Vaiciulyte, 2012). Gauti asimetrinio *t* skirstinio parametrų įverčiai pateikti 2.3 lentelėje.

2.3 lentelė. Asimetrinio *t* skirstinio įverčiai

	(*BMI*, *Bfat*)	(*BMI*, *ssf*)	(*BMI*, *LBM*)	(*Bfat*, *ssf*)	(*Bfat*, *LBM*)	(*ssf*, *LBM*)
$\hat{\mu}_1$	19,60	20,87	19,76	13,52	12,08	59,35
$\hat{\mu}_2$	16,42	57,62	53,08	55,86	54,77	54,86
$\hat{\Omega}_{11}$	33,33	10,58	15,59	19,90	97,81	344,53
$\hat{\Omega}_{12}$	79,21	66,55	43,86	89,18	135,81	94,65
$\hat{\Omega}_{22}$	312,15	891,75	204,12	414,36	440,01	26,00
$\hat{\Theta}_{11}$	99,93	1,88	42,12	24,68	519,00	1086
$\hat{\Theta}_{12}$	56,09	61,09	37,83	186,20	23,60	-14,53
$\hat{\Theta}_{22}$	31,49	1987	33,98	1405	1,07	11,79
$\hat{b}$	12,70	3,05	6,62	2,33	11,23	1,93
$\hat{L}$	515,19	699,95	529,98	663,27	634,54	826,50

2.7 pav. pavaizduotas 100 moterų asimetrinio *t* skirstinio tankio paviršiaus ploto ir kontūro linijos bei duomenų sklaidos diagrama. Pateikti grafikai vaizdžiai iliustruoja asimetrinio *t* skirstinio tankio asimetriją ir pasiskirstymo ypatybes.



2.7. pav. Asimetrinio *t* skirstinio tankio paviršiaus ploto, kontūro linijos bei duomenų sklaidos diagrama

2.5.3. Finansinių duomenų tyrimas

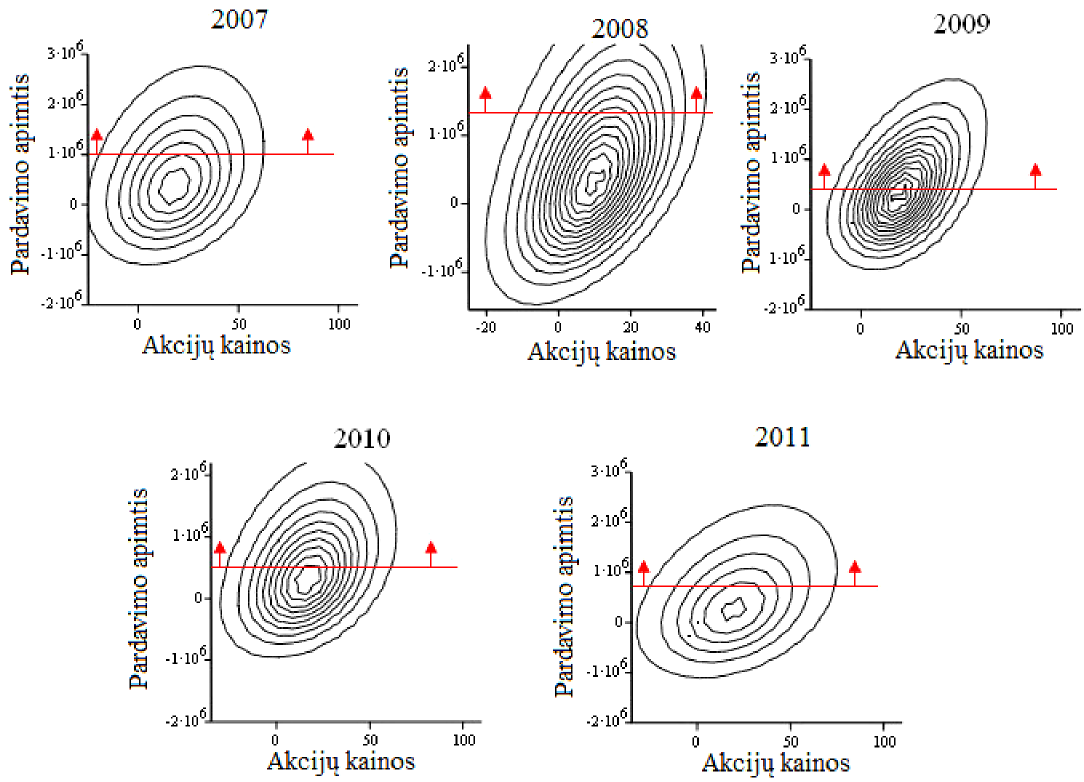
Daugelis statistinių išvadų grindžiamos prielaida, kad stebimas atsitiktinis dydis turi normalųjį skirstinį (vienas iš plačiausiai matematinėje statistikoje naudojamų skirstinių). Žinoma, tariant, jog stebimas atsitiktinis dydis turi normalųjį skirstinį, matematiniame modelyje ignoruojama tai, kad stebimas atsitiktinis dydis gali įgyti labai didelių arba labai mažų reikšmių (Čekanavičius, Murauskas, 2002). Taigi vertinant ekonominius rodiklius normaliojo skirstinio tankio funkcija nėra geriausias variantas ir gana dažnai netinka ekonominių rodiklių analizei (Kabasinskas ir kt., 2009). Beje, šiame dėsnyje neatsižvelgiama į asimetriją tarp investuotojų lūkesčių (tikimybė gauti pelną) ir rezultatų. Investuotojas siekia, kaip galima daugiau žinoti apie investavimo alternatyvas, tačiau tai ne visada įmanoma. Šį skirtumą geriau atspindi asimetrinis *t* skirstinys, nes jame duomenys aprašomi normaliuoju skirstiniu su atsitiktiniu vidurkių vektoriumi ir atsitiktine dispersija, atsižvelgiant į asimetriją ir dideles atsitiktines reikšmes.

Toliau pateikiamas pavyzdys, kuriame asimetrinis *t* skirstinys nagrinėjamas ekonominės situacijos kontekste, skirstinių parametrais aprašant investuotojų nuotaikas ir apsisprendimus, ir pavaizduojant investuotojų lūkesčių srities kitimą kriziniu ir po kriziniu laikotarpiu (Sakalauskas ir kt., 2013, 2014; Vaičiulytė, 2014).

Dvimačio asimetrinio *t* skirstinio parametrų vertinimas atliekamas pasinaudojus įmonių reitingų duomenimis, t. y. nagrinėjamos sveikatos apsaugos industrijai priklausančių atskirų sektorių įmonių kiekvienos akcijų pardavimo dienos kainos (*Close*) ir pardavimo apimtis (*Volume*) nuo 2007 iki 2011 metų. Stebima, kaip įmonių duomenų asimetrinio *t* skirstinio parametrai kinta kiekvienais metais ir pagal tai aptariami investuotojų lūkesčiai.

Duomenys. Disertacijoje tiriamos 130 JAV įmonių finansiniai duomenys imti iš http://finance.yahoo.com/, apima laikotarpį nuo 2007 iki 2011 metų. Tai ekonominio nuosmukio, kuris prasidėjo 2007 m. pabaigoje, laikotarpis. 2008 m. būdingos itin ryškios likvidumo problemos, 2009 m. pabaigoje sunkmetis įgijo kitą pobūdį, o 2011 metai laikomi sunkmečio pabaiga. Buvo sudaryta dvimatė imtis su *Close* ir *Volume* duomenimis ir apskaičiuoti šios imties asimetrinio *t* skirstinio modelio parametrai disertacijoje aprašytu metodu.

Tyrimo rezultatai. 2.8 pav. nubraižytos asimetrinio *t* skirstinio tankio kontūro linijos rodo duomenų poros *Close* ir *Volume* sklaidą kiekvienais metais. Pasinaudojus statistine asimetrinio *t* skirstinio interpretacija (žr. 2.2 skyrelį), galima pavaizduoti investuotojų lūkesčių sritį, kurią sudaro asimetrinio *t* skirstinio tankio integravimo pusplokštumė, pažymėta raudonomis rodyklėmis. 2008 metais pastebimas investuotojų lūkesčių srities sumažėjimas, atspindintis realią tų metų situaciją: investuotojų pasitikėjimas buvo sumažėjęs. Po krizės – 2009 m. – jis išaugo ir vėl priartėjo prie 2007 m. ribos.



2.8 pav. 2007–2011 metų akcijų kainos (*Close*) ir pardavimo apimties (*Volume*) duomenų asimetrinio *t* skirstinio tankio kontūro linijos

Kaip parodė rezultatai 2.8 pav., ekonominė krizė turėjo įtakos investuotojų lūkesčių kitimui. Gautame reitingų duomenų skirstinyje pav. 2.8 pavaizduota investuotojų lūkesčių sritis atspindi jų nuomonę apie tikėtinus įmonių rezultatus. Šis pavyzdys parodo, kad sudarytas modelis tinka finansinių duomenų modeliavimui.

2.6. Skyriaus išvados

Analiziniai ir skaitiniai tyrimai leidžia tvirtinti, kad disertacijoje sudarytas adaptuotas MCMC algoritmas gali būti pritaikytas asimetrinio *t* skirstinio parametrams vertinti norimu tikslumu. Parodyta, kad šis metodas realizuoja logaritminės tikėtinumo funkcijos stochastinę gradientinę paiešką, vykdant ją „fiksuoto taško“ („paprastos“ iteracijos) arba EM algoritmo pavidalu. Sukurtas metodas pasižymi Monte-Karlo imties tūrio parinkimo taisykle bei modeliavimo paklaidų įvertinimu statistiniu būdu atskirose grandyse. Tokiu būdu adaptuotas MCMC algoritmas sutrumpina skaičiavimų apimtį maždaug 2,5 karto. Sukurtas metodas leidžia į optimizuojamus parametrus įvesti ir formos bei integravimo srities parametrus, kurie įvertinami taip pat stochastiniu algoritmu. Sukurtas algoritmas buvo pritaikytas Australijos sporto instituto duomenų ir finansinių 2007–2011 metų JAV įmonių duomenų analizei. Atlikti eksperimentai patvirtino, kad metodo skaitinės savybės atitinka teorinį modelį. Bendru atveju šis algoritmas yra naudingas stochastinio pobūdžio sistemų tyrimui ir kitiems statistikos uždaviniams spręsti MCMC metodu.

3 skyrius. DAUGIAMATIS RETŲ ĮVYKIŲ TIKIMYBIŲ VERTINIMO PUASONO-GAUSO MODELIS

3.1. Įvadas

Šiame skyriuje nagrinėjamas adaptuotas MCMC algoritmas, taikomas dažnių statistikos išvadoms gauti Bajeso metodu. Darbai, akcentuojantys prielaidas, kuriomis remiasi Bajeso teorija, yra aprašyti autorių: Besag (1989), Fishman (2003), Gamerman (2006), ir kt. Bajeso metodas plačiai taikomas įvairiems sprendimo priėmimo uždaviniams spręsti bei verslo ar finansinių rodiklių vertinimui (Bradley ir Thomas, 2000; Clayton ir Kaldor, 1987; Chen, 2009; Liseo ir Loperfido, 2003; Tsutakawa ir kt. 1985). Bajeso analizė pirmiausia priklauso nuo modelio apriorinio skirstinio parametrų pasiskirstymo. Šis skirstinys gali būti neparametrinis arba parametrinis ir priklausyti nuo nežinomų parametrų, kurie savo ruožtu gali būti sudaryti keliais lygiais. Šis parametrų ir apriorinių skirstinių nuoseklumas ir sudaro hierarchinį modelį. Hierarchija turi kokiame nors taške baigtis (sustoti), darant prielaidą, kad visi likusieji pirminiai parametrai jau yra žinomi. Empirinis Bajeso (EB) metodas naudoja stebėjimų duomenis tam, kad būtų galima įvertinti šiuos modelio parametrus, ir tada tęsia procesą taip, lyg visi apriorinio skirstinio parametrai būtų jau žinomi (Tanner ir Wong, 1987). EB metodas nuo Bajeso metodo skiriasi tuo, kad apriorinio skirstinio parametrai yra įvertinami pasinaudojus duomenų imtimi.

Nors iš principo EB gali būti įgyvendinamas hierarchiniams modeliams, turintiems bet kokį lygių skaičių, tam, kad būtų paprasčiau, pateikiamas metodas, skirtas dviejų lygių modeliui.

Disertacijoje, pasinaudojus Puasono-Gauso modeliu, konstruojamas adaptuotas MCMC algoritmas, skirtas kelių retų įvykių tikimybėms vertinti. Retų įvykių tikimybių vertinimo uždaviniai iškyla, nagrinėjant draudiminius įvykius (skirtingų automobilių rūšių avarijų skaičių ir pan.), susirgimų arba mirties atvejus populiacijose ir pan. Aposteriorinis įvykių dažnių skirstinys yra sudaromas Bajeso metodu (žr. 1.5.2 skyrelį).

Tegu Y – atsitiktinis dydis, kurio įgyjamos reikšmės yra 0 ir 1. Disertacijoje taikoma nagrinėjamo įvykio tikimybės P *logit* transformacija:

$$\rho = \ln \frac{P}{1-P}. \tag{3.1}$$

su tikimybėmis $P = P(Y=1) \in (0,1)$, $P(Y=0) = 1-P$ (Altaleb ir Chauveau, 2002; Pearce, 2006). Pasinaudojant (3.1), tikimybė per transformaciją ρ išreiškiama taip:

$$P = \frac{1}{1+e^{-\rho}}. \tag{3.2}$$

Vieno įvykio tikimybės vertinimo empiriniu Bajeso metodu algoritmas buvo sudarytas Tsutakawa ir kt. (1985), iškėlus apriorinę prielaidą, kad *logit* transformacijos yra pasiskirsčiusios pagal normalųjį skirstinį. Šis algoritmas buvo nagrinėjamas autorių: Jakimauskas ir Sakalauskas (2012), Sakalauskas (2010), Bradley ir Thomas (2000), Berger (1985) bei kt. Toliau disertacijoje nagrinėjamas šio algoritmo apibendrinimas daugiamačiu atveju ir pasiūlytas pradinių reikšmių parinkimo būdas daugiamačiam Puasono-Gauso modeliui. Skyriuje gauti rezultatai buvo publikuoti šiuose moksliniuose leidiniuose: Sakalauskas ir Vaičiulytė (2012c), Sakalauskas ir Vaičiulytė (2013), Vaičiulytė ir Sakalauskas (2014).

3.2. Puasono-Gauso modelis

Sakykim, turime K populiacijų aibę $C = (c^1, c^2, \ldots, c^K)$, i-oje populiacijoje c^i yra I_i individų, $i = 1, 2, \ldots, K$. Tarkime, kad populiacijose gali įvykti M kažkokių įvykių (susirgimai, mirtys, draudiminiai įvykiai ir pan.)

$$I_{ij}^{m}=\begin{cases}1, \text{ jei } j \text{ individui įvyko } m \text{ įvykis,}\\ 0, \text{ priešingu atveju;}\end{cases}$$

čia $m=1, 2, \ldots, M$, $j \in c^{i}$. Tikslas yra įvertinti įvykių tikimybes

$$P_{i}^{m}=P\left(Y_{i}^{m}=\sum_{j \in c^{i}} I_{ij}^{m}\right),$$

kai Y_{i}^{m} yra *m*-tojo įvykio pasirodymų skaičius *i*-toje populiacijoje, $i=1, 2, \ldots, K$, $m=1, 2, \ldots, M$.

Dėl didelių populiacijų dydžio I_{i} skirtumų paprastas santykinis įvertinys $\frac{Y_{i}^{m}}{I_{i}}$ ne visada leidžia patikimai įvertinti įvykių tikimybes (Clayton ir Kaldor, 1987). Todėl disertacijoje nagrinėjamas kelių retų įvykių tikimybių empirinis Bajeso Puasono-Gauso modelis, leidžiantis įvertinti koreliacinius sąryšius tarp skirtingų įvykių dažnių.

Empiriniame Bajeso metode laikoma, kad *m*-tojo įvykio *i*-tojoje populiacijoje skaičius Y_{i}^{m} pasiskirstęs pagal Puasono dėsnį su parametru λ_{i}^{m} :

$$Y_{i}^{m} \sim \operatorname{Pois}\left(\lambda_{i}^{m}\right), \ \lambda_{i}^{m}=I_{i} \cdot P_{i}^{m}, \tag{3.3}$$

Tikimybė, kad Y_{i}^{m} įgis tam tikrą reikšmę, lygi (Aksomaitis, 2000):

$$f_{Y_{i}^{m}}\left(\varpi, \lambda_{i}^{m}\right)=P\left(Y_{i}^{m}=\varpi ; \lambda_{i}^{m}\right)=e^{-\lambda_{i}^{m}} \frac{\left(\lambda_{i}^{m}\right)^{\varpi}}{\varpi !}, \ \varpi=0, 1, 2, \ldots$$

Puasono-Gauso modelyje priimama, kad įvykių tikimybių *logit* transformacija $\rho_{i}^{m}=\operatorname{logit} P_{i}^{m}$ populiacijose yra pasiskirsčiusi pagal daugiamatį normalųjį skirstinį su parametrais μ, Ω (Bradley ir Thomas, 2000; Tsutakawa ir kt., 1985):

$$\rho_i^m \sim N(\mu, \Omega), \tag{3.4}$$

t. y. ρ_i^m skirstinio tankis

$$g(\rho_i^m, \mu, \Omega) = \frac{\exp\left(-(\rho_i^m - \mu)^T \Omega^{-1} (\rho_i^m - \mu)\right)}{\sqrt{|\Omega|} \cdot (2\pi)^{\frac{M}{2}}},$$

$i = 1, 2, \ldots, K$, $m = 1, 2, \ldots, M$.

