



J U R G I T A A R N A S T A U S K A I T Ė

**SUDERINAMUMO SU
NORMALIUOJU SKIRSTINIŲ
HIPOTEZIŲ TIKRINIMO
KRITERIJŲ KŪRIMAS IR
GALINGUMO LYGINAMASIS
TYRIMAS TAIKANT MONTE
KARLO METODĄ**

D A K T A R O D I S E R T A C I J A

K a u n a s
2 0 2 2

KAUNO TECHNOLOGIJOS UNIVERSITETAS

JURGITA ARNASTAUSKAITĖ

SUDERINAMUMO SU NORMALIUOJU
SKIRSTINIŲ HIPOTEZIŲ TIKRINIMO
KRITERIJŲ KŪRIMAS IR GALINGUMO
LYGINAMASIS TYRIMAS TAIKANT MONTE
KARLO METODĄ

Daktaro disertacija
Gamtos mokslai, informatika (N 009)

Kaunas, 2022

Disertacija rengta 2018–2022 metais Kauno technologijos universiteto Matematikos ir gamtos mokslų fakulteto Taikomosios matematikos katedroje. Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Doktorantūros teisė Kauno technologijos universitetui suteikta kartu su Vytauto Didžiojo ir Vilniaus Gedimino technikos universitetais.

Mokslinis vadovas:

Doc. dr. Tomas RUZGAS (Kauno technologijos universitetas, fiziniai mokslai, informatika, N 009).

Redagavo: anglų kalbos redaktorė Dovilė Blaudžiūnienė (leidykla „Technologija“), lietuvių kalbos redaktorė Rozita Znamenskaitė (leidykla „Technologija“)

Informatikos mokslo krypties disertacijos gynimo taryba:

Prof. habil. dr. Rimantas BARAUSKAS (Kauno technologijos universitetas, gamtos mokslai, informatika, N 009) – **pirmininkas**;

Dr. Jolita BERNATAVIČIENĖ (Vilniaus universitetas, gamtos mokslai, informatika, N 009);

Prof. dr. Gintaras PALUBECKIS (Kauno technologijos universitetas, gamtos mokslai, informatika, N 009);

Prof. habil. dr. Minvydas Kazys RAGULSKIS (Kauno technologijos universitetas, gamtos mokslai, informatika, N 009);

Dr. Renata RETKUTĖ (Kembridžo universitetas, Anglija, gamtos mokslai, matematika, N 001).

Disertacija bus ginama viešame Informatikos mokslo krypties disertacijos gynimo tarybos posėdyje 2022 m. spalio 14 d. 9 val. Kauno technologijos universiteto „Santakos“ slėnyje, Posėdžių kambaryje.

Adresas: K. Baršausko g. 59 – A228, Kaunas, LT- 51423, Lietuva.

Tel. (370) 37 300 625; el. paštas doktorantura@ktu.lt

Disertacija išsiųsta 2022 m. rugsėjo 14 d.

Su disertacija galima susipažinti interneto svetainėje <http://ktu.edu>, Kauno technologijos universiteto bibliotekoje (K. Donelaičio g. 20, 44239 Kaunas), Vytauto Didžiojo universiteto bibliotekoje (K. Donelaičio g. 52, 44244, Kaunas) ir Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, 10223, Vilnius).

© J. Arnastauskaitė, 2022

KAUNAS UNIVERSITY OF TECHNOLOGY

JURGITA ARNASTAUSKAITĖ

**DEVELOPMENT AND COMPARATIVE
STUDY USING THE MONTE CARLO
METHOD OF GOODNESS-OF-FIT TESTS
FOR NORMALITY**

Summary of Doctoral Dissertation
Natural Sciences, Informatics (N 009)

Kaunas, 2022

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Mathematics and Natural Sciences, Department of Applied Mathematics, during the period of 2018–2022. The studies were supported by the Research Council of Lithuania.

The doctoral right has been granted to Kaunas University of Technology together with Vytautas Magnus University and Vilnius Gediminas Technical University.

Scientific Supervisor:

Assoc. Prof. Dr. Tomas RUZGAS (Kaunas University of Technology, Natural Sciences, Informatics, N 009).

Edited by: English language editor Dovilė Blaudžiūnienė (Publishing House *Technologija*), Lithuanian language editor Rozita Znamenskaitė (Publishing House *Technologija*)

Dissertation Defence Board of Informatics Science Field:

Prof. Dr. habil. Rimantas BARAUSKAS (Kaunas University of Technology, Natural Sciences, Informatics, N 009) – **chairman**;

Dr. Jolita BERNATAVIČIENĖ (Vilnius University, Natural Sciences, Informatics, N 009);

Prof. Dr. Gintaras PALUBECKIS (Kaunas University of Technology, Natural Sciences, Informatics, N 009);

Prof. Dr. habil. Minvydas Kazys RAGULSKIS (Kaunas University of Technology, Natural Sciences, Informatics, N 009);

Dr. Renata RETKUTĖ (University of Cambridge, England, Natural Sciences, Mathematics, N 001).

The official defence of the dissertation will be held at 9 a.m. on 14 October, 2022 at the public meeting of Dissertation Defence Board of Informatics Science Field in Meeting room at Santaka Valley of Kaunas University of Technology (K. Baršausko str. 59 – A228, Kaunas).

Address: K. Baršausko str. 59 – A228, Kaunas, LT- 51423, Lithuania.

Tel. no. (+370) 37 300 625; e-mail doktorantura@ktu.lt

The doctoral dissertation was sent on 14 September, 2022.

The doctoral dissertation is available on the internet <http://ktu.edu>, at the library of Kaunas University of Technology (K. Donelaičio St. 20, 44239 Kaunas, Lithuania), at the library of Vytautas Magnus university (K. Donelaičio St. 52, 44244, Kaunas) and at the library of Vilnius Gediminas Technical University (Saulėtekio al. 14, 10223, Vilnius).

© J. Arnastauskaitė, 2022

TURINYS

Paveikslų sąrašas	7
Lentelių sąrašas	8
SIMBOLIŲ SĄRAŠAS.....	9
ŽYMĖJIMŲ SĄRAŠAS	11
ĮVADAS.....	12
I. MOKSLINĖS LITERATŪROS APŽVALGA IR ANALIZĖ	15
1.1. Suderinamumo hipotezė ir jos kriterijų lyginamoji apžvalga.....	15
1.2. Vienmačių kriterijų tyrimų apžvalga ir analizė	17
1.3. Daugiamačių kriterijų tyrimų apžvalga ir analizė.....	26
1.4. Literatūros apžvalgos bei analizės išvados ir apibendrinimas	33
II. METODOLOGIJA.....	35
2.1. Vienmačiai suderinamumo hipotezės tikrinimo kriterijai.....	35
2.1.1 Vienmačiai kriterijai, paremti empirine pasiskirstymo funkcija.....	35
2.1.2 Vienmačiai kriterijai, paremti regresija ir koreliacija (rangine statistika).....	37
2.1.3 Momentais paremti vienmačiai kriterijai	40
2.1.4 Kiti vienmačiai kriterijai.....	46
2.1.5 Naujas vienmatis kriterijus ($N1$).....	49
2.2. Daugiamačiai suderinamumo hipotezės tikrinimo kriterijai.....	51
2.2.1. Daugiamačiai kriterijai, paremti regresija ir koreliacija (rangine statistika).....	52
2.2.2. Momentais paremti daugiamačiai kriterijai	54
2.2.3. Kiti daugiamačiai kriterijai	56
2.2.4. Naujas daugiamatis kriterijus ($N2$)	57
2.3. Modeliavimo tyrimui taikyti statistiniai skirstiniai.....	63
2.3.1. Modeliavimo tyrimui taikyti vienmačiai skirstiniai.....	64
2.3.2. Modeliavimo tyrimui taikyti daugiamačiai skirstiniai	65
2.4. Kriterijų galios skaičiavimo procedūra.....	66
2.5. Monte Karlo metodo taikymo procedūra.....	69
2.6. Paskirstytųjų skaičiavimų taikymas modeliavimo tyrimui	72
2.7. Metodinės dalies išvados ir apibendrinimas	74
III. TYRIMŲ REZULTATAI IR JŲ APTARIMAS	75

3.1. Vienmačių suderinamumo hipotezės tikrinimo kriterijų galios lyginamoji analizė	75
3.2. Daugiamačių suderinamumo hipotezės tikrinimo kriterijų galios lyginamoji analizė	82
3.3. Praktinis pritaikymas realiems duomenims	90
3.3.1. Prostatos specifinio antigeno rodiklių normalumo prielaidos tikrinimas	90
3.3.2. IQOS ir tradicinių cigarečių išmetamų kietųjų dalelių koncentracijos normalumo prielaidos tikrinimas	92
3.4. Darbo rezultatus apibendrinančios išvados.....	94
IV. BENDROSIOS IŠVADOS IR REKOMENDACIJOS	95
VI. SUMMARY	96
LITERATŪROS ŠALTINIAI	126
AUTORĖS GYVENIMO APRAŠYMAS.....	134
AUTORĖS MOKSLINIŲ PUBLIKACIJŲ SĄRAŠAS.....	135
PADĖKA.....	137
PRIEDAI	138
1 priedas. Darbe naudoti skirstiniai.....	138
2 priedas. Naujo daugiamačio kriterijaus parametrų rinkiniai	141
3 priedas. Daugiamačių kriterijų galios palyginimo grafikai pagal skirstinį.....	143
4 priedas. IQOS ir tradicinių cigarečių išmetamų kietųjų dalelių eksperimentų sąlygos.....	149

Paveikslų sąrašas

1.1 pav. Kriterijų galios palyginimas $0,5N(0,1) + 0,5N(3,1)$ ir $0,5Exp(1) + 0,5Exp(0,5)$ skirstinių atvejais.....	19
1.2 pav. Kriterijų galios palyginimas mažėjančios (a), laiptinės (b), trikampio (c), bukos (d), smailios (e) formų ir bimodaliniams (f) skirstiniams	21
1.3 pav. Kriterijų palyginimas kiekvienam nagrinėtam skirstinių mišinio atvejui	24
1.4 pav. Kriterijų galios palyginimas beta, Laplaso ir gama skirstinių atvejais.....	26
1.5 pav. T_n, h galios palyginimas ($d = 5$) tolygiojo 0,1 ir normaliojo skirstinio mišinio atvejais.....	28
1.6 pav. Kriterijų galios palyginimas ($d = 5, n = 50$) standartinio normaliojo skirstinio atveju	29
1.7 pav. Aštuonių daugiamačių kriterijų empirinės galios palyginimas	30
1.8 pav. Kriterijų palyginimas pagal I rūšies klaidos tikimybę, kai $n = 50$	31
1.9 pav. Kriterijų palyginimas ($d = 5, n = 50$).....	32
1.10 pav. Kriterijų palyginimas 5-mačiams ir 200-mačiams duomenims.....	33
2.1 pav. Pasiūlytos branduolio funkcijos forma su keletu nagrinėtų parametrų rinkinių	51
2.2 pav. Kriterijų galios įvertinimo schema	68
2.3 pav. Statistikos, paremtos laiptine funkcija, reikšmingumo lygmenys	69
2.4 pav. Kriterijaus galios įvertinimo Monte Karlo metodu schema	72
2.5 pav. SAS tinklo skaičiavimo procedūros veikimo schema	73
2.6 pav. Paskirstytojo modeliavimo proceso schema.....	74
3.1 pav. Penkių galingiausių kriterijų galios kitimas simetriniams skirstiniams ...	81
3.2 pav. Penkių galingiausių kriterijų galios kitimas asimetriniams skirstiniams..	82
3.3 pav. Penkių galingiausių kriterijų galios kitimas modifikuotiems normaliesiems skirstiniams.....	82
3.4 pav. Glodumo parametro kitimo priklausomybė nuo triukšmo klasterio svorio ir dimensijos	83
3.5 pav. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia visų tipų skirstinių grupėms.....	89
3.6 pav. Prostatos specifinio antigeno rodiklių duomenų rinkinio struktūra.....	91
3.7 pav. Eksperimentų patalpos schema	93

Lentelių sąrašas

1.1 lentelė. Suderinamumo kriterijų statistikų galios palyginimas	21
1.2 lentelė. Kriterijų rangų palyginimas pagal tipą	23
2.1 lentelė. Statistikos $T * \text{Johnson}$ S_U pasiskirstymo parametrų įverčiai	63
3.1 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis simetrinių skirstinių grupei su skirtingais imties dydžiais.....	76
3.2 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis asimetrinių skirstinių grupei su skirtingais imties dydžiais.....	77
3.3 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis modifikuotų normaliųjų skirstinių grupei su skirtingais imties dydžiais	78
3.4 lentelė. Minimalus imties dydis, kai NI kriterijus tampa galingiausias.....	79
3.5 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis simetrinių skirstinių grupei	84
3.6 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis asimetrinių skirstinių grupei	85
3.7 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis mišriųjų skirstinių grupei	86
3.8 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis normaliojo skirstinio mišinių grupei.....	87

SIMBOLIŲ SĄRAŠAS

\mathbb{R}^d – d -matė realiųjų skaičių aibė.

$X_k = (X_1, X_2, \dots, X_d)^T \in \mathbb{R}^d$ – d -matis vektorius.

d – dimensija.

n – imties dydis.

\bar{X} – imties vidurkis.