Pasinaudojant (3.2) formule ir (3.3) aproksimacija, tikimybių P_i^m įvertiniai apskaičiuojami per (1.8) aposteriorinius tikimybių tankius

$$p(P_i^m | Y) = \frac{\int_{-\infty}^{+\infty} \frac{1}{1+e^{-\rho_i^m}} \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \mu, \Omega) d\rho_i^m}{B_i(\mu, \Omega)}, \tag{3.5}$$

čia

$$B_i(\mu, \Omega) = \int_{-\infty}^{+\infty} \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \mu, \Omega) d\rho_i^m,$$

$i = 1, 2, \ldots, K$, $m = 1, 2, \ldots, M$.

Taikant Bajeso metodą statistikoje dažnai tenka minimizuoti tam tikras funkcijas, išreikštas per aposteriorinio tankio integralą. Nagrinėjamu atveju nežinomi parametrai μ, Ω vertinami didžiausio tikėtinumo metodu (Bradley ir Thomas, 2000; Tsutakawa ir kt., 1985). Nesunku įsitikinti, kad logaritminė tikėtinumo funkcija, atmetus narius, nepriklausančius nuo modelio parametrų, yra

$$\begin{aligned} L(\mu, \Omega) &= -\sum_{i=1}^{K} \ln\left(p(P_i^m | Y)\right) = \\ &= -\sum_{i=1}^{K} \ln\left(\int_{-\infty}^{+\infty} \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \mu, \Omega) d\rho_i^m\right) = -\sum_{i=1}^{K} \ln\left(B_i(\mu, \Omega)\right), \end{aligned} \tag{3.6}$$

kurią minimizavus gaunami parametrų μ, Ω įvertiniai.

3.3. Didžiausio tikėtinumo funkcijos išvestinės

(3.6) tikėtinumo funkcija yra analizinė funkcija, t. y. diferencijuojama be galo daug kartų pagal parametrus μ, Ω. Nesunku įsitikinti, kad šios funkcijos pirmos eilės išvestinės yra:

$$\frac{\partial L(\mu,\Omega)}{\partial \mu} = \sum_{i=1}^{K} \frac{\int_{-\infty}^{+\infty} \Omega^{-1}(\rho_i^m - \mu) \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \mu, \Omega) d\rho_i^m}{B_i(\mu,\Omega)}, \tag{3.7}$$

$$\frac{\partial L(\mu,\Omega)}{\partial \Omega} = \\ = -\sum_{i=1}^{K} \frac{\int_{-\infty}^{+\infty} \left(\Omega^{-1} - \Omega^{-1}(\rho_i^m - \mu)(\rho_i^m - \mu)^T \Omega^{-1}\right) \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \mu, \Omega) d\rho_i^m}{B_i(\mu,\Omega)}. \tag{3.8}$$

Prilyginus nuliui (3.7) ir (3.8) išvestines, gaunamos lygtys, kurias išsprendus gaunami Puasono-Gauso modelio parametrų įverčiai:

$$\hat{\mu} = \frac{1}{K} \sum_{i=1}^{K} \frac{\int_{-\infty}^{+\infty} \rho_i^m \cdot \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \hat{\mu}, \hat{\Omega}) d\rho_i^m}{B_i(\hat{\mu},\hat{\Omega})}, \tag{3.9}$$

$$\hat{\Omega} = \frac{1}{K} \sum_{i=1}^{K} \frac{\int_{-\infty}^{+\infty} (\rho_i^m - \hat{\mu})(\rho_i^m - \hat{\mu})^T \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g(\rho_i^m, \hat{\mu}, \hat{\Omega}) d\rho_i^m}{B_i(\hat{\mu},\hat{\Omega})}, \tag{3.10}$$

kurios bus pritaikytos EM algoritmui sudaryti.

3.4. Puasono-Gauso modelio parametrų įvertiniai

Pasinaudojant (3.9), (3.10) lygybėmis galima sudaryti „fiksuoto taško“ metodą parametrų μ, Ω didžiausio tikėtinumo įvertiniams rasti, statistikoje dar vadinamą EM algoritmu (žr. 1.5 skyrelį):

$$\mu^{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{\int_{-\infty}^{+\infty}\rho_i^m \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g\left(\rho_i^m, \mu^k, \Omega^k\right) d\rho_i^m}{B_i\left(\mu^k, \Omega^k\right)}, \tag{3.11}$$

$$\Omega^{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{\int_{-\infty}^{+\infty}\left(\rho_i^m - \mu^k\right)\left(\rho_i^m - \mu^k\right)^T \prod_{m=1}^{M} f\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) g\left(\rho_i^m, \mu^k, \Omega^k\right) d\rho_i^m}{B_i\left(\mu^k, \Omega^k\right)}, \tag{3.12}$$

Kadangi yra nagrinėjamos retų įvykių tikimybės $\left(P_i^m \approx 0\right)$, o $logit\, P_i^m \sim N(\mu, \Omega)$, galima daryti prielaidą, kad šios tikimybės yra apytiksliai pasiskirsčiusios pagal log-normalinį dėsnį. Pasinaudojant šia prielaida, (3.11) ir (3.12) lygybėse galima paimti tokį euristinį pradinį tašką $\left(\mu^0, \Omega^0\right)$:

$$\mu^0 = \ln(P) - \frac{1}{2}\Omega^0, \tag{3.13}$$

$$\Omega^0 = \ln\left(\frac{\sum_{i=1}^{K}\Psi^{-1}\cdot\left(Y_i - I_i P\right)\cdot\left(Y_i - I_i P\right)^T\cdot\Psi^{-1}}{\left(\sum_{i=1}^{K} I_i\right)^2} + 1\right), \tag{3.14}$$

čia $P = \left(\frac{\sum_{i=1}^{K} Y_i^1}{\sum_{i=1}^{K} I_i}, \frac{\sum_{i=1}^{K} Y_i^2}{\sum_{i=1}^{K} I_i}, \ldots, \frac{\sum_{i=1}^{K} Y_i^M}{\sum_{i=1}^{K} I_i}\right)$, $\Psi = \mathrm{Diag}(P)$, $Y_i = \left(Y_i^1, Y_i^2, \ldots, Y_i^M\right)$,

$i = 1, 2, \ldots, K$.

3.5. Markovo grandinės Monte-Karlo algoritmas

(3.13), (3.14) „fiksuoto taško“ algoritmą galima realizuoti adaptuotu MCMC metodu. Tegu sugeneruota *k* Markovo grandžių ir kiekvienoje apskaičiuoti įverčiai μ^k, Ω^k, paėmus (3.13) ir (3.14) pradines reikšmes. Taigi, kiekvienoje grandyje generuojami atsitiktiniai daugiamačiai vektoriai, pasiskirstę pagal normalųjį dėsnį (žr. 1.3.2 skyrelį),

$$\rho_{i,j}^m \sim N(\mu^k, \Omega^k), \quad j = 1, \ldots, N^k,$$

čia N^k yra imties tūris *k*-toje Markovo grandyje. Siekiant išvengti skaičiavimo problemų, galinčių atsirasti dėl labai mažų tarpinių rezultatų reikšmių, skaičiuojant integralus pagal (3.5)–(3.12) formules įvedama pagalbinė funkcija

$$\begin{aligned} r_i\left(\rho_i^m\right) &= \ln\left\{\prod_{m=1}^{M} f_i\left(Y_i^m, \frac{I_i}{1+e^{-\rho_i^m}}\right) \Big/ \prod_{m=1}^{M} f_i\left(Y_i^m, \frac{I_i}{1+e^{-\mu^m}}\right)\right\} = \\ &= \exp\left\{\sum_{m=1}^{M} \frac{I_i\left(e^{-\rho_i^m} - e^{-\mu^m}\right)}{\left(1+e^{-\mu^m}\right)\left(1+e^{-\rho_i^m}\right)} + Y_i^m \cdot \ln\left(\frac{1+e^{-\mu^m}}{1+e^{-\rho_i^m}}\right)\right\}. \end{aligned} \tag{3.15}$$

Apskaičiuojamos sumos:

$$\tilde{B}_i^k = \sum_{j=1}^{N^k} r_i\left(\rho_{i,j}^m\right), \tag{3.16}$$

$$\tilde{G}_i^k = \sum_{j=1}^{N^k} r_i\left(\rho_{i,j}^m\right)^2, \tag{3.17}$$

$$\tilde{\upsilon}_i^k = \sum_{j=1}^{N^k} \rho_{i,j}^m \cdot r_i\left(\rho_{i,j}^m\right), \tag{3.18}$$

$$\tilde{S}_i^k = \sum_{j=1}^{N^k} \left(\rho_{i,j}^m - \tilde{\upsilon}_i^k\right) \cdot \left(\rho_{i,j}^m - \tilde{\upsilon}_i^k\right)^T \cdot r_i\left(\rho_{i,j}^m\right), \tag{3.19}$$

$$\tilde{p}_{i,m}^{k} = \sum_{j=1}^{N^k} \frac{r_i\left(\rho_{i,j}^{m}\right)}{1+e^{-\rho_{i,j}^{m}}}, \tag{3.20}$$

kuriomis pasinaudojama apskaičiuojant pagal (3.13) ir (3.14) kitos iteracijos modelio parametrų įvertinius

$$\mu^{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{\tilde{\upsilon}_i^k}{\tilde{B}_i^k}, \tag{3.21}$$

$$\Omega^{k+1} = \frac{1}{K}\sum_{i=1}^{K}\frac{\tilde{S}_i^k}{\tilde{B}_i^k}, \tag{3.22}$$

ir tikėtinumo funkcijos įvertinį

$$L^k\left(\mu,\Omega\right) = -\sum_{i=1}^{K}\ln\left(\frac{\tilde{B}_i^k}{N^k}\right). \tag{3.23}$$

Tikėtinumo funkcijos dispersija aproksimuojama taip (žr. (1.21)):

$$DL^k = \sum_{i=1}^{K}\left(\frac{\tilde{G}_i^k \cdot N^k}{\left(\tilde{B}_i^k\right)^2} - 1\right). \tag{3.24}$$

Naudojantis (3.5), populiacijose galima įvertinti įvykių tikimybes:

$$P_{i,m}^k = \frac{\tilde{p}_{i,m}^k}{\tilde{B}_i^k}. \tag{3.25}$$

Algoritmo stabdymui įvestas kriterijus

$$\begin{aligned} H^k = {} & \frac{K}{\frac{1}{K}\sum_{i=1}^{K}\frac{\tilde{G}_i^k}{\left(\tilde{B}_i^k\right)^2}} \times \\ & \times\left(-\ln\frac{\left|\Omega^{k+1}\right|}{\left|\Omega^k\right|} + \mathrm{tr}\left(\Omega^{k+1}\cdot\left(\Omega^k\right)^{-1}\right) + \left(\mu^{k+1}-\mu^k\right)^T\cdot\left(\Omega^k\right)^{-1}\cdot\left(\mu^{k+1}-\mu^k\right) - M\right). \end{aligned} \tag{3.26}$$

skirtas patikrinti hipotezei apie vidurkių ir kovariacijų matricų sutapimą dviejose gretimose iteracijose: $\mu^{k+1} = \mu^k$, $\Omega^{k+1} = \Omega^k$ (žr. (1.25), (1.27)).

Šis kriterijus, apskaičiuotas tikėtinumo funkcijos maksimumo taške, yra pasiskirstęs pagal Fišerio dėsnį su $p = \frac{M(M+3)}{2}$ laisvės laipsniais, jei imties tūris yra pakankamai didelis (1.7 skyrelį; Krishnaiah, 1984).

Algoritmas stabdomas, jei stabdymo kriterijus neprieštarauja hipotezei apie vidurkių ir kovariacijų matricų sutapimą dviejose gretimose iteracijose:

$$H^k \le F_{\delta,p}, \tag{3.27}$$

tikėtinumo funkcijos pasikliautinojo intervalo ilgis yra mažesnis už pasirinktą mažą reikšmę ε, $\varepsilon > 0$:

$$2 \cdot \tau_\gamma \cdot \sqrt{\frac{D^k X}{N^k}} \le \varepsilon, \tag{3.28}$$

o kovariacijų matricų įvertis nėra singuliarus. Čia $F_{\delta,p}$ ir τ_γ yra atitinkami Fišerio ir normaliojo skirstinio kvantiliai, δ, γ – pasikliovimo lygmuo. Vienmačio empirinio Bajeso modelio singuliarumo sąlygos buvo ištirtos Sakalausko (2010) ir Jakimausko (2014). Siekiant išvengti singuliarumo nagrinėjamu atveju, skaičiavimai būdavo nutraukiami, kai kovariacijų matricos įverčio didžiausios ir mažiausios tikrinių reikšmių santykis viršydavo pasirinktą kritinę reikšmę, kuri paprastai būna 10–30 (Lin ir Zhu, 2004). Disertacijoje ši reikšmė pasirinkta lygia 10. Pastaruoju atveju reikėtų taikyti analizės metodus, atsižvelgiančius į kovariacinės matricos singuliarumą, kuris disertacijoje nenagrinėjamas.

Jei nors viena stabdymo sąlyga nėra tenkinama, generuojama nauja Monte-Karlo imtis. Imties tūrį galima reguliuoti pagal taisyklę, analogišką (2.40) taisyklei:

$$N^{k+1} \geq \frac{N^k \cdot \nu}{H^k} \cdot F_{\nu,p}, \tag{3.29}$$

čia $F_{\nu,p}$ – Fišerio skirstinio kvantilis, ν – pasikliovimo lygmuo. Šios taisyklės taikymas leidžia racionaliai parinkti Monte-Karlo imčių tūrį Markovo grandinėje ir kartu užtikrina (3.21) ir (3.22) sekų konvergavimą į (3.6) tikėtinumo funkcijos minimumą su tikimybe 1 (Sakalauskas, 2000).

Nagrinėjamo uždavinio sprendimo algoritmas pateiktas 3.1 lentelėje.

3.1 lentelė. Puasono-Gauso modelio MCMC metodo tyrimo statistinio modeliavimo algoritmas

Puasono-Gauso modelio parametrų vertinimo algoritmas
Tikslas: MCMC metodu įvertinti nepalankių įvykių tikimybes. *Įėjimo parametrai*: populiacijų skaičius *K*, jose esančių individų skaičius I, tiriamų įvykių skaičius *M*, vidurkio vektorius μ^0, kovariacijų matrica Ω^0, pradinis minimalus ir maksimalus Markovo grandies Monte-Karlo imties tūriai N^0, $N_{\min}$, $N_{\max}$, maksimalus grandžių skaičius *k*. *Išėjimo parametrai*: logaritminės tikėtinumo funkcijos *L*, vidurkio vektoriaus bei kovariacijų matricos, įvykių tikimybių įverčių sekos, kiekvienos grandies pasikliautinojo intervalo ilgis *cf*, stabdymo statistika *H* ir imties tūris *N*. *1 žingsnis.* Priskiriamos pradinės parametrų reikšmės *2 žingsnis.* Generuojami atsitiktiniai daugiamačiai vektoriai, pasiskirstę pagal normalųjį skirstinį (3.4) *3 žingsnis.* Pagal Puasono dėsnį sugeneruojamas įvykių skaičius populiacijoje (3.3) *4 žingsnis.* Apibrėžiama logaritminė tikėtinumo funkcija (3.6) *5 žingsnis.* Sudaromas ciklas *for* iteracijų skaičius *iter* kinta nuo 1 iki *k* *6 žingsnis.* Ciklo kintamajam *n* priskiriamas 1 *7 žingsnis.* Pasikliautinojo intervalo ilgiui *cf* priskiriama didelė reikšmė

Puasono-Gauso modelio parametrų vertinimo algoritmas
8 žingsnis. *while* $n \le N$ ir $cf > \varepsilon$
9 žingsnis. *for i* kinta nuo 1 iki *K*
10 žingsnis. *for m* kinta nuo 1 iki *M*
11 žingsnis. Apskaičiuojamos sumos (3.16)–(3.20)
12 žingsnis. Kintamajam *n* priskiriama *n*+1 reikšmė
13 žingsnis. Apskaičiuojamas tikėtinumo funkcijos pasikliautinojo intervalo ilgis *cf* ir ciklas *while* vėl kartojamas (*8 žingsnis*)
14 žingsnis. *for i* kinta nuo 1 iki *K*
15 žingsnis. *for m* kinta nuo 1 iki *M*
16 žingsnis. Apskaičiuojami Monte-Karlo įverčiai (3.21)–(3.25)
17 žingsnis. Tikrinama stabdymo statistika (3.27)
18 žingsnis. Įvedamas imties tūrio reguliavimas (3.29)
19 žingsnis. Sudaromas masyvas duomenims išvesti ir toliau tęsiamas iteracijų ciklas (*5 žingsnis*)

3.6. Kompiuterinis modeliavimas

3.6.1. Skaitinis eksperimentas

Sudaryto algoritmo elgsenos tyrimui atlikti eksperimentai su parinktais pradiniais duomenimis (Sakalauskas ir Vaičiulytė, 2013) ir realiais duomenimis (Vaičiulytė ir Sakalauskas, 2014).