$\hat{\sigma}^2$ – imties dispersija.

$\hat{\sigma}$ – imties standartinis nuokrypis.

ω – imties mediana.

$M_{(i)}$ – transformuotoji mediana.

Σ – kovariacinė matrica.

R – koreliacinė matrica.

r_{ij} – koreliacija tarp i -ojo ir j -ojo stebinių.

α – reikšmingumo lygmuo (I rūšies klaidos tikimybė).

β – II rūšies klaidos tikimybė.

w – svorio parametras.

h – glodinimo parametras.

s – asimetrijos koeficientas.

k – eksceso koeficientas.

m_i – i -tasis centrinis momentas.

Z_i – standartinis normalusis atsitiktinis dydis.

η_1 ir η_3 – regresijos lygties parametrai.

q_i – normaliojo skirstinio kvantilis.

J_n – robustinis sklaidos matas.

$\hat{\Sigma}$ – pastovusis kovariacinės matricos $\{\hat{g}_i\}$ įvertinys.

$h_l(\cdot)$ – normalizuotasis l laipsnio polinomas.

\hat{S}_b^2 – dvisvoris sklaidos įvertis.

$K(x)$ – branduolio funkcija.

$\Phi(\cdot)$ – standartinio normaliojo pasiskirstymo funkcija.

$\Phi^{-1}(\cdot)$ – atvirkštinė standartinio normaliojo pasiskirstymo funkcija.

ϕ – standartinio normaliojo skirstinio tikimybės tankio funkcija.

$\Gamma(\cdot)$ – gama pasiskirstymo funkcija.

$\hat{F}^{(k)}$ – auto-kovariacinė funkcija.

$\#T$ – aibės T elementų skaičius.

V_i – standartizuotųjų dažnių vektorius.

N_{in} – atsitiktinių vektorių skaičius.

$J = J(\theta)$ – Fišerio informacijos matrica.

Q – svorio parametrų $d(d+1)/2 \times d(d+1)/2$ dydžio kovariacinė matrica.

$\hat{P}(t)$ – daugiamačio standartinio normaliojo skirstinio charakteristinė funkcija.

$\hat{Q}(t)$ – standartizuotų stebėjimų empirinė charakteristinė funkcija.

$\varphi_\beta(t)$ – branduolio svorinė funkcija.

D_{ij} – Mahalanobio atstumas tarp i -ojo ir j -ojo stebinių.

w_j – normalizavimo transformacija.

$\mu_{i,\tau}$ – suprojektuotų duomenų i -ojo komponento vidurkis.

$\Sigma_{i,\tau}^2$ – suprojektuotų duomenų i -ojo komponento dispersija.

$f_i(X, \theta_i)$ – i -ojo tipo pasiskirstymo funkcija.

$\psi(t)$ – atsitiktinių reikšmių charakteristinė funkcija.

p_i ($i = 1, \dots, q$) – apriorinė tikimybė.

\mathcal{R} – apskritimo spindulys (apribotas sfera).

δ ir γ – formos parametrai.

λ – sklaidos parametras.

ξ – padėties parametras.

ŽYMEJIMŲ SĄRAŠAS

KS – Kolmogorovo ir Smirnovo kriterijus.
AD – Andersono ir Darlingo kriterijus.
CVM – Kramerio ir von Mises kriterijus.
LF – Lillieforso kriterijus.
GLB – Gleno, Leemio ir Barro kriterijus.
ZW1 ir ZW2 – Zhang ir Wu kriterijai.
SW – Šapiro ir Vilko kriterijus.
DA – D'Agostino kriterijus.
SF – Šapiro ir Francijos kriterijus.
Filli – Fillibeno kriterijus.
ChenS – Chen ir Šapiro kriterijus.
SWRG – Rahmano ir Govindarajulo kriterijus.
ZQ, ZQstar ir ZQQstar – Zhango kriterijai.
BCMR – Barrio, Cuestos, Alberto, Matráno ir Rodríguezo kriterijus.
Coin – Coino kriterijus.
DAP – D'Agostino ir Pearsono kriterijus.
JB – Jarque ir Beros kriterijus.
H1, H2, H3 ir H4 – Hoskingo ir Wallio kriterijai.
CC1 ir CC2 – Kabaños ir Kabaños kriterijai.
BS – Bonetto ir Seiero kriterijus.
BM1 ir BM2 – Bontempsio ir Meddahio kriterijai.
DH – Doorniko ir Hanseno kriterijus.
GG – Gelo ir Gastwirtho kriterijus.
BHS – Bryso, Huberto ir Struyfo kriterijus.
BHSBS – Bryso, Huberto, Struyfo, Bonetto ir Seiero kriterijus.
DLDMLRn, DLDMXAPD ir DLDMZEPD – Desgagné Lafaye de Micheaux Leblanco kriterijai.
CHI2 – chi-kvadrato kriterijus.
MI – Martinezo ir Iglewiczo kriterijus.
EP – Eppsio ir Pullio kriterijus.
GMG – Gelio, Miao ir Gastwirtho kriterijus.
Baksh – Bakšajevo kriterijus.
N1 – naujai pasiūlytas vienmatis suderinamumo hipotezės tikrinimo kriterijus.
NRR – Nikulino, Rao ir Robsono kriterijus.
DN – Dzhaparidzės ir Nikulino kriterijus.
Mar1 ir Mar2 – Mardijos kriterijai.
Roy – Roystono kriterijus.
HZ – Henzės ir Zirklerio kriterijus.
LV – Lobato ir Velasco kriterijus.
Energy – „Energy“ kriterijus.
N2 – naujai pasiūlytas daugiamatis suderinamumo hipotezės tikrinimo kriterijus.

ĮVADAS

Darbo aktualumas ir svarba

Šiuolaikinėje duomenų analitikoje sudėtingesnių sprendimų priėmimas neįmanomas be hipotezių tikrinimo. Duomenų analitikai stengiasi panaudoti statistinę ir apriorinę informaciją. Paprastai jie savo tyrimus pradeda nuo hipotezių apie duomenų pasiskirstymą tikrinimo. Tai daroma dėl labai svarbių priežasčių:

- žinomas duomenų skirstinys gali paaiškinti duomenis generuojantį procesą;
- skirstinio charakteristikos gali būti susijusios su pradinio modelio parametrais;
- pagal žinomą duomenų skirstinį galima tikslingai taikyti tinkamus modelius ir sumažinti skaičiavimo kompiuteriu resursus (atmintį, skaičiavimo laiką ir kt.).

Dėl aptartų priežasčių darbe nagrinėjama suderinamumo hipotezės tikrinimo problematika. Tikrinamos nulinė ir alternatyvi hipotezės:

H_0 : Duomenų pasiskirstymas yra suderintas su normaliuoju skirstiniu;

H_A : Duomenų pasiskirstymas nėra suderintas su normaliuoju skirstiniu.

Šių hipotezių apie normalų duomenų pasiskirstymą tikrinimas taikomas (kaip prielaida) daugeliui parametrinių duomenų analizės metodų (pvz., koreliacinė, dispersinė, regresinė analizės). Normalumo prielaida yra viena iš svarbiausių dėl daugybės praktinių taikymo sričių. Vienas iš normalumo prielaidos tikrinimo būdų yra grafinis (t. y. braižant histogramą ir įvertinant pasiskirstymą), tačiau tai nėra patikimas būdas dėl klaidos tikimybės. Patikimesnis būdas – patikrinti suderinamumo hipotezę taikant kriterijų. Nuo 1900 m. suderinamumo hipotezės tikrinimui yra sukurta daug vienmačių ir daugiamačių kriterijų. 1900 m. Karlas Pearsonas pristatė pirmąjį suderinamumo hipotezės tikrinimo kriterijų, kurį pavadino chi-kvadratu [1]. Šis kriterijus laikomas šiuolaikinės statistikos pagrindu. Vėliau buvo pristatomi vis nauji kriterijai. Šiuo metu vieni populiariausių ir dažniausiai taikomų kriterijų yra Kolmogorovo ir Smirnovo [2], Andersono ir Darlingo [3], Kramerio ir von Mises [4], taip pat jų modifikacijos. Kriterijai yra paremti įvairiais matais arba metrikomis, dėl to skirtingai reaguoja į duomenų charakteristikas. Todėl tinkamo suderinamumo hipotezės kriterijaus pasirinkimas išlieka svarbus.

Darbo objektas, subjektas ir matas

Šio tyrimo *objektu* laikoma suderinamumo hipotezė, skirta normalumo prielaidos tikrinimui. Darbo *subjektai* yra suderinamumo hipotezės tikrinimo vienmačiai ir daugiamačiai kriterijai. Suderinamumo hipotezės kriterijų palyginimui naudojamas galios *matas*.

Darbo tikslas

Sukurti ir iširti suderinamumo hipotezės tikrinimo vienmačius ir daugiamačius kriterijus, kurie būtų efektyvūs normalumo prielaidos atveju.

Darbo uždaviniai

1. Ištirti pagrindines problemas, sprendimus bei kliūtis atliekant suderinamumo hipotezių tikrinimą.
2. Sudaryti naujus vienmatį ir daugiamatį suderinamumo hipotezės tikrinimo kriterijus, skirtus normalumo prielaidos tikrinimui.
3. Atlikti sudarytų ir populiarių vienmačių ir daugiamačių suderinamumo hipotezės tikrinimo kriterijų galios lyginamąją analizę moksliniuose tyrimuose taikomų duomenų pasiskirstymų atvejais.
4. Pritaikyti sudarytus suderinamumo hipotezės tikrinimo kriterijus realių duomenų rinkiniams.

Darbo metodika

Disertacijoje taikomi įvairūs tikimybių teorijos, matematinės statistikos, lygiagrečiųjų paskirstytųjų skaičiavimų ir daugiamačių duomenų vizualizavimo metodai. Pasiūlytas vienmatis suderinamumo kriterijus, kuris paremtas N -metrika. Pristatytas daugiamatis kriterijus yra paremtas apvertimo formule. Empirinėje dalyje pateikti Monte Karlo modeliavimo metodu gauti rezultatai. Šio darbo programiniam realizavimui buvo pasitelkti *SAS* ir *R* statistiniai paketai.

Ginamieji teiginiai

- Pasiūlytas N -metrikos teorija paremtas suderinamumo hipotezės tikrinimo kriterijus, kuris yra galingesnis už kitus galingiausius vienmačius kriterijus dideliems imties dydžiams.
- Pasiūlytas pasiskirstymo tankių skirtumo vertinimu ir apvertimo formulės taikymu paremtas daugiamatis suderinamumo hipotezės tikrinimo kriterijus, kuris pasižymi kur kas didesne galia nei kiti galingiausi kriterijai simetrinių ir mišriųjų skirstinių grupėms.

Darbo mokslinis naujumas ir praktinė reikšmė

Pasiūlyti suderinamumo hipotezės tikrinimo kriterijai pratęsia, apibendrina ir patikslina A. Bakšajevo [5, 6], L. Klebanovo [7], M. Kavaliausko [8] ir T. Ruzgo [9] rezultatus. Pasiūlyti vienmatis ir daugiamatis kriterijai buvo pristatyti [**P1**, **K1**, **P2**, **K4**, **K5**]. Pasiūlytas vienmatis kriterijus galingiausias simetriniams duomenų rinkiniams, kurių imties dydis $n \geq 112$, asimetriniams duomenų rinkiniams, kurių imties dydis $n \geq 118$, ir modifikuotų normaliųjų skirstinių duomenų rinkiniams, kurių imties dydis yra $n \geq 88$. Naują daugiamatį kriterijų pakankamo dydžio imtims geriausia naudoti 2-mačiams duomenims, o didesnės dimensijos duomenims šį kriterijų rekomenduojama naudoti su dideliais imties dydžiais ($n \geq 512$). Kriterijai buvo pritaikyti realių duomenų rinkiniams. Pagal gautus rezultatus galima padaryti išvadą, kad šiame darbe pristatyti statistiniai kriterijai gali būti sėkmingai taikomi realių duomenų analizės suderinamumo hipotezės tikrinimo uždaviniuose. Šiame darbe pristatytus kriterijus ir jų taikymo rekomendacijas Hugas Hernandezas [10], Silviu Gabrielis Szentesi ir kt. [11], Sergejus Gričaras ir kt. [12], Atif Avdović ir kt. [120], Hanan Haj Ahmad su Ehab M. Almetwally [121], Daniel Doz ir kt. [122] naudojo savo tyrimuose.

Darbo rezultatų aprobavimas

Disertacijos tema yra išspausdinti trys moksliniai straipsniai, kurie publikuoti informacijos instituto duomenų bazės (ISI) leidiniuose su citavimo indeksu, o vienas straipsnis publikuotas recenzuojamame mokslo leidinyje (be citavimo indekso). Viena publikacija išspausdinta nacionalinėje ir trys – tarptautinėje leidykloje.

Atliktų tyrimų rezultatai pristatyti šešiose tarptautinėse ir nacionalinėse konferencijose. Skaityti pranešimai: 24-ojoje tarptautinėje „Mathematical modelling and analysis“ 2019 (Talinas, Estija), 11-ojoje tarptautinėje „Data Analysis Methods for Software Systems“ 2019 (Druskininkai, Lietuva), 9-ojoje nuotolinėje, tarptautinėje „International conference on applied analysis and mathematical modeling“ 2021 (Stambulas, Turkija), 62-ojoje nacionalinėje Lietuvos matematikų draugijos konferencijoje, 1-ojoje nuotolinėje, tarptautinėje „SAB-ATAN: International Conference on Mathematical Sciences and Applications“ 2021 (Bauio, Filipinai) ir 12-ojoje tarptautinėje „Data Analysis Methods for Software Systems“ 2021 (Druskininkai, Lietuva).