Šiame eksperimente sugeneruota imtis, sudaryta iš $K = 10$ populiacijų, kuriose galėjo įvykti $M = 3$ įvykiai, kai įvykių tikimybių *logit* transformacija pasiskirsčiusi pagal daugiamatį normalųjį skirstinį su parametrais

$$\mu = \begin{pmatrix} -3 \\ -4 \\ -5 \end{pmatrix}, \ \Omega = \begin{pmatrix} 0{,}25 & 0 & 0 \\ 0 & 0{,}25 & 0 \\ 0 & 0 & 0{,}25 \end{pmatrix}.$$

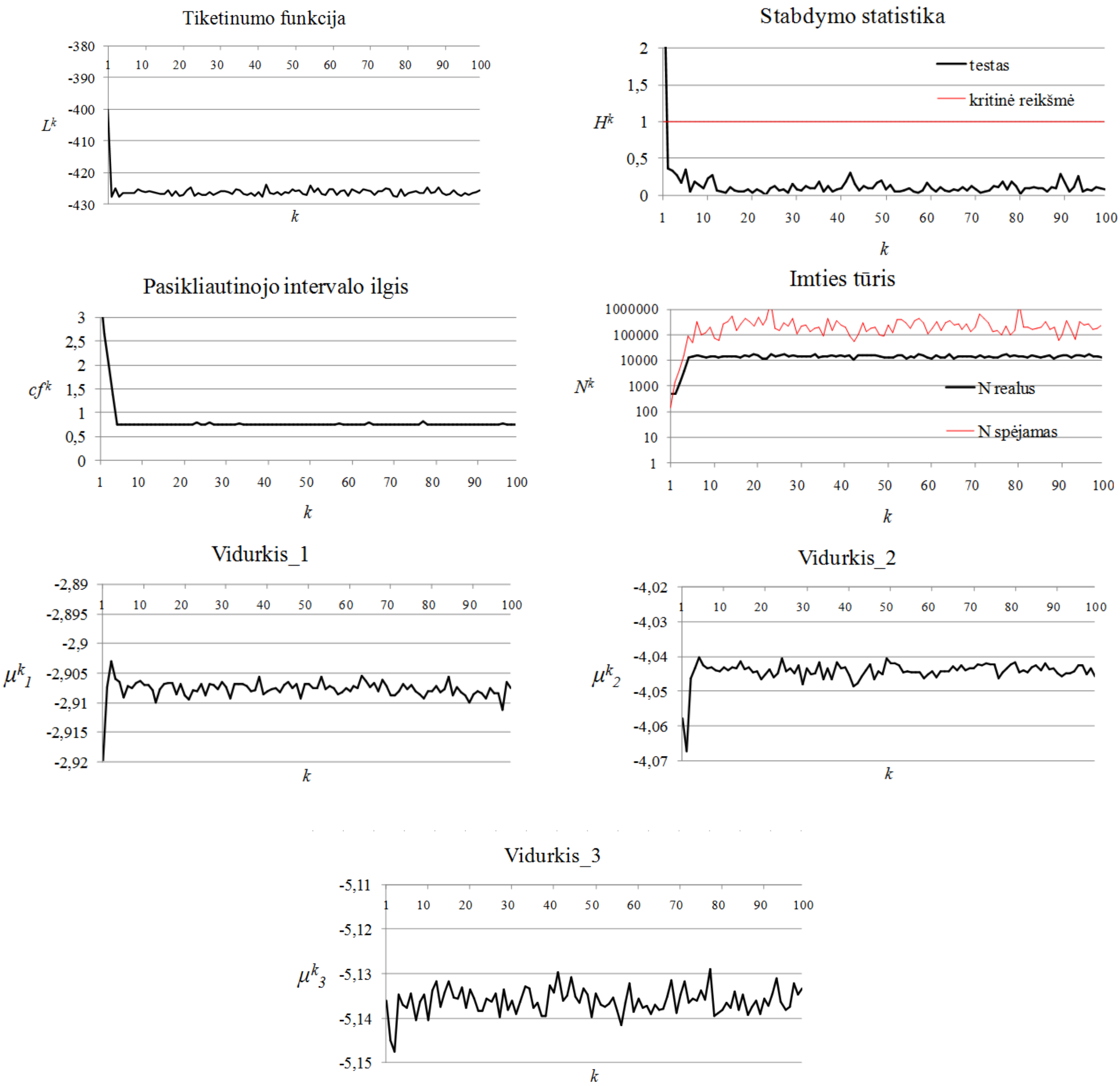
Sugeneruota $k = 100$ Markovo grandžių pagal (3.15)–(3.20), (3.29) išraiškas. Siekiant išvengti labai mažų arba labai didelių imties tūrio reikšmių, buvo taikomos ribos: $500 \le N^k \le 17000$.

Gautos modelio charakteristikos pateiktos 3.2 lentelėje.

3.2 lentelė. Skaitinio eksperimento Puasono-Gauso modelio įverčiai

Iteracija	μ_1	μ_2	μ_3	Tikslo funkcija	Pasikliautinojo intervalo ilgis	Imties tūris	Stabdymo statistika
1	–2,92	–4,06	–5,14	–400,12	3,56	500	3,51
2	–2,91	–4,07	–5,14	–427,64	2,65	500	0,37
3	–2,90	–4,05	–5,15	–424,93	2,01	1 359	0,33
4	–2,91	–4,04	–5,13	–427,53	1,37	4 122	0,28
5	–2,91	–4,04	–5,14	–426,53	0,75	13 602	0,17
6	–2,91	–4,04	–5,14	–426,59	0,75	14 779	0,35
7	–2,91	–4,04	–5,13	–426,57	0,75	15 322	0,05
8	–2,91	–4,04	–5,14	–426,47	0,75	14 867	0,17
9	–2,91	–4,04	–5,14	–425,28	0,75	13 098	0,14
10	–2,91	–4,04	–5,13	–425,74	0,75	14 131	0,08

3.1 pav. pavaizduotos priklausomybės nuo generuotų grandžių skaičiaus, kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}75$. Sprendinys pradėjo konverguoti jau po $k = 5$ iteracijų. 3.1 pav. pastebima, kad tikėtinumo funkcijos reikšmė mažėja, kol pasiekiama uždavinio sprendinių zona. Analogiškai algoritmo stabdymo kriterijus $\frac{H^k}{F_{\delta,p}}$ mažėja tol, kol pasiekiama nutraukimo kritinė reikšmė 1, čia $F_{\delta,p}$ yra Fišerio skirstinio su $p = 9$ laisvės laipsniais 0,95-kvantilio reikšmė. Grafike pateikta imties tūrio priklausomybė apskaičiuota pagal (3.29) taisyklę (pažymėtas *N spėjamas*). Šiame paveiksle pavaizduotas realus imties tūris (pažymėtas *N realus*) yra gautas nutraukus imties generavimą taip, kad (3.24) pasikliautinojo intervalo ilgis neviršytų pasirinktos kritinės reikšmės $\varepsilon = 0{,}75$. Iš pateiktų priklausomybių matyti, jog grandinės pradžioje pasikliautinasis intervalas yra didelis, o po to sumažėja iki reikiamos reikšmės ε.



3.1 pav. Skaitinio eksperimento Puasono-Gauso modelio vertinimo MCMC algoritmu charakteristikos

3.6.2. Socialinių duomenų tyrimas

Disertacijoje modeliuojami 2003 m. Lietuvos gyventojų savižudybių (*suic*) ir nužudymų (*hom*) duomenys, gauti iš Lietuvos sveikatos apsaugos ministerijos Higienos instituto (5 priedas). Nagrinėjami $M = 2$ įvykiai populiacijų aibėje sudarytoje iš $K = 60$ savivaldybių. Pradiniai parametrai apibrėžiami pagal (3.13) ir (3.14) formules:

$$\mu_{\text{visų}} = \begin{pmatrix} -7,77 \\ -9,25 \end{pmatrix}, \qquad \Omega_{\text{visų}} = \begin{pmatrix} 0,08 & 0 \\ 0,03 & 0,06 \end{pmatrix},$$

$$\mu_{\text{vyrų}} = \begin{pmatrix} -7,20 \\ -8,85 \end{pmatrix}, \qquad \Omega_{\text{vyrų}} = \begin{pmatrix} 0,08 & 0 \\ 0,03 & 0,07 \end{pmatrix},$$

$$\mu_{\text{moterų}} = \begin{pmatrix} -8,88 \\ -9,82 \end{pmatrix}, \qquad \Omega_{\text{moterų}} = \begin{pmatrix} 0,10 & 0 \\ 0,02 & 0,08 \end{pmatrix}.$$

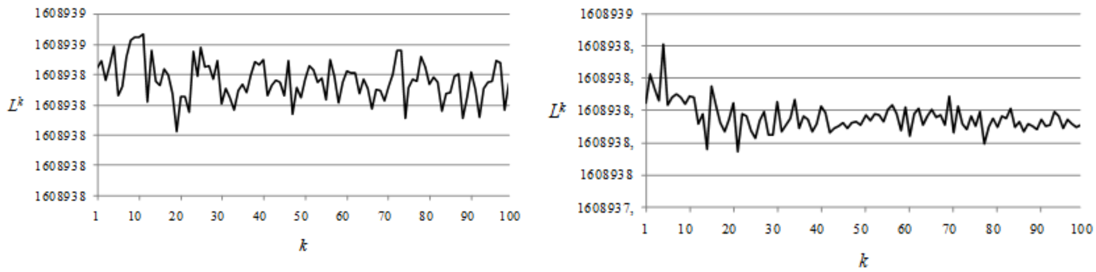
Pasinaudojus aprašytu Puasono-Gauso modelio MCMC algoritmu sugeneruota 100 Markovo Monte-Karlo grandžių ir įvertintos nepalankių įvykių – *suic* ir *hom* tikimybės P_i^m, kai Y_i^m yra *m*-tojo stebinio pasirodymų skaičius, $i = 1, 2, \ldots, 60$ ir $m = 1, 2$.

Gautos modelio charakteristikos pateiktos 3.3 lentelėje. 5 priede pateikti disertacijoje gauti vyrų *suic* ir *hom* tikimybių įverčiai $P_{\text{suic}} \cdot 10^5$ ir $P_{\text{hom}} \cdot 10^5$ ir palyginimui Sakalausko (2010) įverčiai $P_{\text{suic\_1}} \cdot 10^5$ ir $P_{\text{hom\_1}} \cdot 10^5$, gauti neatsižvelgiant į koreliaciją tarp šių įvykių. Įvedus koreliaciją yra gaunama mažesnė tikėtinumo funkcijos reikšmė. Kadangi gauti įverčiai yra skirtingi, galima daryti išvadą, kad koreliaciją įvesti būtina.

3.3 lentelė. Socialinių duomenų Puasono-Gauso modelio įverčiai

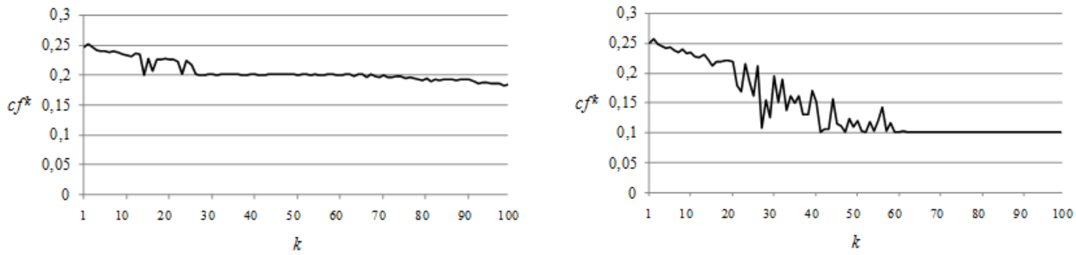
	suic vidurkis	*hom* vidurkis	Koreliacija	Tikėtinumo funkcija
Visi gyventojai	-7,79	-9,26	0,34	3444176
Vyrai	-7,25	-8,80	0,40	1608939
Moterys	-8,89	-9,83	0,24	1835223

3.2–3.6 pav. pavaizduotos priklausomybės nuo generuotų grandžių skaičiaus, gautos tiriant vyrų duomenis, kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}2$ (grafikai (*a*)) ir, kai neviršija $\varepsilon = 0{,}1$ (grafikai (*b*)).



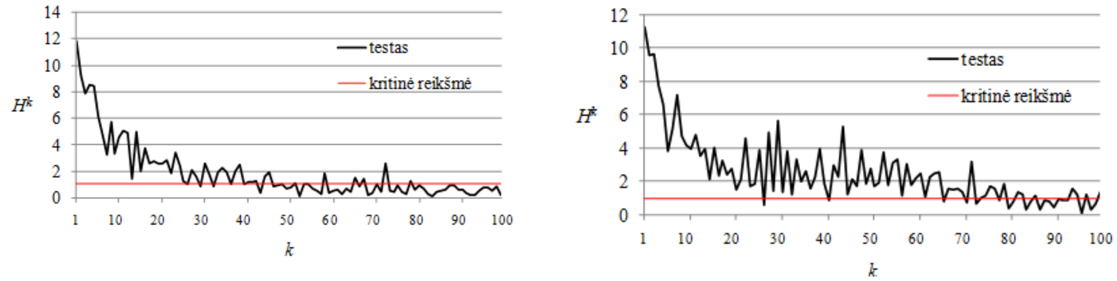
(*a*) (*b*)

3.2 pav. Puasono-Gauso modelio tikėtinumo funkcija L^k



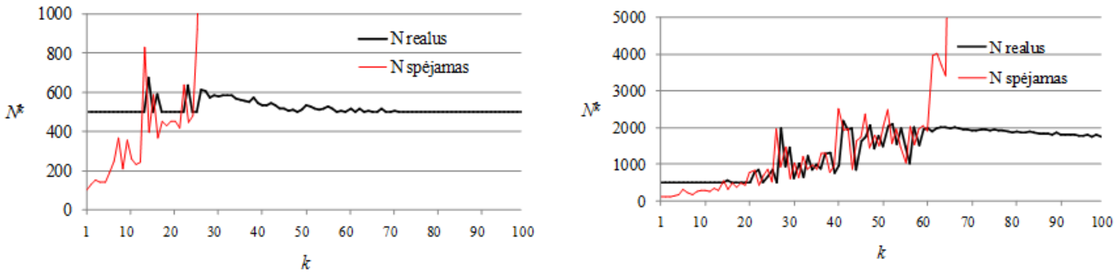
(*a*) (*b*)

3.3 pav. Puasono-Gauso modelio pasikliautinojo intervalo ilgis cf^k



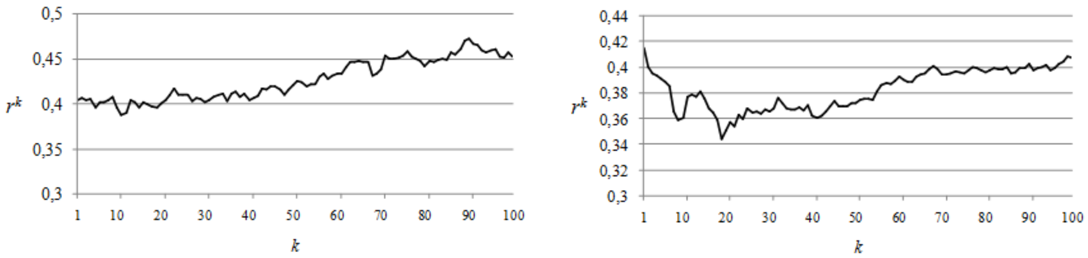
(*a*) (*b*)

3.4 pav. Puasono-Gauso modelio stabdymo statistika H^k



(*a*) (*b*)

3.5 pav. Puasono-Gauso modelio imties tūris N^k



(*a*) (*b*)

3.6 pav. Puasono-Gauso modelio vyrų *suic* ir *hom* koreliacijos įvertis r^k

3.2 pav. pastebima, kad mažiausia tikėtinumo funkcijos reikšmė pasiekta jau pirmose iteracijose. 3.3 pav. pavaizduotas pasikliautinojo intervalo ilgis priklausomai nuo grandžių skaičiaus. Iš pateiktų priklausomybių matyti, jog grandinės pradžioje pasikliautinasis intervalas yra didelis, o po to sumažėja iki reikiamos reikšmės (0,2 arba 0,1). Algoritmo stabdymo kriterijus $\dfrac{H^k}{F_{\delta,p}}$ (3.4 pav. *testas*) mažėja tol, kol pasiekiama nutraukimo kritinė reikšmė 1, čia $F_{\delta,p}$

yra Fišerio skirstinio su $p = 5$ laisvės laipsniais 0,95-kvantilio reikšmė. 3.5 pav. pateikta imties tūrio, apskaičiuoto pagal (3.29) taisyklę (pažymėtas *N spėjamas*), priklausomybė. Šiame paveiksle pavaizduotas realus imties tūris (pažymėtas *N realus*) yra gautas nutraukus imties generavimą taip, kad (3.24) pasikliautinojo intervalo ilgis neviršytų pasirinktos kritinės reikšmės ε. 3.6 pav. pateiktas Puasono-Gauso modelio vyrų *suic* ir *hom* koreliacijos įvertis.