Daktaro disertacija pristatyta:

- 2022 m. sausio 14 d. Kauno technologijos universiteto Matematikos ir gamtos mokslų fakulteto Taikomosios matematikos katedros posėdyje;
- 2022 m. vasario 15 d. Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakulteto Matematinio modeliavimo katedros seminare;
- 2022 m. birželio 2 d. Informatikos mokslo krypties doktorantūros komiteto posėdyje.

Disertacijos struktūra

Disertaciją sudaro įvadas, literatūros apžvalga, tyrimo metodika bei tyrimo rezultatų aptarimo dalys, darbo išvados, literatūros šaltinių sąrašas, trumpas autorės gyvenimo aprašymas su mokslinių publikacijų sąrašu, padėka ir priedai. Disertacijos apimtis – 150 puslapių. Disertacijos pagrindinėje dalyje yra 23 paveikslai, 11 lentelių ir 122 cituojamų literatūros šaltinių sąrašas (iš jų 18,33 proc. ir 34,17 proc. per pastaruosius penkerius ir dešimt metų).

I. MOKSLINĖS LITERATŪROS APŽVALGA IR ANALIZĖ

Šioje dalyje pristatomas tyrimo objektas – suderinamumo hipotezė. Aptariami suderinamumo hipotezės, skirtos patikrinti normalumo prielaidą, kriterijų lyginamieji tyrimai. Pabaigoje pateikiamos literatūros apžvalga ir analizę apibendrinančios išvados.

1.1. Suderinamumo hipotezė ir jos kriterijų lyginamoji apžvalga

Suderinamumo hipotezė (angl. *goodness of fit*) yra neparimetrinė hipotezė apie stebimo atsitiktinio dydžio X pasiskirstymo funkciją:

$$H_0: F_X \equiv F_0(X).$$

$$H_A: F_X \neq F_0(X).$$

1900 m. šią hipotezę imta tikrinti, kai buvo sukurtas pirmasis chi-kvadrato kriterijus. Tačiau patį terminą 1984 m. pasiūlė Lazarusas ir Folkmanas [13], o 1987 m. išplėtojo Forsythes ir Compass [14].

Dažniausiai suderinamumo hipotezė taikoma tikrinant prielaidą apie duomenų suderinamumą su normaliuoju skirstiniu. Šios prielaidos patikrinimas yra svarbus duomenų analizės metodų parinkimui. Pavyzdžiui, dispersinėje analizėje (taikyti Fišerio apibendrinantį tiesinį modelį ar Kruskalo Voliso neparimetrinį metodą) [15, 16], regresinėje analizėje (taikyti mažiausių kvadratų tiesinį modelį ar neparimetrinį splainų, arba branduolinės regresijos metodą) [17, 18]. Duomenų pasiskirstymo nustatymas svarbus ir neuroniniuose tinkluose tiek parenkant aktyvacijos funkcijas, tiek konstruojant patį neuroninį tinklą [19, 20].

Turimai imčiai X_1, X_2, \dots, X_n , gautai stebint atsitiktinį dydį X , kurio skirstinys priklauso žinomai skirstinių šeimai. Laikydami, kad H_0 yra teisinga, parenkama kriterijaus statistika $\gamma_n = \gamma_n(X_1, \dots, X_n)$, kurios skirstinys yra žinomas arba aproksimuojamas žinomu skirstiniu. Statistikos galimų reikšmių aibė \mathcal{R} skaidoma į dvi sritis \mathcal{K} ir $\mathcal{R} \setminus \mathcal{K}$. Hipotezė H_0 tikrinama pagal tokią procedūrą: jei sudarius konkrečią imtį X_1, X_2, \dots, X_n , statistikos γ_n reikšmė patenka į kritinę sritį \mathcal{K} , tai H_0 yra atmestina, o jei patenka į hipotezės priėmimo sritį $\mathcal{R} \setminus \mathcal{K}$, tai H_0 yra priimtina.

Kriterijaus statistikos reikšmė turi vienareikšmę atitikį su tikimybe, vadinama empiriniu reikšmingumo lygmeniu arba p -reikšme. Praktikoje, naudojant programinę įrangą, tam, kad būtų priimtas sprendimas dėl nulinės hipotezės priėmimo arba atmetimo, dažnai yra užsiduodamas teorinis reikšmingumo lygmuo α (pvz., $\alpha = 0,05$) ir atliekamas jo palyginimas su p -reikšme:

- jei p -reikšmė $> \alpha$, tai nulinė hipotezė yra priimama;
- jei p -reikšmė $\leq \alpha$, tai nulinė hipotezė yra atmetama.

Suderinamumo hipotezės kriterijai pagal duomenų dimensiją gali būti skirstomi į vienmačius ir daugiamačius. Pirmiausia buvo sukurti ir analizuoti vienmačiai kriterijai. 1900 m. pradėtas prielaidos apie duomenų pasiskirstymą nagrinėjimas [1]. Šis kriterijus ir dabar yra plačiai taikomas. Tobulėjant technologijoms, mokslininkai kūrė, tikslino ir modifikavo vienmačius suderinamumo hipotezės tikrinimo kriterijus [10, 21–34]. Vieni iš pirmųjų buvo

sukurti Kolmogorovo ir Smirnovo [2], Anderseno ir Darlingo [3], Kramerio ir von Mises [4, 35] bei Šapiro ir Vilko [36] kriterijai. Šie kriterijai ir dabar yra plačiai taikomi bei realizuoti daugelyje statistinių paketų (*SAS*, *SPSS*, *R* ir kt.). Plačiau šie vienmačių kriterijų tyrimai aptariami 1.2 skyriuje.

Nuo 1981 m. pradėti pirmieji tyrimai su daugiamačiais suderinamumo hipotezės tikrinimo kriterijais. 1981 m. Moore ir Stubblebine pristatė pirmą daugiamatį Chi-kvadrato kriterijų [37]. Nuo tada prasidėjo daugiamačių kriterijų kūrimo ir lyginimo tyrimai [37–46]. Vieni iš pirmųjų daugiamačių suderinamumo kriterijų yra Henzės ir Zirklerio [39], Kramerio ir von Mises [40], Roystono [45]. Daugiausiai daugiamačių suderinamumo hipotezės tikrinimo kriterijų yra realizuoti statistinių skaičiavimų *R* ir kituose statistiniuose paketuose. Plačiau šie daugiamačių kriterijų tyrimai aptariami 1.3 skyriuje.

Taikant statistinius kriterijus galima padaryti šias klaidas.

- Atmesti H_0 , nors ji yra teisinga. Ji vadinama I rūšies klaida. Tikimybė, kad ši klaida bus padaryta, yra

$$\alpha = P(\gamma_n \in \mathcal{K} | H_0).$$

- Priimti H_0 , nors ji yra klaidinga. Ji vadinama II rūšies klaida. Tikimybė, kad ši klaida bus padaryta, yra

$$\beta = P(\gamma_n \notin \mathcal{K} | H_A).$$

Dažniausias kriterijų lyginimo matas buvo galia [5, 6, 10, 22, 25, 32, 34, 38, 47, 48, 57–60, 62–66, 69]. Kriterijaus *galia* vadinama tikimybė atmesti neteisingą H_0 hipotezę. Galia yra priešinga II rūšies klaidos tikimybei. Mažinant I klaidos tikimybę α didėja II klaidos tikimybė β , o kriterijaus galia mažėja. Kriterijus yra tuo geresnis, kuo yra mažesnės abiejų rūšių klaidos. Praktikoje kriterijai sudaromi taip, kad fiksuotai I rūšies klaidai II rūšies klaida būtų kiek galima mažesnė. Dažniausiai pasirenkama α reikšmė, lygi 0,05. Paprastai yra skaičiuojama priešingo įvykio tikimybė

$$1 - \beta = P(\gamma_n \in \mathcal{K} | H_A),$$

t. y. *kriterijaus galia*. β yra tikimybė neatmesti nulinės hipotezės, kai ji klaidinga. Kriterijaus galia leidžia palyginti du kriterijus, turinčius tą patį α ir taikomus tokio pat dydžio imtims. Šis matas yra nedimensinis normuotas dydis, priklausantis intervalui (0; 1), taip pat dar naudojama ir procentinė išraiška [5, 6]. Galingesniu kriterijumi laikomas tas, kuriam $1 - \beta$ yra didesnis. Didinant imties dydį kriterijaus galia paprastai didėja [108, 109].

Suderinamumo hipotezės tikrinimo kriterijų palyginimas atliekamas dviem būdais.

- *Taikant matematinius išvedimus ir įrodymus*. Šio tipo tyrimai yra gana reti. Jiems atlikti reikalingos išsamios tikimybių teorijos ir matematinės statistikos žinios. Vieną iš išsamiausių statistinių hipotezių tyrimų atliko A. Bakšajevs [5, 6]. Mokslininkas pristatė *N*-metrika paremtą kriterijų. Jį matematiškai

išvedė ir atliko jo tyrimą taikydamas imitacinį modeliavimą [6]. Krauczis savo darbuose [47, 48] nagrinėjo suderinamumo kriterijus, paremtus empirine pasiskirstymo funkcija ir koreliacija. Foudjo susitelkė ties robustinių kriterijų modifikavimu ir lyginimu [49]. A. Desgagné ir P. Lafaye de Micheaux nagrinėjo suderinamumo kriterijus, paremtus eksceso ir asimetrijos koeficientais, teoriniu ir modeliavimo būdais [25]. Górecki su kolegomis tyrė suderinamumo kriterijus, tinkančius funkciniams duomenims [27]. N. Henze su kolegomis daug dėmesio skyrė daugiamačių kriterijų kūrimui ir lyginimui [39, 40]. Jis pristatė ir įrodė kriterijų invariantiškumo ir suderinamumo savybes [40]. D. Moore ir J. Stubblebine modifikavo Chi-kvadrato kriterijų ir jį pritaikė daugiamačiams duomenims [37]. Plačiau šių ir kitų mokslininkų darbai aptariami 1.2 ir 1.3 skyriuose.

- *Taikant imitacinį modeliavimą.* Tai viena populiariausių suderinamumo hipotezės tikrinimo kriterijų palyginimo metodikų. Itin išsamius ir plačius tyrimus atliko H. Hernandezas [10], F. Ahmad [22], A. Desgagné [25], I. Agu Friday [26], T. Górecki [27], X. Romão [28], O. Kesemenas [41] ir daugelis kitų. M. Steele pristatė suderinamumo hipotezės tikrinimo kriterijus diskretiesiems duomenims ir atliko jų palyginimą su kitais populiariais kriterijais. Tyrimus atliko daugelio įvairių duomenų skirstinių atveju [21]. Michael su kolegomis tyrimus vykdė diskrečiųjų suderinamumo hipotezės kriterijų palyginimo tematika [50, 51]. Plačiau šių ir kitų mokslininkų darbai aptariami 1.2 ir 1.3 skyriuose.

Mokslinius straipsnius, kuriuose nagrinėjama suderinamumo hipotezės tikrinimo kriterijų problematika, galima išskirti į tris dalis.

1. Teoriniai suderinamumo hipotezės tikrinimo kriterijų išvedimai ir įrodymai [5, 6, 25, 37, 39, 40, 47–49].
2. Suderinamumo hipotezės tikrinimo kriterijų tyrimai taikant imitacinį modeliavimą [10, 21–34, 37–46, 50–52].
3. Suderinamumo hipotezės tikrinimo kriterijų panaudojimas sprendžiant realaus pasaulio uždavinius [15–20, 53, P4].

1.2. Vienmačių kriterijų tyrimų apžvalga ir analizė

XX a. pradžioje imta kurti ir nagrinėti vienmačius suderinamumo hipotezės tikrinimo kriterijus [1]. Nuo tada iki šių dienų vis dar atliekamas vienmačių suderinamumo hipotezės naujų kriterijų kūrimas ir lyginimas su kitų mokslininkų pristatytais kriterijais. Praktiniu aspektu aktualiausias išlieka normalumo prielaidos tikrinimas. Vienmačiai kriterijai ypač dažnai taikomi atliekant duomenų analizę. XXI a. šios tematikos tyrimai buvo itin aktyviai atliekami [5, 6, 10, 21–34, 47–51]. Mažoji dalis šios tematikos tyrimų rėmėsi matematiniais įrodymais, o didžioji – imitaciniu modeliavimu. Mokslininkai savo tyrimuose palygino didelį kiekį klasikinių suderinamumo hipotezės kriterijų (iki 50). Tyrimų metu kriterijai buvo taikomi įvairiems imčių dydžiams (nuo 10 iki 2000), naudojant įvairias reikšmingumo lygmens α vertes (dažniausiai pasirenkamas $\alpha = 0,05$). Modeliavimo tyrimai atlikti daugeliui simetrinių, asimetrinių ir modifikuotų

normaliųjų skirstinių bei jų mišinių duomenims. Kriterijai buvo lyginami naudojant galios matą. Tyrėjai dažnai naudojo R programinę įrangą su `POWER` biblioteka.