Algoritmo stabdymo sąlygos (*a*) atveju patenkintos po $k = 16$ iteracijų, o jei pasikliautinasis intervalas mažesnis, t. y. (*b*) atveju – po $k = 66$ iteracijų. Disertacijoje atliktas tyrimas siekiant palyginti sukurto MCMC algoritmo, kai imties tūris reguliuojamas, efektyvumą su standartiniu algoritmu, kuriame visose iteracijose imties tūris yra fiksuojamas. 3.4 lentelėje pateikti skaičiavimų rezultatai, gauti kai Monte Karlo imties tūris yra reguliuojamas pagal (3.29) taisyklę ir, kai šis tūris yra fiksuotas. Algoritmų efektyvumas lyginamas pagal imties tūrį reikalingą uždavinio konvergavimo sąlygoms patenkinti.

3.4 lentelė. Puasono-Gauso modelio standartinio ir adaptuoto MCMC algoritmų palyginimas

ε	Imčių tūris	k	Paskutiniosios iteracijos N^k	Bendras N
0,2	reguliuojamas	16	705	10 109
	fiksuotas	23	700	16 100
0,1	reguliuojamas	66	1 976	74 367
	fiksuotas	70	2 000	140 000

Jeigu Monte-Karlo imties tūris nebūtų reguliuojamas pagal (3.29) taisyklę ir pasirinktas pavyzdžiui $N = 2000$, tai konvergavimas būtų pasiektas tik po $k = 70$ iteracijų. Bendras Monte-Karlo imčių tūris, reikalingas sustojimo sąlygoms patenkinti, (*a*) atveju yra 10 109, kai paskutiniojoje grandyje buvo sugeneruota 705, (*b*) atveju tam reikia 74 367, kai paskutiniojoje grandyje buvo sugeneruota 1 976 dydžio imtis. Jei būtų taikytas šiam uždaviniui standartinis MCMC metodas su fiksuota imtimi, tai viso būtų reikėję

$23 \cdot 700 = 16100$ (kitu atveju $70 \cdot 2000 = 140\,000$) Monte-Karlo bandymų uždaviniui išspręsti arba būtų tekę nutraukti skaičiavimus anksčiau ir nepasiekus sprendinių reikiamu tikslumu. Tokiu būdu adaptuotas MCMC metodas leido beveik dvigubai sutrumpinti skaičiavimų apimtį.

3.7. Skyriaus išvados

Disertacijoje sudarytas adaptuotas MCMC algoritmas, skirtas kelių retų įvykių tikimybėms vertinti empiriniu Bajeso metodu. Algoritmas pasižymi statistiniu stabdymo kriterijumi ir imties tūrio reguliavimu. Pasiūlytas pradinių reikšmių parinkimo būdas ir įvesta modifikuota tikėtinumo funkcija, tuo siekiant išvengti jos labai mažų arba labai didelių reikšmių. Taip pat, siekiant išvengti modelio singuliarumo, Monte-Karlo grandys yra sustabdomos, jei *logit* transformacijos kovariacijos matricos apibrėžtumo santykis viršija kritinę reikšmę (disertacijoje pasirinkta kritinė reikšmė lygi 10). Atlikus skaičiavimus, gaunami modelio nežinomų parametrų ir tikimybių atskirose populiacijose įverčiai. Sudarytas algoritmas buvo pritaikytas 2003 m. Lietuvos savižudybių ir nužudymų skaičiaus duomenų analizei. Šio algoritmo palyginimas su standartiniu fiksuotos imties MCMC algoritmu parodė, kad jis leidžia gauti retų įvykių tikimybių Bajeso įverčius reikiamu tikslumu per mažesnį grandžių skaičių. Atliktas eksperimentas parodo, kad adaptuotas MCMC metodas leido beveik dvigubai sutrumpinti skaičiavimus. Sudarytas algoritmas gali būti taikomas ir kitokių socialinių bei medicininių duomenų analizei.

4 skyrius. STABILIOJO SIMETRINIO VEKTORIAUS SKIRSTINIO PARAMETRŲ VERTINIMAS

4.1. Įvadas

Normalusis skirstinys yra vienas plačiausiai naudojamų skirstinių matematinėje statistikoje, dažnai taikomas verslo ir ekonomikos duomenų analizei. Tačiau finansiniai duomenys neretai pasižymi asimetrija, ekscesu ir dideliais nuokrypiais, todėl normalusis skirstinys ne visuomet tinkamas – pavyzdžiui, akcijų grąža ar rizikos faktoriai gali nebūti pasiskirstę pagal normalųjį skirstinį (Kabasinskas ir kt., 2009; Belovas ir kt., 2006). Todėl normalieji skirstiniai keičiami bendresniais – stabiliaisiais modeliais, kurie vis plačiau taikomi verslo, draudimo, finansinių rinkų modeliavimui (Rachev ir Mittnik, 2000; Kabasinskas ir kt., 2012; Sakalauskas ir kt., 2013, 2014). Tačiau jų taikymą praktikoje riboja tai, jog stabiliųjų skirstinių pasiskirstymo bei tankio funkcijos negali būti išreiškiamos per elementariąsias funkcijas, išskyrus keletą atskirų atvejų, pvz., Koši ir Levi dėsniai. Taip pat stabiliųjų skirstinių tankio funkcijos, kai $\alpha = \frac{1}{3}$ ir $\beta = -1$, gali būti išreiškiamos modifikuota Beselio funkcija $K_{\frac{1}{3}}(x)$, o kai $\alpha = \frac{2}{3}$, $\alpha = \frac{3}{2}$ – Vitakerio funkcija $W_{\frac{1}{2},\frac{1}{6}}(x)$ arba hipergeometrines funkcijas (Janicki ir Weron, 1993; Rachev ir Mittnik, 2000; Belovas ir kt., 2006). Beje, stabilieji skirstiniai neturi baigtinės dispersijos (išskyrus normalųjį atvejį, kai stabilumo parametras $\alpha = 2$). Daugiamačių stabiliųjų dėsnių parametrų vertinimas statistiniais metodais ir iškilusios su tuo problemos aptariami šiuose moksliniuose darbuose: Press (1972), Rachev ir Xin (1993), Cheng ir Rachev (1995), Nikias ir Shao (1995), Ravishanker ir Qiou (1999), Nolan (2001), Nolan ir kt. (2001). Šiame skyriuje nagrinėjamas stabiliojo simetrinio vektoriaus, išreiškiamo per atsitiktinį vektorių, pasiskirsčiusį pagal normalųjį skirstinį, ir α -stabiliuosius dydžius, skirstinio parametrų vertinimas didžiausio tikėtinumo metodu, taikant MCMC algoritmą. Šio skyriaus rezultatai buvo publikuoti moksliniuose darbuose (Sakalauskas ir kt., 2013, 2014; Sakalauskas ir Vaičiulytė, 2014; Vaičiulytė ir Sakalauskas, 2014).

Disertacijoje naudojama Zolotariovo stabiliojo skirstinio tankio išraiška (Золотарев, 1983). Atsitiktinis dydis X pasiskirstęs pagal stabilųjį dėsnį $S_\alpha(\sigma,\beta,\mu)$, jei jo charakteringoji funkcija:

$$\psi(x|\alpha,\sigma,\beta,\mu)=\begin{cases}\textbf{(1)}\\ \textbf{(2)}\end{cases}$$

$$\textbf{(1)}\quad \exp\left\{-|\sigma\cdot\theta|^\alpha\cdot\left(1-i\beta\,\mathrm{sgn}(\theta)\tan\left(\frac{\pi\alpha}{2}\right)\right)+i\mu\theta\right\},\ \text{jeigu}\ \alpha\in(0;1)\cup(1;2], \tag{4.1}$$

$$\textbf{(2)}\quad \exp\left\{-|\sigma\cdot\theta|\cdot\left(1+\frac{2i\beta}{\pi}\ln|\theta|\,\mathrm{sgn}(\theta)\right)+i\mu\theta\right\},\quad \text{jeigu}\ \alpha=1,$$

čia $\theta\in\Re$ ir

$$\mathrm{sgn}(\theta)=\begin{cases}1, & \text{jei}\ \theta>0,\\ 0, & \text{jei}\ \theta=0,\\ -1, & \text{jei}\ \theta<0.\end{cases}$$

Vienmatis (4.1) atsitiktinis stabilusis dydis yra aprašomas keturiais parametrais: stabilumo $\alpha\in(0;2]$, asimetrijos $\beta\in[-1;1]$, mastelio $\sigma>0$, poslinkio $\mu\in\Re^1$. Stabilumo indeksas α yra esminis modeliuojant finansines sekas, o mastelio (sklaidos) parametras σ stabiliuoju atveju gali būti taikomas rizikai matuoti. Atsitiktiniai dydžiai, kurie yra stabilūs fiksuoto skaičiaus atsitiktinių dėmenų sudėties atžvilgiu, vadinami α-stabiliaisiais.

Vienmačiu atveju yra žinoma, kad

$$s=s_1\cdot s_2,$$

čia

s_1 – stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta = 1$, o formos parametras $\alpha_1 < 1$;

s_2 – kitas stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta = 0$, o formos parametras α_2;

s – stabilusis atsitiktinis dydis, kurio asimetrijos parametras $\beta = 0$, o formos parametras $\alpha = \alpha_1 \cdot \alpha_2$ (Rachev ir Mittnik, 1993; Ravishanker ir Qiou, 1999).

Taikant šį būdą paprastai parenkama, kad s_2 būtų atsitiktinis dydis, pasiskirstęs pagal normalųjį dėsnį, t. y. $\alpha_1 = \dfrac{\alpha}{2}$ ir $\alpha_2 = 2$. Atsitiktinis stabilusis dydis, kai $\alpha_1 < 1$, o $\beta = 1$, yra vadinamas subordinatoriumi ir jis visada įgyja tik teigiamas reikšmes (Rachev ir Mittnik, 1993; Ravishanker ir Qiou, 1999). Pritaikius šį metodą daugiamačiu atveju gaunamas su priklausomomis komponentėmis atsitiktinis vektorius, kuriuo galima modeliuoti duomenis su didelėmis didelių reikšmių tikimybėmis (Nolan, 2007; Sakalauskas ir Vaičiulytė, 2014).

Taikant šį metodą galima gauti daugiamatį stabilųjį simetrinį vektorių (Rachev ir Mittnik, 1993; Ravishanker ir Qiou, 1999):

$$X = \mu + \sqrt{s_1} \cdot s_2, \tag{4.2}$$

čia

s_1 – subordinatorius su parametru α,

s_2 – atsitiktinis vektorius, pasiskirstęs pagal normalųjį dėsnį $N(0,\Omega)$,

μ – atsitiktinis vidurkio vektorius.

4.2. Didžiausio tikėtinumo metodu gaunami įvertiniai

Disertacijoje nagrinėjamas stabilaus atsitiktinio dydžio apibendrinimas daugiamačiu atveju. (4.2) stabilųjį vektorių apibūdina parametrai: stabilumo indeksas α, vidurkio vektorius μ ir kovariacijų matrica Ω. Šiame skyrelyje nagrinėjamas šių parametrų vertinimas didžiausio tikėtinumo metodu (žr. 1.5.1 skyrelį).

Tegul $X=\left(X^1, X^2, \ldots, X^K\right)$ yra nepriklausomų d-mačių stabiliųjų vektorių imtis. Šios imties tikėtinumo funkcija užrašoma taip (Ravishanker ir Qiou, 1999):

$$\tilde{L}(X,\mu,\Omega,\alpha)=\frac{\left(\dfrac{\alpha}{2-\alpha}\right)^K}{(2\pi)^{\frac{K}{2}}\cdot|\Omega|^{\frac{K}{2}}}\cdot\prod_{i=1}^{K}\int_0^{\infty}\int_1^{1}\exp\left[-\left|\frac{t_\alpha(y_i)}{s_i}\right|^{\frac{\alpha}{2-\alpha}}-\frac{1}{2}\left(X^i-\mu\right)^T\frac{\Omega^{-1}}{s_i}\left(X^i-\mu\right)\right]\times$$

$$\times\left|\frac{t_\alpha(y_i)}{s_i}\right|^{\frac{\alpha}{2-\alpha}}\cdot\frac{1}{s_i^{\frac{d}{2}+1}}\,dy_i ds_i,$$

čia

$$t_\alpha(y)=\frac{\sin\left(\dfrac{\pi\cdot\alpha\cdot y}{2}\right)\cdot\sin\left(\dfrac{\pi\cdot(2-\alpha)\cdot y}{2}\right)^{\frac{2-\alpha}{\alpha}}}{(\sin\pi\cdot y)^{\frac{2}{\alpha}}\cdot\left(\cos\dfrac{\pi\cdot\alpha}{4}\right)^{\frac{2}{\alpha}}}.$$

Įvedus keitinį

$$z_i=\left|\frac{t_\alpha(y_i)}{s_i}\right|^{\frac{\alpha}{2-\alpha}},$$

logaritminė tikėtinumo funkcija yra:

$$L(X,\mu,\Omega,\alpha)=\frac{K}{2}\ln|\Omega|-\sum_{i=1}^{K}\ln\left(\int_0^\infty\int_0^1 B\left(X^i,y_i,z_i,\mu,\Omega,\alpha\right)\cdot\exp\{-z_i\}dy_i dz_i\right),\quad (4.3)$$

čia

$$B\left(X^i,y_i,z_i,\mu,\Omega,\alpha\right)=\exp\left\{-\frac{z_i^{\frac{2-\alpha}{\alpha}}\left(X^i-\mu\right)^T\Omega^{-1}\left(X^i-\mu\right)}{2\cdot t_\alpha(y_i)}\right\}\cdot\frac{z_i^{\frac{2-\alpha}{\alpha}\cdot\frac{d}{2}}}{\sqrt{|\Omega|}\cdot t_\alpha(y_i)^{\frac{d}{2}}}.$$

Norint rasti daugiamačio α-stabilaus dėsnio parametrų μ,Ω įverčius didžiausio tikėtinumo metodu, reikia išspręsti šią lygčių sistemą (žr. 1.5.1 skyrelį):

$$\begin{cases}\dfrac{\partial L(X,\mu,\Omega,\alpha)}{\partial\mu}=-\sum\limits_{i=1}^{K}\dfrac{\partial f\left(X^i,\mu,\Omega,\alpha\right)}{\partial\mu}\cdot\dfrac{1}{f\left(X^i,\mu,\Omega,\alpha\right)}=0,\\ \dfrac{\partial L(X,\mu,\Omega,\alpha)}{\partial\Omega}=-\sum\limits_{i=1}^{K}\dfrac{\partial f\left(X^i,\mu,\Omega,\alpha\right)}{\partial\Omega}\cdot\dfrac{1}{f\left(X^i,\mu,\Omega,\alpha\right)}=0.\end{cases}$$

Pagal fiksuoto taško metodą parametrų įverčiai gaunami iš lygčių:

$$\hat{\mu}=\frac{\sum\limits_{i=1}^{K}\dfrac{X^i\cdot g_i}{f_i}}{\sum\limits_{i=1}^{K}\dfrac{g_i}{f_i}},\quad (4.4)$$

$$\hat{\Omega}=\frac{1}{K}\cdot\sum_{i=1}^{K}\frac{\left(X^i-\hat{\mu}\right)\left(X^i-\hat{\mu}\right)^T g_i}{f_i},\quad (4.5)$$

čia

$$g\left(X,\hat{\mu},\hat{\Omega},\alpha\right)=\int_0^\infty\int_0^1\frac{z^{\frac{2-\alpha}{\alpha}}}{t_\alpha(y)}\cdot B\left(X,y,z,\hat{\mu},\hat{\Omega},\alpha\right)\cdot\exp\{-z\}dydz,\quad (4.6)$$

$$f\left(X,\hat{\mu},\hat{\Omega},\alpha\right)=\int_0^\infty\int_0^1 B\left(X,y,z,\hat{\mu},\hat{\Omega},\alpha\right)\cdot\exp\{-z\}dydz.\quad (4.7)$$

Monte-Karlo metodu apskaičiavus (4.6), (4.7) integralus (žr. 1.6 skyrelį), galima taikyti EM algoritmą (4.4), (4.5) lygtims išspręsti (žr. 1.5 skyrelį).

Vertinant parametrą α galima naudoti tokį pat metodą kaip ir antrame skyriuje: rasti tikėtinumo funkcijos integralinį pavidalą, priklausantį nuo formos parametro, tada diferencijuojant pagal tą parametrą gauti lygtis, kurios analogiškos (2.4) ir (2.6) lygtims. Bet šiame skyriuje remiamasi tuo, jog, esant fiksuotiems μ, Ω, formos parametro įvertį galima gauti, sprendžiant vieno kintamojo minimizavimo uždavinį. Šis būdas detaliau aptariamas kitame skyrelyje.

4.3. Markovo grandinės Monte-Karlo algoritmas

Daugiamačio stabiliojo simetrinio vektoriaus parametrų įvertiniams gauti gali būti pritaikomas EM algoritmas, apskaičiuojant Monte-Karlo metodu (4.3), (4.6), (4.7) integralus, įeinančius į įvertinių išraiškas.