A. Bakšajevs pasiūlė suderinamumo ir kitų statistinių hipotezių kriterijų statistikų konstravimo būdą, parentą N -metrika. N -metrika (dar vadinama N -atstumu) yra tikimybinis matas¹, kuris apibrėžiamas atstumu tarp dviejų tikimybinių matų q ir ν Hilberto erdvėje \mathcal{H} . Šios metrikos erdvė yra izometrinė išgautos Hilberto erdvės pogrupiui. Hilberto erdvė, pavadinta D . Hilbert garbei, apibendrina euklidinės erdvės sąvoką. Ji išplečia vektorių algebrą iš dviejų arba trijų matavimų Euklido erdvėje į daugelio matmenų ar net begalinę erdves. Metrines erdve vadinama pora (X, ρ) , susidedanti iš tam tikros aibės (erdvės) X elementų ir atstumo, t. y. vienareikšmės, neneigiamos, realiosios funkcijos $\rho(x, y)$, apibrėžtos bet kuriems x ir y iš X ir tenkinančios tokias tris aksiomas:

1. $\rho(x, y) = 0$ tada ir tik tada, kai $x = y$;
2. $\rho(x, y) = \rho(y, x)$ (simetrijos aksioma);
3. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ (trikampio aksioma).

Laikoma, kad (X, ρ) ir (Y, d) yra dvi metrinės erdvės. Abipusiškai vienareikšmis atvaizdis $f : X \rightarrow Y$ vadinamas izometrija, jei $\rho(x, y) = d(f(x), f(y))$ visiems $x, y \in X$. Erdvės (X, ρ) ir (Y, d) , tarp kurių galima sukonstruoti izometrinį atvaizdį, vadinamos izometrinėmis. Erdvių (X, ρ) ir (Y, d) izometrija reiškia, kad metriniai ryšiai tarp jų elementų yra tie patys. Gali skirtis tik elementų kilmė (prigimtis), o tai metrinė erdvių požiūriu yra neesminga. Kitaip sakant, izometrinės erdves galima tapatinti.

N -metriką galima apibrėžti [7]:

$$N(q, \nu) = 2 \int_{\mathcal{H}} \int_{\mathcal{H}} K(x, y) dq(x) d\nu(y) - \int_{\mathcal{H}} \int_{\mathcal{H}} K(x, y) dq(x) dq(y) - \int_{\mathcal{H}} \int_{\mathcal{H}} K(x, y) d\nu(y) d\nu(x); \quad (1.1)$$

čia $K(x, y)$ yra stipriai neigiama branduolio funkcija². L. Klebanovo pristatytai N -metrikai A. Bakšajevs pritaikė branduolio funkcijas: t. y. $K1 = K(x, y) = |x - y|$, $K2 = K(x, y) = 1 - \exp(-\|x - y\|^2)$ ir kt. Plačiau A. Bakšajevs pasiūlytas kriterijus yra apibrėžiamas 2.1.4 poskyryje. Sudarytas kriterijus palygintas remiantis Bahaduro [54] pasiūlytu, Nikitino [55] išplėtotu asimptotiniu santykinio efektyvumu. Bahaduro metodu nustatoma kriterijaus galia $1 - \beta$ ir palyginamas kriterijų galios mažėjimo greitis α , kai didėja imties dydis n . Tegul U_n ir V_n yra dvi statistikos sekos,

¹ Funkcija $P: \mathfrak{S} \rightarrow [0, 1]$, abibrėžta aibės Ω (t. y. elementariųjų įvykių erdvės) poaibių σ -algebroje \mathfrak{S} , vadinama tikimybiniu matu, jei $P(\Omega)=1$, $P(\emptyset)=0$; $A \subset B \Rightarrow P(A) \leq P(B)$; čia $A, B \in \mathfrak{S}$; $A_i \cap A_j = \emptyset$, čia $i \neq j$, kai $i, j \in N \Rightarrow P(\sum_{j \in N} A_j) = \sum_{j \in N} P(A_j)$, $A_j \in \mathfrak{S}$; $A_1 \supset A_2 \supset \dots$, $\bigcap_{j=1}^{\infty} A_j = \emptyset \Rightarrow \lim_{j \rightarrow \infty} P(A_j) = 0$.

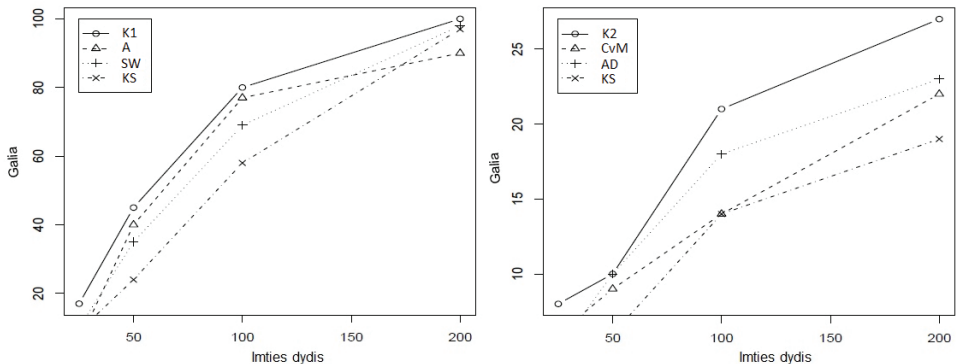
² Jeigu $\forall n \in \mathbb{N}, \forall x_1, \dots, x_n \in \mathcal{X}$, kai \mathcal{X} – bet kokia aibė, ir $\forall c_1, \dots, c_n \in \mathbb{C}$ su sąlyga, kad $\sum_{i=1}^n c_i = 0$, funkcija $K(x, y)$ tenkina nelygybę $\sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) c_i \bar{c}_j \leq 0$, tai K yra neigiamai apibrėžtas branduolys.

paremtos imties dydžiu n . Jos yra skirtos nulinės ir alternatyvios hipotezių palyginimui. Tada fiksuotiems $\beta \in (0,1)$ ir poslinkio parametru θ egzistuoja riba [6]:

$$e_{U,V}(\beta, \theta) = \lim_{\alpha \rightarrow 0} e_{U,V}(\alpha, \beta, \theta); \quad (1.2)$$

čia $e_{U,V}$ – santykinis asimptotinis sekos U_n efektyvumas sekos V_n atžvilgiu.