Tegul parinktos kurios nors pradinės reikšmės $\mu^0, \Omega^0, \alpha^0$ ir sugeneruota k Markovo grandžių bei kiekvienoje grandyje apskaičiuoti įverčiai $\mu^k, \Omega^k, \alpha^k$. Tegu

$$\begin{aligned} Y_j &\sim U(0,1), \\ Z_j &\sim -\ln(Y_j), \end{aligned} \tag{4.8}$$

čia $j = 1, 2, \ldots, N^k$, N^k – yra Monte-Karlo imties tūris k-toje grandyje. Apskaičiuojamos sumos:

$$P_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} B\left(X^i, Y_j, Z_j, \mu^k, \Omega^k, \alpha^k\right), \tag{4.9}$$

$$PP_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} \left(B\left(X^i, Y_j, Z_j, \mu^k, \Omega^k, \alpha^k\right)\right)^2, \tag{4.10}$$

$$V_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\frac{Z_j^{\frac{2-\alpha^k}{\alpha^k}}}{t_{\alpha^k}(Y_j)} B\left(X^i, Y_j, Z_j, \mu^k, \Omega^k, \alpha^k\right), \tag{4.11}$$

$$VV_i^k = \frac{1}{N^k}\sum_{j=1}^{N^k}\left(\frac{Z_j^{\frac{2-\alpha^k}{\alpha^k}}}{t_{\alpha^k}(Y_j)} B\left(X^i, Y_j, Z_j, \mu^k, \Omega^k, \alpha^k\right)\right)^2, \tag{4.12}$$

kurios reikalingos įvertiniams apskaičiuoti kitoje iteracijoje pagal (4.4) ir (4.5) bei EM algoritmą:

$$\mu^{k+1} = \frac{\sum_{i=1}^{K} X^i \frac{V_i^k}{P_i^k}}{\sum_{i=1}^{K}\frac{V_i^k}{P_i^k}}, \tag{4.13}$$

$$\Omega^{k+1} = \frac{1}{K}\sum_{i=1}^{K}\left(X^i - \mu^k\right)\left(X^i - \mu^k\right)^T \frac{V_i^k}{P_i^k}, \tag{4.14}$$

Tada logaritminės tikėtinumo funkcijos Monte-Karlo įvertinys yra:

$$L^k = -\sum_{i=1}^{K}\ln\left(P_i^k\right). \tag{4.15}$$

Tikėtinumo funkcijos 95% pasikliautinasis intervalas (žr. (1.21)):

$$\left[L^k - \frac{2}{\sqrt{N^k}}\sqrt{N^k\cdot\sum_{i=1}^{K}\frac{PP_i^k}{\left(P_i^k\right)^2} - K}, L^k + \frac{2}{\sqrt{N^k}}\sqrt{N^k\cdot\sum_{i=1}^{K}\frac{PP_i^k}{\left(P_i^k\right)^2} - K}\right]. \tag{4.16}$$

Parametro α^{k+1} reikšmė gaunama sprendžiant (4.15) tikėtinumo funkcijos vienmačio minimizavimo pagal α uždavinį. Šiam vienmačiam minimizavimo uždaviniui išspręsti buvo taikomas aukso pjūvio metodas (Čiegis, Būda, 1997).

Monte-Karlo Markovo grandinės generuojamos pagal (4.13), (4.14) formules tol, kol pasiekiamas norimo tikslumo (4.16) pasikliautinasis intervalas ir statistinė hipotezė apie (4.4), (4.5) lygybes neatmetama. Pastarajai hipotezei tikrinti naudojamas kriterijus (Anderson, 1958; žr. 1.7.2 skyrelį):

$$H^k = \frac{K}{\frac{1}{K}\sum_{i=1}^{K}\frac{W_i^k}{\left(P_i^k\right)^2}} \times \left(-\ln\frac{\left|\Omega^{k+1}\right|}{\left|\Omega^k\right|} + \left(\mu^{k+1}-\mu^k\right)^T \cdot \left(\Omega^{k+1}\right)^{-1} \cdot \left(\mu^{k+1}-\mu^k\right) + \sum_{i=1}^{d}\left(\Omega^{k+1}\cdot\left(\Omega^k\right)^{-1}\right)_{i,i} - d\right). \tag{4.17}$$

Gerai žinoma, kad imties tūriui didėjant, (4.17) stabdymo kriterijaus skirstinys artėja į χ_p^2 skirstinį su $p = \frac{1}{2}d(d+3)$ laisvės laipsniais (Krishnaiah, 1984). Todėl stabdymo statistinė hipotezė pagal (4.17) kriterijų yra atmetama, jei

$$H^k > \Psi_{\delta,p}, \tag{4.18}$$

čia $\Psi_{\delta,p}$ – yra χ_p^2 skirstinio kvantilis, δ – pasikliovimo lygmuo. Jei ši hipotezė yra neatmetama ir (4.16) pasikliautinojo intervalo ilgis tampa mažesnis už pasirinktą reikšmę ε, $\varepsilon > 0$, tai Monte-Karlo grandinės generavimą galima nutraukti ir priimti paskutinėje iteracijoje gautus įverčius.

Jei nors viena stabdymo sąlyga nėra tenkinama, generuojama nauja Monte-Karlo imtis. Imties tūrį galima reguliuoti pagal taisyklę (Sakalauskas, 2000):

$$N^{k+1} \geq \frac{N^k}{H^k}\cdot\Psi_{\nu,p}, \tag{4.19}$$

čia atskiru atveju ν gali sutapti su δ. Šios taisyklės taikymas leidžia racionaliai parinkti Monte-Karlo imčių tūrį Markovo grandinėje ir kartu užtikrina (4.13) ir (4.14) sekų konvergavimą į (4.3) tikėtinumo funkcijos minimumą su tikimybe 1 (Sakalauskas, 2000).

Nagrinėjamo uždavinio sprendimo algoritmas pateiktas 4.1 lentelėje.

4.1 lentelė. Stabiliojo vektoriaus MCMC metodo tyrimo statistinio modeliavimo algoritmas

Stabiliojo simetrinio vektoriaus skirstinio parametrų vertinimo algoritmas

Tikslas: stabiliojo vektoriaus skirstinio parametrų įvertinimas MCMC metodu.
Įėjimo parametrai: stabilumo indeksas α^0, vidurkio vektorius μ^0 ir kovariacijų matrica Ω^0, stebinių skaičius *K*, pradinis minimalus ir maksimalus Markovo grandies Monte-Karlo imties tūriai N^0, $N_{\min}$, $N_{\max}$, maksimalus grandžių skaičius *k*.
Išėjimo parametrai: logaritminės tikėtinumo funkcijos *L* bei skirstinio parametrų α, μ, Ω įverčių sekos, kiekvienos grandies pasikliautinojo intervalo ilgis *cf*, stabdymo statistika *H* ir imties tūris *N*.
1 žingsnis. Priskiriamos pradinės parametrų reikšmės
2 žingsnis. Apibrėžiama logaritminė tikėtinumo funkcija (4.3)
3 žingsnis. Sugeneruojami atsitiktiniai vektoriai (4.8)
4 žingsnis. Sprendžiamas vienmatis minimizavimo uždavinys parametro α įverčiui ir minimaliai tikėtinumo funkcijos reikšmei rasti.
5 žingsnis. *for* iteracijų skaičius *iter* kinta nuo 1 iki *k*
    6 žingsnis. Ciklo kintamajam *n* priskiriamas 1
    7 žingsnis. (4.16) pasikliautinojo intervalo ilgiui *cf* priskiriamas 1
    8 žingsnis. *while* $n \le N$ (arba $n \le N$ ir $cf > \varepsilon$)
        9 žingsnis. Sugeneruojami atsitiktiniai vektoriai *Y* ir *Z* (4.8)
            10 žingsnis. *for i* kinta nuo 1 iki *K*
                11 žingsnis. Apskaičiuojamos sumos (4.9) – (4.12)
        12 žingsnis. Apskaičiuojamas (4.16) pasikliautinojo intervalo ilgis *cf*
        13 žingsnis. Kintamajam *n* priskiriama *n*+1 reikšmė ir ciklas *while* vėl kartojamas (*8 žingsnis*)

Stabiliojo simetrinio vektoriaus skirstinio parametrų vertinimo algoritmas	
14 žingsnis.	Gaunami parametrų įverčiai (4.13)–(4.14)
15 žingsnis.	Tikrinama stabdymo statistika (4.18)
16 žingsnis.	Įvedamas imties tūrio reguliavimas (4.19)
17 žingsnis.	Pradiniam artiniui priskiriami gauti Monte-Karlo įverčiai
18 žingsnis.	Sudaromas masyvas duomenims išvesti ir toliau tęsiamas iteracijų ciklas (*5 žingsnis*)

4.4. Kompiuterinis modeliavimas

1. Eksperimentas. Modeliuoti duomenys

Algoritmo elgsenos tyrimui buvo atlikti eksperimentai su parinktais pradiniais duomenimis ir realiais duomenimis. Pirmame eksperimente buvo sugeneruota $K = 50$ dvimačių atsitiktinių vektorių su šiais pradiniais duomenimis:

$$d = 2\,,\ \alpha = 1{,}25,\ \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix},\ \Omega = \begin{pmatrix} 3{,}67 & 0{,}86 \\ 0{,}86 & 2{,}55 \end{pmatrix}.$$

Toliau buvo sugeneruota pagal (4.13), (4.14), (4.19) išraiškas $k = 50$ Markovo Monte-Karlo grandžių, paėmus pradines reikšmes:

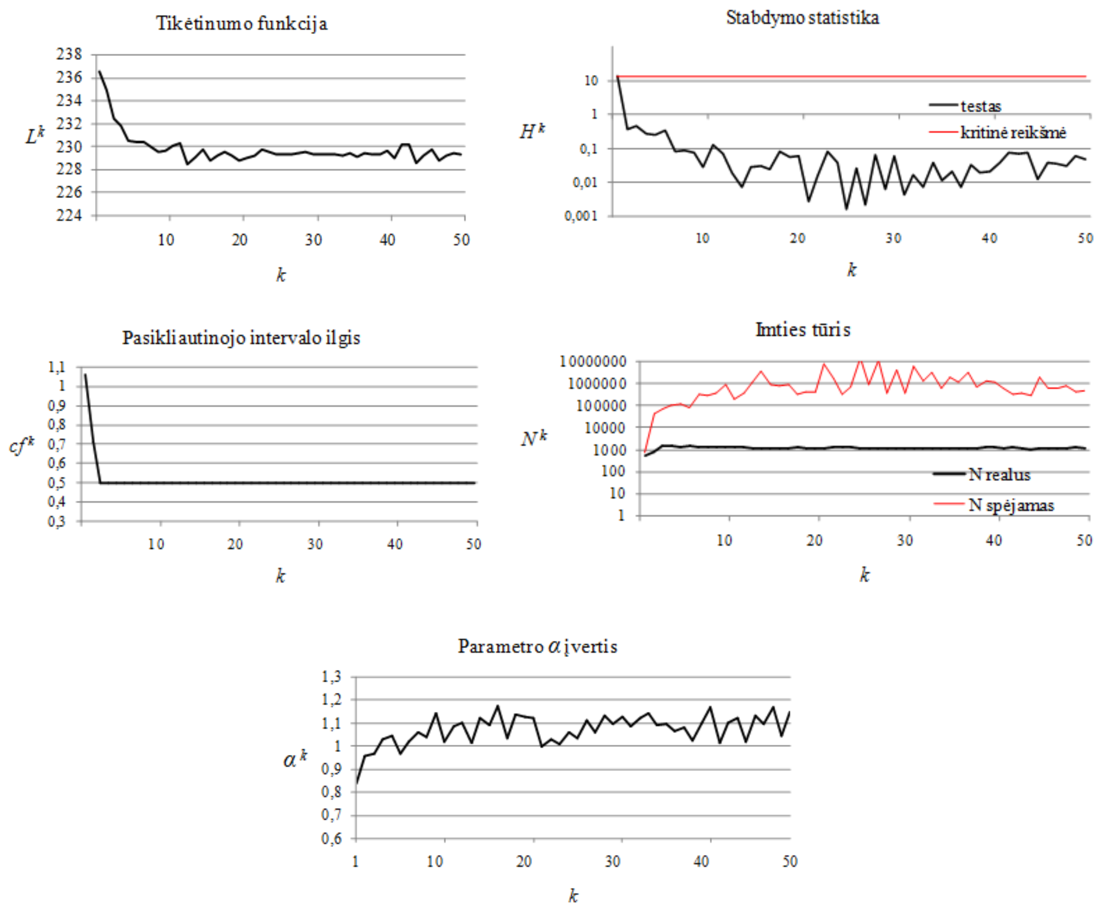
$$\mu^0 = 1{,}1 \cdot \mu, \quad \Omega^0 = 0{,}9 \cdot \Omega.$$

Siekiant išvengti labai mažų arba labai didelių imties tūrio reikšmių, buvo taikoma riba $N^k \geq 500$, ir generavimas buvo nutraukiamas, kai (4.3) tikėtinumo funkcijos (4.16) pasikliautinojo intervalo ilgis tapdavo mažesnis už pasirinktą reikšmę $\varepsilon = 0{,}5$. Stabdymo taisyklės pasiektos po $k = 3$ iteracijų. Gauti šie įverčiai:

$$\alpha = 0{,}97\,, \quad \mu = \begin{pmatrix} 0{,}90 \\ 2{,}32 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 5{,}26 & 2{,}45 \\ 2{,}45 & 3{,}88 \end{pmatrix},$$

$$L = 232{,}42\,, \qquad H = 0{,}46\,.$$

4.1 pav. pavaizduotos priklausomybės nuo generuotų grandžių skaičiaus, kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}5$.



4.1 pav. Skaitinio eksperimento α-stabiliojo dėsnio vertinimo MCMC algoritmu charakteristikos

Stabdymo statistikos *testas* parodo (4.17) stabdymo statistikos kitimo priklausomybę, čia *kritinė reikšmė* yra χ^2_p skirstinio su $p = 5$ laisvės laipsniais 0,999-kvantilio reikšmė. Pateiktame imties tūrio, reguliuojamo pagal (4.19) taisyklę (pažymėtas *N spėjamas*), grafike pavaizduotas *N realus* apskaičiuotas taip, kad (4.16) pasikliautinojo intervalo ilgis neviršytų pasirinktos kritinės reikšmės. Šios priklausomybės iliustruoja sudaryto algoritmo konvergavimą į (4.3) uždavinio sprendinį.

2. Eksperimentas. Trijų įmonių finansiniai duomenys

Disertacijoje sudarytas algoritmas ištestuotas su trijų telekomunikacinių įmonių AT&T, BellSouth ir CenturyLink akcijų duomenimis nuo 2012-01-20 iki 2012-04-01 paimtais iš http://finance.yahoo.com/ (Vaičiulytė ir Sakalauskas, 2014).

Nagrinėjama $K = 50$ trimačių vektorių su šiais pradiniais duomenimis:

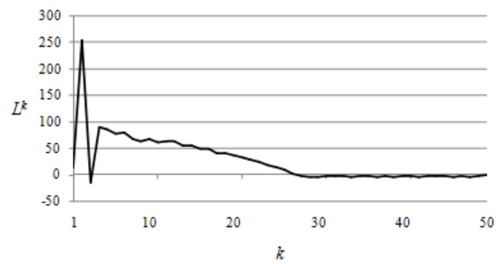
$$d = 3, \quad \alpha = 1{,}25, \quad \mu = \begin{pmatrix} 30{,}609 \\ 25{,}266 \\ 28{,}154 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 0{,}58 & 0{,}089 & 0{,}689 \\ 0{,}089 & 0{,}14 & 0{,}174 \\ 0{,}689 & 0{,}174 & 1{,}246 \end{pmatrix}.$$

Pasinaudojus aprašytu α-stabiliojo dėsnio MCMC algoritmu sugeneruota $k = 50$ Markovo Monte-Karlo grandžių.

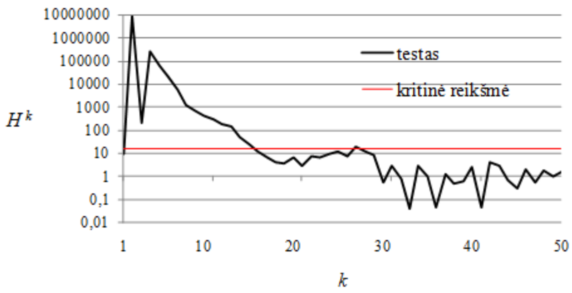
4.2–4.6 pav. pavaizduotos priklausomybės nuo generuotų grandžių skaičiaus, kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}2$. Tikėtinumo funkcija (4.2 pav.) mažėja tol, kol pasiekiama didžiausio tikėtinumo uždavinio sprendinių zona. 4.3 pav. pavaizduota stabdymo statistikos *kritinė reikšmė* yra χ_p^2 skirstinio su $p = 9$ laisvės laipsniais 0,999-kvantilio reikšmė (lygi 27,88). 4.4 pav. *N realus* yra gautas nutraukus imties generavimą taip, kad pasikliautinojo intervalo ilgis neviršytų pasirinktos kritinės reikšmės $\varepsilon = 0{,}2$. *N spėjamas* imties tūris, apskaičiuotas pagal (4.19) taisyklę. Siekiant išvengti labai mažų arba labai didelių imties tūrio reikšmių, buvo taikoma riba $N^k \geq 500$ ir algoritmo stabdymo sąlygos patenkintos po $k = 28$ iteracijų. Gauti šie įverčiai:

$$\alpha = 1{,}538, \qquad \mu = \begin{pmatrix} 30{,}676 \\ 25{,}265 \\ 38{,}154 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0{,}226 & 0{,}037 & 0{,}273 \\ 0{,}037 & 0{,}049 & 0{,}072 \\ 0{,}273 & 0{,}072 & 0{,}449 \end{pmatrix},$$

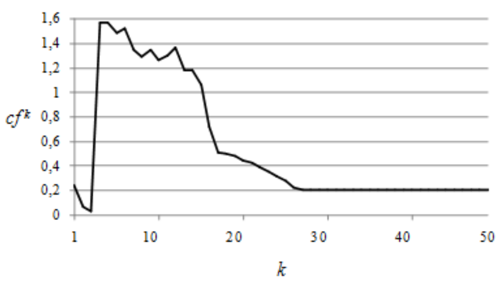
$$L = -2{,}42, \qquad H = 12{,}929844.$$



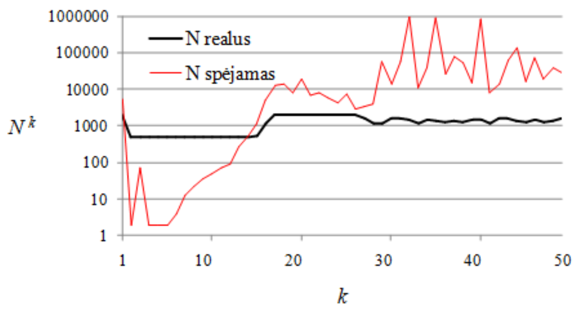
4.2 pav. Tikėtinumo funkcija L^k



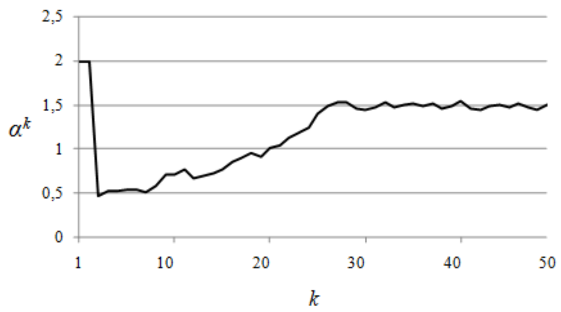
4.3 pav. Stabdymo statistika H^k



4.4 pav. Pasikliautinojo intervalo ilgis cf^k



4.5 pav. Imties tūris N^k



4.6 pav. α-stabiliojo dėsnio parametro α įvertis

4.2 lentelėje pateiktas fiksuotas ir pagal (4.19) taisyklę reguliuojamas Monte Karlo imties tūris reikalingas algoritmo stabdymo sąlygoms patenkinti, kai pasikliautinojo intervalo ilgis lygus $\varepsilon = 0{,}2$ ir $\varepsilon = 0{,}1$. Jeigu Monte-Karlo imties tūris būtų pasirinktas fiksuotas, tai konvergavimas būtų pasiektas po $k = 29$ iteracijų. Bendras adaptuoto MCMC algoritmo Monte-Karlo imčių tūris, reikalingas sustojimo sąlygoms patenkinti, yra 32 346, kai paskutiniojoje grandyje buvo sugeneruota 1 646 dydžio imtis. Taikant šiam uždaviniui standartinį MCMC metodą su fiksuota imtimi, tai iš viso būtų reikėję $29 \cdot 2\,000 = 58\,000$ Monte-Karlo bandymų uždaviniui išspręsti arba būtų tekę nutraukti skaičiavimus anksčiau, t. y. dar nepasiekus sprendinių reikiamu tikslumu.

4.2 lentelė. 3-mačio stabiliojo vektoriaus standartinio ir adaptuoto MCMC algoritmų palyginimas

ε	Imčių tūris	k	Paskutiniosios iteracijos N^k	Bendras N
0,2	reguliuojamas	28	1 646	32 346
	fiksuotas	29	2 000	58 000
0,1	reguliuojamas	28	6 478	88 004
	fiksuotas	30	7 000	210 000

Jeigu Monte-Karlo imties tūris nebūtų reguliuojamas pagal (4.19) taisyklę ir pasirinktas lygus $N = 7\,000$, kai $\varepsilon = 0{,}1$, tai konvergavimas būtų pasiektas po $k = 30$ iteracijų. Tokiu atveju iš viso būtų reikėję $30 \cdot 7\,000 = 210\,000$ Monte-Karlo bandymų uždaviniui išspręsti. Taikant adaptuotą MCMC metodą, bendras Monte-Karlo imčių tūris, reikalingas sustojimo sąlygoms patenkinti, yra 88 004, kai paskutiniojoje grandyje buvo sugeneruota 6 478 dydžio imtis. Tokiu būdu adaptuotas MCMC metodas leido dvigubai sutrumpinti skaičiavimus.

3. Eksperimentas. Penkių įmonių finansiniai duomenys

Analogiškai atliktas tyrimas ir su 5 įmonių akcijų kainomis: AT&T, BellSouth, CenturyLink, CBS ir Sprint (Vaičiulytė, Sakalauskas, 2014).

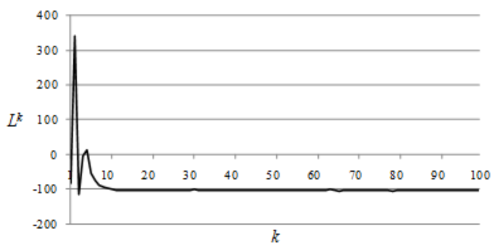
Sudarytu adaptuotu α-stabiliojo dėsnio MCMC algoritmu nagrinėjamos $K = 50$ tūrio vektorių imtys dviem atvejais:

(a) kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}2$;

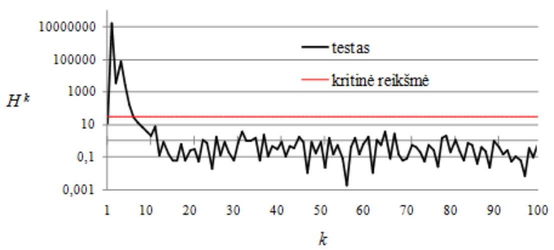
(b) kai pasikliautinojo intervalo ilgis neviršija $\varepsilon = 0{,}1$.

Algoritmo stabdymo sąlygos (*a*) atveju patenkintos po $k = 12$ iteracijų, o jei pasikliautinasis intervalas mažesnis, t. y. (*b*) atveju – po $k = 23$ iteracijų.

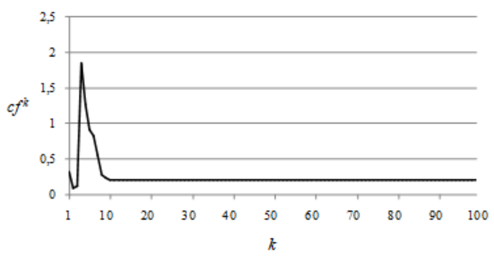
4.7–4.11 pav. pavaizduotos priklausomybės nuo sugeneruotų $k = 100$ Markovo grandžių skaičiaus (*a*) atveju ir imties tūriui taikoma riba $N^k \geq 500$.



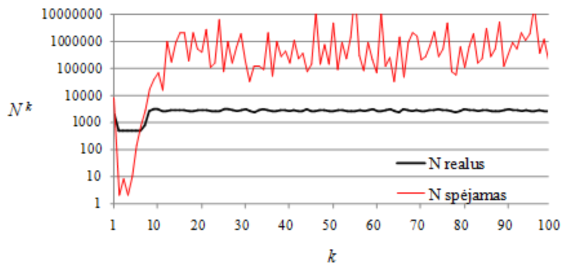
4.7 pav. Tikėtinumo funkcija L^k



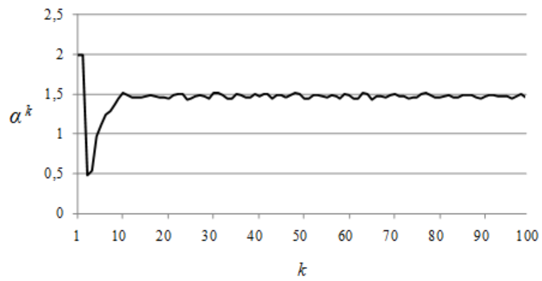
4.8 pav. Stabdymo statistika H^k



4.9 pav. Pasikliautinojo intervalo ilgis cf^k



4.10 pav. Imties tūris N^k



4.11 pav. α-stabiliojo dėsnio parametro α įvertis

4.3 lentelėje pateiktas fiksuotas ir pagal (4.19) taisyklę reguliuojamas Monte Karlo imties tūris reikalingas algoritmo stabdymo sąlygoms patenkinti. Jeigu Monte-Karlo imties tūris nebūtų reguliuojamas pagal (4.19) taisyklę ir pasirinktas lygus $N = 3\,000$, kai $\varepsilon = 0{,}2$, arba $N = 12\,000$, kai $\varepsilon = 0{,}1$, tai konvergavimas būtų pasiektas po $k = 19$ iteracijų. Bendras Monte-Karlo imčių tūris, reikalingas sustojimo sąlygoms patenkinti, (*a*) atveju yra 16 904, kai paskutiniojoje grandyje buvo sugeneruota 2 519, (*b*) atveju tam reikia 156 567, kai paskutiniojoje grandyje buvo sugeneruota 11 724 dydžio imtis. Taikant šiam uždaviniui standartinį MCMC metodą su fiksuota imtimi, tai iš viso būtų reikėję $19 \cdot 3\,000 = 57\,000$ ir kitu atveju $19 \cdot 12\,000 = 228\,000$ Monte-Karlo bandymų uždaviniui išspręsti arba būtų tekę nutraukti skaičiavimus anksčiau, dar nepasiekus sprendinių reikiamu tikslumu. Pastebima, kad adaptuotas MCMC algoritmas leido 1,5–3 kartus sutrumpinti skaičiavimus.

4.3 lentelė. 5-mačio stabiliojo vektoriaus standartinio ir adaptuoto MCMC algoritmų palyginimas

ε	Imčių tūris	k	Paskutiniosios iteracijos N^k	Bendras N
0,2	reguliuojamas	12	2 519	16 904
	fiksuotas	19	3 000	57 000
0,1	reguliuojamas	23	11 724	156 567
	fiksuotas	19	12 000	228 000

4.5. Skyriaus išvados

Skirstinių, susijusių su α-stabiliaisiais, tarp jų tokių kaip daugiamačiais stabiliaisiais simetriniais skirstiniais, tyrimai yra šiuo metu aktualūs, kadangi jie dažnai pasitaiko analizuojant verslo bei finansų duomenis ir informacinius srautus kompiuterių tinkluose. Disertacijoje sudarytas statistinis stabiliųjų simetrinių vektorių adaptuotas MCMC tyrimo algoritmas, kuris leidžia įvertinti šių vektorių skirstinių parametrus, pasinaudojant didžiausio tikėtinumo metodu. α-stabiliojo dėsnio parametrų įvertinimo MCMC algoritmas stabdomas pagal statistinį kriterijų, imties tūris yra reguliuojamas. Sukurtas algoritmas ištirtas telekomunikacinių bendrovių akcijų duomenimis nuo 2012-01-20 iki 2012-04-01. Šio algoritmo palyginimas su standartiniu fiksuotos imties MCMC algoritmu parodė, kad jis leidžia gauti α-stabiliojo dėsnio įvertinius reikiamu tikslumu per mažesnį grandžių skaičių. Atlikto testo atveju adaptuotas MCMC algoritmas leido maždaug du kartus sumažinti skaičiavimo apimtį.

REZULTATAI IR IŠVADOS

1. Disertacijoje sudaryti ir ištirti Markovo grandinės Monte-Karlo adaptavimo metodai.
2. Pasiūlytos taisyklės Monte-Karlo imčių tūrio reguliavimui Markovo grandyse, įvertinių tikslumo vertinimui ir Markovo proceso grandinės stabdymui.
3. Sudarytas algoritmas asimetrinio *t* skirstinio parametrams vertinti adaptuotu MCMC metodu. Parodyta, kad šis metodas realizuoja logaritminės tikėtinumo funkcijos stochastinę gradientinę paiešką, vykdant ją EM algoritmu. Atlikti eksperimentai su Australijos sportininkių duomenimis ir sveikatos industrijai priklausančių įmonių finansiniais duomenimis patvirtino, kad metodo skaitinės savybės atitinka teorinį modelį. Algoritmas gali būti pritaikytas stochastinio pobūdžio sistemų tyrimui MCMC metodu.
4. Sukonstruotas adaptuotas MCMC algoritmas, skirtas kelių retų įvykių tikimybėms vertinti empiriniu Bajeso metodu. Pasiūlytas pradinių reikšmių parinkimo būdas daugiamačiam Puasono-Gauso modeliui. Įvesta modifikuota tikėtinumo funkcija, siekiant išvengti jos labai mažų arba labai didelių reikšmių. Sukurtas modelis pritaikytas socialinių duomenų analizei.
5. Sukonstruotas statistinis stabiliųjų simetrinių vektorių skirstinių parametrų tyrimo adaptuotas MCMC algoritmas. Šis algoritmas pritaikytas telekomunikacinių bendrovių akcijų duomenų modeliui sudaryti. Sudarytas modelis gali būti taikomas akcijų rinkų duomenų analizei.
6. Išnagrinėtos MCMC metodo asimetrinio *t* skirstinio, Puasono-Gauso modelio ir stabiliojo dėsnio parametrų vertinimo skaitmeninimo problemos, kurios yra būdingos daugeliui MCMC metodo taikymo uždavinių. Tokiu būdu gauti rezultatai gali būti sėkmingai pritaikyti ir kitiems statistikos uždaviniams spręsti adaptuotu MCMC metodu.
7. Statistinio modeliavimo būdu ištirtas MCMC algoritmų efektyvumas. Atlikti algoritmų elgsenos tyrimai parodė, kad adaptuotas MCMC algoritmas leidžia

gauti nagrinėjamų skirstinių parametrų įvertinius per mažesnį grandžių skaičių ir maždaug du kartus sumažinti skaičiavimų apimtį.

LITERATŪRA

[1] Aksomaitis, J. (2000). Tikimybių teorija ir statistika. Kaunas: Technologija.

[2] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

[3] Archambeau, C., Lee J. A., Verleysen, M. (2003). On Convergence Problems of the EM Algorithm for Finite Gaussian Mixtures. *Artificial Neural Networks*, 99–106.

[4] Altaleb, A., Chauveau, D. (2002). Bayesian Analysis of the Logit Model and Comparison of Two Metropolis-Hastings Strategies. *Computational Statistics and Data Analysis*, 39, 137–152.

[5] Azzalini, A., Capitanio, A. (2003). Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution. *Journal of the Royal Statistical Society*: Series B (Statistical Methodology), 65, 367–389.

[6] Azzalini, A., Genton, M. G. (2008). Robust Likelihood Methods Based on the Skew-t and Related Distributions. *International Statistical Review*, 76 (1), 106–129.

[7] Bayesian-Inference. Prieiga per internetą [žiūrėta 2014-06-18]: <http://www.bayesian-inference.com/mcmc#algorithms>

[8] Bartkutė, V., Sakalauskas, L. (2007). Simultaneous perturbation stochastic approximation of nonsmooth functions. *European Journal of Operational Research*, 181(3), 1174-1188.

[9] Belovas, I., Kabasinskas, A., Sakalauskas, L. (2006). A Study of Stable Models of Stock Markets. *Information Technology and Control*, 35(1), 34–56. Prieiga per internetą [žiūrėta 2014-06-12]: <http://itc.ktu.lt/itc351/Belov351.pdf>

[10] Bentkus, V., Bloznelis, M., ir Gotze, F. (1996). A Berry-Esseen bound for Student's statistic in the non-i.i.d. case. *Journal Theoretical Probability*, 9(3), 765–796.

[11] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

[12] Besag, J. E. (1989). Towards Bayesian Image Analysis. *Journal of the Applied Statistics*, 16, 395–407.

[13] Boyles, R. A. (1983). On the Convergence of the EM Algorithm. *Journal of the Royal Statistical Society*, 45(B), 47–50.

[14] Bradley, P. C., Thomas, A. L. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall.

[15] Cabral, C. R. B., Bolfarine, H., Pereira, J. R. G. (2008). Bayesian Density Estimation Using Skew Student-t-normal Mixtures. *Computational Statistics and Data Analysis*, 52 (12), 5075–5090.

[16] Carlin, B. P. (1992). A Simple Monte Carlo Approach to Bayesian Graduation. *Transactions of the Society of Actuaries*, 44, 55–76.

[17] Casella, G., George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46(3), 167–174. Prieiga per internetą [žiūrėta 2014-06-12]:
<http://biostat.jhsph.edu/~mmccall/articles/casella_1992.pdf>

[18] Chen, F. (2009). *Bayesian Modeling Using the MCMC Procedure*. Prieiga per internetą [žiūrėta 2014-06-12]:
<http://support.sas.com/resources/papers/proceedings09/257-2009.pdf>

[19] Cheng, B. N., Rachev, S. T. (1995). Multivariate Stable Future Prices. *Mathematical finance*, 5, 133–153.

[20] Clayton, D., Kaldor, J. (1987). Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping. *Biometrics*, 43(3), 671–681.

[21] Cook, R. D., Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.

[22] Čekanavičius, V., Murauskas, G. (2000). *Statistika ir jos taikymai I* Vilnius: TEV.

[23] Čekanavičius, V., Murauskas, G. (2002). *Statistika ir jos taikymai II*. Vilnius: TEV.

[24] Čiegis, R., Būda, V. (1997). *Skaičiuojamoji matematika*. Vilnius: TEV. Prieiga per internetą [žiūrėta 2014-06-17]: <http://uosis.mif.vu.lt/~olgas/SM/SMvadovelisBC.pdf>

[25] Dennis, J. E., Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia: Classics in Applied Mathematics. Prieiga per internetą [žiūrėta 2014-06-12]: <http://books.google.lt/books?id=RtxcWd0eBD0C&pg=PA15&hl=lt&source=gbs_toc_r&cad=3#v=onepage&q&f=false>

[26] Durrett R. (2010). *Probability: Theory and Examples*. New York: Cambridge University Press.

[27] Fishman, G. S. (2003). *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer.

[28] Gamerman, D. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.

[29] Gelfand, A. E., Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410), 398–409. Prieiga per internetą [žiūrėta 2014-06-12]: <http://home.gwu.edu/~stroud/classics/GelfandSmith90.pdf>

[30] Brooks, S. P., Gelman, A. (1998). General Methods for Monotoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7, 434–455.

[31] Geman, S., Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images. *IEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

[32] Gentle, J., E. (2003). *Random Number Generation and Monte Carlo Methods*. New York: Springer.

[33] Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

[34] Hastings, W. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57(1), 97–109. Prieiga per internetą [žiūrėta 2014-06-12]: <http://www.isds.duke.edu/~scs/Courses/Stat376/Papers/Basic/Hastings1970.pdf>

[35] Jakimauskas, G. (2014). *Duomenų tyrybos empirinių Bajeso metodų tyrimas ir taikymas*. *Daktaro disertacija*. Vilnius: Vilniaus universitetas. Prieiga per internetą [žiūrėta 2014-05-20]: <http://www.mii.lt/files/mii_dis_2014_jakimauskas.pdf>

[36] Jakimauskas, G., Sakalauskas, L. (2012). Empirical Bayesian regression model for estimation of small rates. *Lietuvos matematikos rinkinys*, 53A, 42–47.

[37] Janicki, A., Weron, A. (1993). *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*. New York: Marcel Dekker.

[38] Yahoo Finance. Prieiga per internetą [žiūrėta 2014-06-12]: <http://finance.yahoo.com>

[39] Kabašinskas, A. (2007). *Finansinių rinkų statistinė analizė ir statistinio modeliavimo metodai*. *Daktaro disertacija*. Vilnius: Vytauto Didžiojo universitetas. Prieiga per internetą [žiūrėta 2013-03-15]: < http://www.mii.lt/files/disert_08_akabasinskas.pdf>

[40] Kabasinskas, A., Rachev, S., Sakalauskas, L., Sun W., Belovas, I. (2009). Stable Paradigm in Financial Markets. *Journal of Computational Analysis and Applications*, 11(3), 642–688.

[41] Kabasinskas, A., Sakalauskas, L., Sun, E. W., Belovas, I. (2012). Mixed-Stable Models for Analyzing High-Frequency Financial Data. *Journal of Computational Analysis and Applications*, 14(7), 1210–1226.

[42] Kaklauskas, L. (2012). *Fraktalinių procesų kompiuterių tinkluose stebėsenos ir valdymo metodų tyrimas*. *Daktaro disertacija*. Vilnius: Vilniaus universitetas. Prieiga per internetą [žiūrėta 2013-03-15]: < http://www.mii.lt/files/mii_dis_2012_kaklauskas.pdf>

[43] Kaklauskas, L., Sakalauskas, L. (2009) Applications of Chaos Theory to Analysis of Computer Network Traffic. *Proceedings of the XIII Intern. Conference "Applied Stochastic Models and Data Analysis" (ASMDA-2009)*, Vilnius, 407–411.

[44] Kim, H. M., Mallick, B. K. (2003). Moments of Random Vectors with Skew *t* Distribution and their Quadratic Forms. *Statistics & Probability Letters*, 63, 417–423.

[45] Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability* (2nd ed.). New York: Chelsea.

[46] Korolev, V., Shevtsova, I. (2012). An improvement of the Berry-Essen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal*, 2, 81-105.

[47] Krishnaiah, P. R. (1984). *Handbook of Statistics 1: Analysis of Variance*. New York: Elsevier Science & Technology Books.

[48] Kubilius J. (1996). *Tikimybių teorija ir matematinė statistika*. Vilnius.

[49] Li, N., Stephens, M. (2003). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-nucleotide Polymorphism Data. *Genetics*, 165 (4), 2213–2233.

[50] Lin, X., Zhu, Y. (2004). Degenerate Expectation-Maximization Algorithm for Local Dimension Reduction. *Studies in Classification, Data Analysis, and Knowledge Organization*, 46, 259–268.

[51] Lietuvos sveikatos apsaugos ministerijos Higienos institutas. Prieiga per internetą [žiūrėta 2014-06-12]: < http://www.hi.lt/>

[52] Liseo, B., Loperfido, N. (2003). A Bayesian Interpretation of the Multivariate Skew-normal Distribution. *Statistics & Probability Letters*, 61(4), 395–401.

[53] Metropolis, N., Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44, 335–341.

[54] Nikias, C. L., Shao, M. (1995). *Signal Processing with Alpha-Stable Distributions and Applications*. New York: Wiley.

[55] Nolan, J. P. (2007). *Stable Distributions – Models for Heavy Tailed Data*. Boston: Birkhauser.

[56] Nolan, J. P. (2001). Maximum Likelihood Estimation and Diagnostics for Stable Distributions. *Lévy Processes: Theory and Applications*, 379–400.

[57] Nolan, J. P., Panorska, A. K., McCulloch, J. H. (2001). Estimation of Stable Spectral Measures. *Mathematical and Computer Modelling*, 34, 1113–1122.

[58] Owen, A. (2002). *Empirical Likelihood*. New York: Chapman and Hall.

[59] Panagiotelis, A., Smith, M. (2008). Bayesian Density Forecasting of Intraday Electricity Prices Using Multivariate Skew *t* Distributions. *International Journal of Forecasting*, 24, 710–727.

[60] Pearce, D. W. (2006). *Aiškinamasis ekonomikos anglų–lietuvių kalbų žodynas*. Vilnius: TEV.

[61] Petersen, K. B., Pedersen, M. S. (2006). *The Matrix Cookbook*. Prieiga per internetą [žiūrėta 2014-06-18]: <http://www.mit.edu/~wingated/stuff_i_use/matrix_cookbook.pdf>

[62] Polyak, B. T. (1987). *Introduction to Optimization*. New York: Optimization Software.

[63] Press, S. J. (1972). Estimation in Univariate and Multivariate Stable Distributions. *Journal of the American Statistical Association*, 67(340), 842–846.

[64] Rachev, S. T., Mittnik, S. (1993). Modeling Asset Returns with Alternative Stable Distributions. *Econometric Reviews*, 12(3), 261– 330.

[65] Rachev, S. T., Mittnik, S. (2000). *Stable Paretian Models in Finance*. New York: Wiley.

[66] Rachev, S. T., Xin, H. (1993). Test for Association of Random Variables in the Domain of Attraction of Multivariate Stable Law. *Probability and Mathematical Statistics*, 14(1), 125–141.

[67] Ravishanker, N., Qiou, Z. (1999). Monte Carlo EM Estimation for Multivariate Stable Distributions. *Statistics & Probability Letters*, 45(4), 335–340.

[68] Rubinstein, R. Y., Kroese, D. P. (2007). *Simulation and the Monte Carlo Method* (2nd ed.). New York: Wiley.

[69] Sakalauskas, L. (2000). Nonlinear Stochastic Optimization by Monte-Carlo Estimators. *Informatica*, 11(4), 455–468.

[70] Sakalauskas, L. (2002). Nonlinear Stochastic Programming by Monte-Carlo Estimators. *European Journal of Operational Research*, 137(3), 558–573.

[71] Sakalauskas, L. (2010). On the Empirical Bayesian Approach for the Poisson-Gaussian Model. *Methodology and Computing in Applied Probability*, 12(2), 247–259.

[72] Sakalauskas, L. (2013). *Statistinis modeliavimas ir analizė*. Šiauliai: Šiaulių universitetas.

[73] Sakalauskas, L., Kalsyte, Z., Vaiciulyte, I., Kupciunas, I. (2013). The application of stable and skew *t*-distributions in predicting the change in accounting and governance risk ratings. *Proceedings of the 8th International Conference „Electrical and Control Technologies“*, 53–58.

[74] Sakalauskas, L., Kalsyte, Z., Vaiciulyte, I., Kupciunas, I. (2014). The relationship between the transparency in provision of financial data and the change in investors' expectations. *Ekonominė inžinerija – Engineering Economics*.

[75] Sakalauskas, L., Vaiciulyte, I. (2010a). Estimation of skew *t*-distribution by the Monte-Carlo Markov chain approach. *Proceedings of 1th International Conference „Stochastic Modeling Techniques and Data Analysis“*, 747–753.

[76] Sakalauskas, L., Vaiciulyte, I. (2010b). Estimation of skew *t*-distribution by Monte-Carlo Markov chain approach. *Proceedings of 9th International Conference „Computer Data Analysis and Modeling“*, 207–210.

[77] Sakalauskas, L., Vaiciulyte, I. (2011). Estimation of shape parameter of the multivariate *t*-skew distribution. *Proceedings of 14th International Conference „Applied Stochastic Models and Data Analysis"*, 1211–1218.

[78] Sakalauskas, L., Vaiciulyte, I. (2012a). Maximum likelihood estimation of multivariate skew *t*-distribution. *Proceedings of 1st International Conference on Operations Research and Enterprise Systems*, 200–203, Vilamoura: SciTePress.

[79] Sakalauskas, L., Vaiciulyte, I. (2012b). Adaptive Monte-Carlo Markov chain. *Proceedings of 2nd International Conference „Stochastic Modeling Techniques and Data Analysis“*, 653–660.

[80] Sakalauskas, L., Vaičiulytė, I. (2012c). Daugiamatis mažų dažnių vertinimo algoritmas. *Lietuvos matematikos rinkinys, Lietuvos matematikų draugijos darbai*, 53B, 260–263. Prieiga per internetą [žiūrėta 2014-06-11]: <ftp://ftp.science.mii.lt/pub/publications/53_TOMAS%282012%29/Serija_B/INFORMATIKA/Sakalauskas_Vaiciulyte.pdf>

[81] Sakalauskas, L., Vaičiulytė, I. (2013). Multidimensional rare event probability estimation algorithm. *Computational Science and Techniques*, 1(2), 222–228. Prieiga per internetą [žiūrėta 2014-09-28]: <http://journals.ku.lt/index.php/CST/article/view/93>

[82] Sakalauskas, L., Vaičiulytė, I. (2014). Sub-gausinio vektoriaus skirstinio parametrų vertinimas Monte-Karlo Markovo grandinės metodu. *Jaunųjų mokslininkų darbai*, 1(41), 104–107. Prieiga per internetą [žiūrėta 2014-06-30]: <http://www.su.lt/bylos/mokslo_leidiniai/jmd/2014_1_41/sakalauskas.pdf>

[83] Scollnik, D. P. M. (1996). An Introduction to Markov Chain Monte Carlo Methods and Their Actuarial Applications. *Proceedings of the Casualty Actuarial Society*, *LXXXIII*. Prieiga per internetą [žiūrėta 2014-06-12]: <http://www.casact.org/pubs/proceed/proceed96/96114.pdf>

[84] Shevtsova, I. (2010). An improvement of convergence rate estimates in the Lyapunov theorem. *Doklady Mathematics*, 82(3), 862–864.

[85] Smith, A. F. M., Robert, G. O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*, Series B, 55(1), 3–23.

[86] Smyth, GK (2011). *Australasian Data and Story Library* (OzDASL). Prieiga per internetą [žiūrėta 2014-06-12]: <http://www.statsci.org/data/oz/ais.txt>

[87] Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. New York: Wiley.

[88] Stepanauskas, G. (2006). *Markov Chain Monte Carlo*. Lectures for the students of the faculty of Mathematics and Informatics. Vilnius: Vilnius University.

[89] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer.

[90] Tanner, M. A., Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation (with discussion). *Journal of the American Statistical Association*, 82(398), 528–540. Prieiga per internetą [žiūrėta 2014-06-12]: <http://www.stat.cmu.edu/~brian/905-2009/all-papers/tanner-wong-1987-with-disc.pdf>

[91] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4), 1701–1762. Prieiga per internetą [žiūrėta 2014-06-12]: <http://projecteuclid.org/download/pdf_1/euclid.aos/1176325750>

[92] Tsutakawa, R. K., Shoop, G. L., Marienfeld, C. J. (1985). Empirical Bayes Estimation of Cancer Mortality Rates. *Statistics in Medicine*, 4(2), 201–212.

[93] Vaiciulyte, I. (2012). Adaptive Monte-Carlo Markov Chain for Multivariate Statistical Estimation. *Proceedings of International Workshop „Stochastic programming for implementation and advanced applications“*, 119–124. Prieiga per internetą [žiūrėta 2014-06-18]:

<http://www.moksloperiodika.lt/STOPROG_2012/abstract/119-124-Vaic.pdf>

[94] Vaičiulytė, I. (2014). Antisimetrinio *t* skirstinio taikymas tiriant finansinius duomenis. *Jaunųjų mokslininkų darbai*, 1(41), 147–151. Prieiga per internetą [žiūrėta 2014-06-30]: <http://www.su.lt/bylos/mokslo_leidiniai/jmd/2014_1_41/vaiciulyte.pdf>

[95] Vaičiulytė, I., Sakalauskas, L. (2011a). Daugiamačio antisimetrinio *t*-skirstinio parametrų vertinimas Monte-Karlo Markovo grandinių metodu. *Jaunųjų mokslininkų darbai*, 1(30), 137–141. Prieiga per internetą [žiūrėta 2014-06-18]: <http://www.su.lt/bylos/mokslo_leidiniai/jmd/11_01_30/vaiciulyte_sakalauskas.pdf>

[96] Vaičiulytė, I., Sakalauskas, L. (2011b). Daugiamačio antisimetrinio *t*-skirstinio parametrinis įvertinimas. *Jaunųjų mokslininkų darbai*, 4(33), 157–163. Prieiga per internetą [žiūrėta 2014-06-18]: <http://www.su.lt/bylos/mokslo_leidiniai/jmd/11_04_33/vaiciulyte_sakalauskas.pdf>

[97] Vaičiulytė, I., Sakalauskas, L. (2014). Markovo grandinės Monte Karlo metodo taikymas tiriant sociologinius duomenis. *Jaunųjų mokslininkų darbai*, 2(42).

[98] Wang, D., Chen, S. X. (2009). Combining Quantitative Trait Loci Analyses and Microarray Data: an Empirical Likelihood Approach. *Computational Statistics and Data Analysis*, 53, 1661–1673.

[99] Wu, C. F. J. (1983). On Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11, 95–103.

[100] Золотарев, В. М. (1983). *Одномерные устойчивые распределения*. Москва: Наука.

PRIEDAI

1 **Priedas.** Asimetrinio *t* skirstinio tankio funkcija

Pasinaudojant gama funkcija

$$\Gamma(z)=\int_0^{\infty} t^{z-1}\cdot e^{-t}dt$$

asimetrinio *t* skirstinio tankis vienlypiu integralu išreiškiamas taip:

$$p(x|\mu,\Omega,\Theta,\alpha,\eta)=\int_0^{\infty}\int_{\eta\cdot(z-\mu)\geq 0}\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot t^{\frac{\alpha}{2}+d-1}e^{-t\cdot A}dzdt=$$

$$=\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\int_0^{\infty}\int_{\eta\cdot(z-\mu)\geq 0}\frac{1}{A\cdot A^{\frac{\alpha}{2}+d-1}}\cdot(t\cdot A)^{\frac{\alpha}{2}+d-1}\cdot e^{-t\cdot A}dzd(tA)=$$

$$=\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot\Gamma\left(\frac{\alpha}{2}+d\right)\int_{\eta\cdot(z-\mu)\geq 0}\frac{1}{A^{\frac{\alpha}{2}+d}}dz,$$

čia $A=(x-z)^T\cdot\Omega^{-1}\cdot(x-z)+(z-\mu)^T\cdot\Theta^{-1}\cdot(z-\mu)+1$.

Kadangi $\Gamma(z+1)=z\cdot\Gamma(z)$, tai

$$\Gamma\left(\frac{\alpha}{2}+d\right)=\prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\cdot\Gamma\left(\frac{\alpha}{2}\right)$$

ir

$$p(x|\mu,\Omega,\Theta,\alpha,\eta)=\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot\prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\cdot\Gamma\left(\frac{\alpha}{2}\right)\int_{\eta\cdot(z-\mu)\geq 0}\frac{1}{A^{\frac{\alpha}{2}+d}}dz=$$

$$=\int_{\eta\cdot(z-\mu)\geq 0}\frac{\prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot A^{\frac{\alpha}{2}+d}}dz\,.$$

2 Priedas. Asimetrinio *t* skirstinio tankio funkcijos išvestinės parametrų atžvilgiu

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial\mu}=\left(\int_0^\infty\int_{\eta\cdot z\geq 0}\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot t^{\frac{\alpha}{2}+d-1}\times\right.$$

$$\left.\times e^{-t\cdot\left((x-z-\mu)^T\cdot\Omega^{-1}\cdot(x-z-\mu)+z^T\cdot\Theta^{-1}\cdot z+1\right)}d(z+\mu)dt\right)'_\mu=$$

$$=\int_0^\infty\int_{\eta\cdot z\geq 0}\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot t^{\frac{\alpha}{2}+d-1}\times$$

$$\times e^{-t\cdot\left((x-z-\mu)^T\cdot\Omega^{-1}\cdot(x-z-\mu)+z^T\cdot\Theta^{-1}\cdot z+1\right)}\times$$

$$\times\left(2\cdot t\cdot\Omega^{-1}\cdot(x-z-\mu)\right)d(z+\mu)dt=$$

$$=4\cdot\int_0^\infty\int_{\eta\cdot(z-\mu)\geq 0}t\cdot\Omega^{-1}\cdot(x-z)\cdot f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)dzdt;$$

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial\Omega}=-\int_0^\infty\int_{\eta\cdot(z-\mu)\geq 0}\frac{2}{\pi^d\cdot|\Omega|^{\frac{1}{2}}\cdot|\Theta|^{\frac{1}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\cdot t^{\frac{\alpha}{2}+d-1}\times$$

$$\times e^{-t\cdot\left((x-z)^T\cdot\Omega^{-1}\cdot(x-z)+(z-\mu)^T\cdot\Theta^{-1}\cdot(z-\mu)+1\right)}\times$$

$$\times\left(\frac{1}{2}\ln|\Omega|+t\cdot\left((x-z)^T\cdot\Omega^{-1}\cdot(x-z)+(z-\mu)^T\cdot\Theta^{-1}\cdot(z-\mu)+1\right)\right)'_\Omega dzdt=$$

$$=2\cdot\int_0^\infty\int_{\eta\cdot(z-\mu)\geq 0}\left(-\frac{1}{2}\Omega^{-1}+t\cdot\Omega^{-1}\cdot(x-z)\cdot(x-z)^T\cdot\Omega^{-1}\right)\times$$

$$\times f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)dzdt;$$

$$\frac{\partial p(x|\mu,\Omega,\Theta,\alpha,\eta)}{\partial\Theta} = -\int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} \frac{2}{\pi^d \cdot |\Omega|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot \Gamma\left(\frac{\alpha}{2}\right)} \cdot t^{\frac{\alpha}{2}+d-1} \times$$

$$\times e^{-t\cdot\left(\left(x-z\right)^T \cdot \Omega^{-1} \cdot \left(x-z\right) + \left(z-\mu\right)^T \cdot \Theta^{-1} \cdot \left(z-\mu\right)+1\right)} \times$$

$$\times\left(\frac{1}{2}\ln|\Theta| - t\cdot\left(\left(x-z\right)^T \cdot \Omega^{-1} \cdot \left(x-z\right) + \left(z-\mu\right)^T \cdot \Theta^{-1} \cdot \left(z-\mu\right)+1\right)\right)_\Theta^/ dzdt =$$

$$= 2\cdot\int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} \left(-\frac{1}{2}\Theta^{-1} + t\cdot\Theta^{-1}\cdot(z-\mu)\cdot(z-\mu)^T \cdot \Theta^{-1}\right)\times$$

$$\times f(x|z,t,\Omega)\cdot f(z|\mu,t,\Theta)\cdot f_1(t|\alpha)dzdt$$

(Petersen ir Pedersen, 2006). Kadangi

$$\frac{\partial\left(\prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\right)}{\partial\alpha} = \prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\cdot\sum_{i=0}^{d-1}\frac{1}{\alpha+2\cdot i},$$

tai

$$\frac{\partial p\left(x|\mu,\Omega,\Theta,\alpha,\eta\right)}{\partial\alpha} = \int_{\eta\cdot(z-\mu)\geq 0} \frac{2\cdot\prod_{i=0}^{d-1}\left(\frac{\alpha}{2}+i\right)\cdot\left(-\frac{1}{2}\ln(A)+\sum_{i=0}^{d-1}\frac{1}{\alpha+2\cdot i}\right)}{\pi^d \cdot |\Omega|^{\frac{1}{2}} \cdot |\Theta|^{\frac{1}{2}} \cdot A^{\frac{\alpha}{2}+d}} dz =$$

$$= \int_0^\infty \int_{\eta\cdot(z-\mu)\geq 0} \left(-\ln(A) + \sum_{i=0}^{d-1}\frac{1}{\frac{\alpha}{2}+i}\right)\cdot f\left(x|z,t,\Omega\right)\cdot f\left(z|\mu,t,\Theta\right)\cdot f_1\left(t|\alpha\right)dzdt.$$

3 **Priedas.** Sugeneruota dvimačių atsitiktinių vektorių imtis

X_1	1,502	1,029	3,181	1,125	-2,059	2,013	1,811	1,254	0,477	1,052
X_2	0,184	2,114	4,022	3,354	1,051	2,329	4,832	2,214	2,339	3,356

0,528	1,522	0,498	-0,112	3,134	0,158	4,501	1,043	2,346	1,475
1,338	3,252	3,002	3,731	3,164	3,540	-0,368	2,526	5,603	1,489

2,487	1,927	2,707	1,425	2,030	1,460	1,452	1,076	2,330	1,485
1,405	4,982	0,698	2,501	2,575	2,458	2,579	2,978	1,684	1,551

3,251	3,251	0,935	2,839	1,924	2,349	2,682	1,247	0,875	2,238
3,338	3,559	3,426	2,984	2,279	2,691	1,516	3,531	2,660	2,340

1,128	0,904	3,662	0,360	1,647	0,580	0,765	1,222	1,409	1,448
4,500	3,191	4,395	2,179	1,928	1,365	2,870	2,507	1,633	3,421

2,079	0,878	2,357	3,408	2,946	2,040	1,764	0,812	0,825	1,725
2,680	2,498	2,834	3,176	1,951	3,474	2,168	2,199	1,693	-1,401

0,339	1,778	1,202	3,844	0,214	1,130	3,732	2,815	-0,449	2,328
2,335	2,499	2,159	4,583	3,814	2,412	5,489	2,971	2,558	0,182

0,469	1,792	3,566	1,841	-0,317	0,928	0,493	3,470	1,474	2,184
3,421	4,242	3,051	5,823	3,962	4,084	2,629	0,558	2,199	3,582

2,341	1,836	1,533	1,533	1,979	1,774	1,379	1,323	0,978	1,696
3,247	1,901	1,961	1,856	2,349	3,065	1,913	0,594	1,493	2,389

1,273	1,605	0,376	2,047	1,838	0,876	2,006	3,408	1,036	1,350
2,363	2,996	2,594	2,931	2,152	-1,316	2,589	4,913	2,810	3,960

4 Priedas. Australijos sportininkių duomenys

Nr.	Kūno masės indeksas (BMI)	Kūno riebalų masė (Bfat) %	Liesa kūno masė (LBM)	Ūgis (Ht)
1	20,56	19,75	63,32	195,9
2	20,67	21,3	58,55	189,7
3	21,86	19,88	55,36	177,8
4	21,88	23,66	57,18	185
5	18,96	17,64	53,2	184,6
6	21,04	15,58	53,77	174
7	21,69	19,99	60,17	186,2
8	20,62	22,43	48,33	173,8
9	22,64	17,95	54,57	171,4
10	19,44	15,07	53,42	179,9
11	25,75	28,83	68,53	193,4
12	21,2	18,08	61,85	188,7
13	22,03	23,3	48,32	169,1
14	25,44	17,71	66,24	177,9
15	22,63	18,77	57,92	177,5
16	21,86	19,83	56,52	179,6
17	22,27	25,16	54,78	181,3
18	21,27	18,04	56,31	179,7
19	23,47	21,79	62,96	185,2
20	23,19	22,25	56,68	177,3
21	23,17	16,25	62,39	179,3
22	24,54	16,38	63,05	175,3
23	22,96	19,35	56,05	174
24	19,76	19,2	53,65	183,3
25	23,36	17,89	65,45	184,7
26	22,67	12,2	64,62	180,2
27	24,24	23,7	60,05	180,2
28	24,21	24,69	56,48	176
29	20,46	16,58	41,54	156
30	20,81	21,47	52,78	179,7
31	20,17	20,12	52,72	180,9
32	23,06	17,51	61,29	179,5
33	24,4	23,7	59,59	178,9
34	23,97	22,39	61,7	182,1
35	22,62	20,43	62,46	186,3
36	19,16	11,29	53,14	176,8
37	21,15	25,26	47,09	172,6
38	21,4	19,39	53,44	176
39	21,03	19,63	48,78	169,9
40	21,77	23,11	56,05	183
41	21,38	16,86	56,45	178,2
42	21,47	21,32	53,11	177,3
43	24,45	26,57	54,41	174,1
44	22,63	17,93	55,97	173,6
45	22,8	24,97	51,62	173,7
46	23,58	22,62	58,27	178,7
47	20,06	15,01	57,28	183,3
48	23,01	18,14	57,3	174,4
49	24,64	26,78	54,18	173,3
50	18,26	17,22	42,96	168,6
51	24,47	26,5	54,46	174
52	23,99	23,01	57,2	176
53	26,24	30,1	54,38	172,2
54	20,04	13,93	57,58	182,7
55	25,72	26,65	61,46	180,5
56	25,64	35,52	53,46	179,8
57	19,87	15,59	54,11	179,6
58	23,35	19,61	55,35	171,7

Nr.	Kūno masės indeksas (BMI)	Kūno riebalų masė (Bfat) %	Liesa kūno masė (LBM)	Ūgis (Ht)
59	22,42	14,52	55,39	170
60	20,42	11,47	52,23	170
61	22,13	17,71	59,33	180,5
62	25,17	18,48	61,63	173,3
63	23,72	11,22	63,39	173,5
64	21,28	13,61	60,22	181
65	20,87	12,78	55,73	175
66	19	11,85	48,57	170,3
67	22,04	13,35	51,99	165
68	20,12	11,77	51,17	169,8
69	21,35	11,07	57,54	174,1
70	28,57	21,3	68,86	175
71	26,95	20,1	63,04	171,1
72	28,13	24,88	63,03	172,7
73	26,85	19,26	66,85	175,6
74	25,27	19,51	59,89	171,6
75	31,93	23,01	72,98	172,3
76	16,75	8,07	45,23	171,4
77	19,54	11,05	55,06	178
78	20,42	12,39	46,96	162
79	22,76	15,95	53,54	167,3
80	20,12	9,91	47,57	162
81	22,35	16,2	54,63	170,8
82	19,16	9,02	46,31	163
83	20,77	14,26	49,13	166,1
84	19,37	10,48	53,71	176
85	22,37	11,64	53,11	163,9
86	17,54	12,16	46,12	173
87	19,06	10,53	53,41	177
88	20,3	10,15	51,48	168
89	20,15	10,74	53,2	172

Nr.	Kūno masės indeksas (BMI)	Kūno riebalų masė (Bfat) %	Liesa kūno masė (LBM)	Ūgis (Ht)
90	25,36	20,86	56,58	167,9
91	22,12	19,64	56,01	177,5
92	21,25	17,07	46,52	162,5
93	20,53	15,31	51,75	172,5
94	17,06	11,07	42,15	166,7
95	18,29	12,92	48,76	175
96	18,37	8,45	41,93	157,9
97	18,93	10,16	42,95	158,9
98	17,79	12,55	38,3	156,9
99	17,05	9,1	34,36	148,9
100	20,31	13,46	39,03	149

5 **Priedas.** 2003 m. savižudybių ir nužudymų tyrimai Lietuvoje

Savivaldybės	Vyrų sk.	*suic*	*hom*	$P_{suic} \cdot 10^5$	$P_{hom} \cdot 10^5$	$P_{suic\_1} \cdot 10^5$	$P_{hom\_1} \cdot 10^5$
Akmenės r.	13920	10	3	71,13	14,96	79,32	14,99
Alytaus r.	15741	12	1	71,31	15,10	81,06	14,20
Alytaus m.	34085	17	3	72,91	15,28	61,65	14,00
Anykščių r.	16099	19	4	70,01	14,83	103,87	15,23
Birštono	2467	4	0	71,07	14,96	99,45	14,53
Biržų r.	16462	12	5	73,24	15,29	79,20	15,58
Druskininkų	11543	8	1	70,69	14,94	78,98	14,41
Jonavos r.	24587	16	6	70,34	15,14	73,09	15,48
Joniškio r.	14815	15	2	71,02	15,22	94,10	14,59
Jurbarko r.	17627	20	2	70,71	15,07	101,85	14,45
Kalvarijų r.	6556	9	2	71,07	15,13	103,99	15,04
Kauno r.	39525	18	7	72,89	15,36	57,70	15,03
Kauno m.	167308	85	27	72,40	15,02	54,08	15,19
Kazlų Rūdos	7037	2	1	71,22	15,01	68,86	14,65
Kėdainių r.	30590	31	5	69,33	14,77	96,57	14,82
Kelmės r.	19334	20	2	70,37	14,86	96,32	14,37
Klaipėdos m.	88308	45	16	70,48	15,40	56,58	15,49
Neringos	1212	1	0	71,25	15,04	87,42	14,60
Klaipėdos r.	22598	17	3	71,01	14,82	79,47	14,54
Palangos m.	8062	8	0	70,70	15,02	91,42	14,24
Kretingos r.	21776	26	1	71,72	14,81	107,07	13,92
Kupiškio r.	11366	11	1	71,44	15,01	91,17	14,42
Lazdijų r.	12788	14	4	72,68	15,33	97,73	15,42
Marijampolės	33536	30	6	71,60	14,99	88,33	15,00
Mažeikių r	31594	37	6	70,02	15,02	108,12	15,10
Molėtų r.	11899	15	2	69,65	14,77	105,57	14,74
Pagėgių	5770	10	1	70,73	14,92	114,79	14,71
Pakruojo r.	13855	11	2	72,02	15,26	82,95	14,64
Panevėžio r.	20587	24	1	70,77	15,16	104,83	13,97
Pasvalio r.	16365	13	4	71,21	15,17	82,62	15,22

Savivaldybės	Vyrų sk.	*suic*	*hom*	$P_{suic} \cdot 10^5$	$P_{hom} \cdot 10^5$	$P_{suic\_1} \cdot 10^5$	$P_{hom\_1} \cdot 10^5$
Plungės r.	20967	18	3	71,43	15,01	85,91	14,62
Prienų r.	16714	15	0	72,45	15,46	88,13	13,82
Radviliškio r.	24469	25	3	70,77	14,95	96,40	14,45
Raseinių r.	20562	21	4	69,48	14,81	95,75	14,99
Rokiškio r.	19442	21	7	70,76	14,82	99,13	16,14
Šakių r.	18271	17	1	73,33	15,37	90,05	14,08
Šalčininkų r.	18534	15	3	71,80	14,98	83,20	14,75
Šilalės r.	15183	15	2	71,23	15,19	92,87	14,57
Šilutės r.	26257	22	1	70,73	14,94	84,51	13,72
Širvintų r.	9392	10	3	70,53	14,85	94,82	15,24
Skuodo r.	12099	11	1	71,46	15,05	88,54	14,38
Švenčionių r.	15174	7	1	69,05	14,70	66,78	14,23
Tauragės r.	24604	19	3	70,14	15,23	80,46	14,44
Telšių r.	26855	22	7	70,71	14,94	83,28	15,71
Trakų r.	17400	14	3	71,84	15,01	83,05	14,81
Ukmergės r.	22335	23	3	72,43	15,22	96,59	14,56
Utenos r.	23205	29	0	69,96	15,42	111,32	13,54
Varėnos r.	14611	25	1	69,74	14,91	134,16	14,26
Vilniaus r.	43452	23	8	74,27	15,33	61,96	15,16
Vilniaus m.	246412	110	22	72,63	15,68	47,50	12,03
Šiaulių m.	60221	43	14	66,89	14,39	74,21	16,26
Šiaulių r.	24563	20	3	71,74	15,28	83,07	14,45
Rietavo r.	5044	13	1	70,38	14,89	141,06	14,75
Zarasų r.	10498	6	1	71,28	15,32	74,57	14,46
Visagino	13653	4	2	73,86	15,46	60,78	14,65
Ignalinos r.	10578	14	0	70,28	15,03	107,42	14,11
Kaišiadorių r.	18366	13	1	70,45	14,83	77,66	14,08
Elektrėnų r.	13644	17	4	70,42	14,80	106,13	15,37
Panevėžio m.	53913	26	9	69,39	14,92	57,25	14,98
Vilkaviškio r.	23786	21	3	69,95	14,77	87,40	14,48