Šią ribą Bahaduras pavadino sekos U_n asimptotiniu santykiniu efektyvumu sekos V_n atžvilgiu. Naudojantis asimptotiniu santykiniu efektyvumu iš 10 stipriai neigiamai apibrėžtų branduolio funkcijų buvo pasirinkta viena branduolio funkcija. Pagal Bahaduro teoriją ji buvo asimptotiškai optimali, ir šios funkcijos tikslios krypties (angl. *exact slope*) ir lokalsios parametų reikšmės buvo didžiausios. Lygiagrečiai su teoriniais tyrimais, taikant Monte Karlo³ metodą, buvo atliekamas empirinis kriterijų galios palyginimas. Autoriaus pasiūlytas kriterijus buvo lyginamas su D'Agostino (A), Lillieforso (KS), Andersono ir Darlingo (AD), Pirson (P), Kramerio ir von Mises (CvM), Šapiro ir Vilko (SW) bei Šapiro ir Francijos (SF). Modeliavimo tyrimui pasirinkti imties dydžiai $n = 25, 50, 100, 200$ ir reikšmingumo lygmuo $\alpha = 0,05$. Kriterijų galia buvo vertinama generuojant imtis pagal 200 skirtingų skirstinių su įvairiais parametų rinkiniais. A. Bakšajevo keletu galingiausių kriterijų galios kitimas didėjant imties dydžiu $0,5N(0,1) + 0,5N(3,1)$ ir $0,5Exp(1) + 0,5Exp(0,5)$ skirstinių atvejais pateikiamas 1.1 paveiksle. Pagal gautus rezultatus nustatyta, kad pasiūlytas kriterijus yra galingas klasikinių kriterijų konkurentas [5, 6].



1.1 pav. Kriterijų galios palyginimas $0,5N(0,1) + 0,5N(3,1)$ ir $0,5Exp(1) + 0,5Exp(0,5)$ skirstinių atvejais

A. Desgagné ir P. Lafaye de Micheaux pristatė Jarquès ir Beros (JB) kriterijaus alternatyvas, paremtas asimetrijos ir eksceso koeficientais naudojant Rao įnašo kriterijų (dar žinomą kaip Lagranžo daugiklių metodas⁴) asimetriniams, laipsniniams

³ Monte Karlo metodas – skaičiavimo metodas, paremtas statistiniu modeliavimu ir gautų rezultatų apdorojimu taikant statistinius metodus.

⁴ Lagranžo daugiklių metodas yra funkcijos arba funkcionalo sąlyginio ekstremumo radimo būdas.

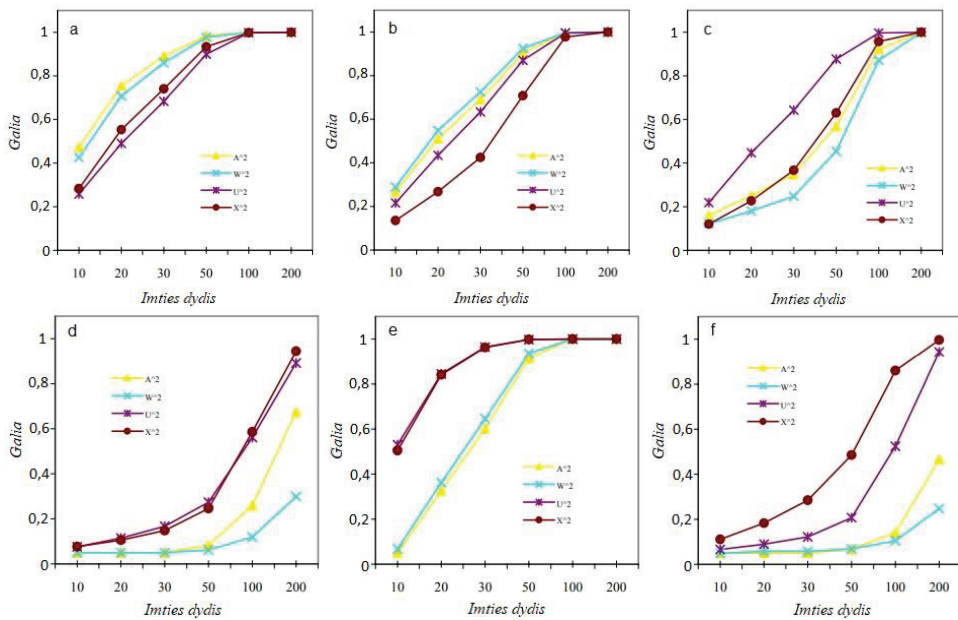
skirstiniams (APD). Autoriai sudarė paprastą kriterijaus statistiką X_{APD}^a (a reiškia asimptotinis), kurios forma panaši į JB, bet joje yra įtraukti 3-iojo laipsnio asimetriškumo ir 4-ojo laipsnio eksceso koeficientai. Ši statistika sudaryta naudojant Rao įnašo kriterijų, kuris taikomas eksponentiniams ir laipsniniams skirstiniams (EPD). X_{APD}^a statistika yra asimptotiškai pasiskirsčiusi pagal Chi-kvadrato pasiskirstymą su dviem laisvės laipsniais. Taip pat pasiūlyta kriterijaus statistika Z_{EPD}^a baigtiniams imties dydžiams $n \geq 10$. Ji yra paremta 2-ojo/ laipsnio eksceso koeficientu. Z_{EPD}^a statistika yra asimptotiškai pasiskirsčiusi pagal $N(0,1)$ skirstinį. A. Desgagné ir P. Lafaye de Micheaux pasiūlyti kriterijai plačiau apibrėžiami 2.1.3 poskyryje. Autorių pasiūlyti kriterijai palyginti su 33 kitų mokslininkų pristatytais kriterijais. Taikyti kriterijai paremti regresija, koreliacija, empirine pasiskirstymo funkcija, asimetrijos ir eksceso koeficientais. Modeliavimo tyrimui buvo 85 duomenų skirstiniai: 33 simetriniai skirstiniai su sunkiomis uodegomis, 26 simetriniai skirstiniai su lengvomis uodegomis ir 26 asimetriniai skirstiniai. Empirinė galia buvo skaičiuojama imties dydžiams $n = 10, 20, 50, 100$ ir 200 su reikšmingumo lygmeniu $\alpha = 0,05$. Modeliavimo tyrimas atliktas generuojant 1 000 000 nepriklausomų imčių. Modeliavimui atlikti buvo naudojama R programinė įranga su POWER biblioteka [56]. Pagal gautus rezultatus nustatyta, kad X_{APD}^a yra galingesnis už asimetrijos ir eksceso koeficientais paremtus Jarques ir Beros, D'Agostino ir Pearsono, Doorniko ir Hanseno kriterijus. Coino ir Z_{EPD}^a kriterijai yra galingiausi simetriškiems skirstiniams, nagrinėjant imties dydžius $n = 10, 20, 50, 100$. Apibendrinant visą tyrimą nustatyta, kad didžiausią galią visais nagrinėtais atvejais turėjo Cheno ir Šapiro, X_{APD}^a bei Šapiro ir Vilko kriterijai [25].

É. O. Krauczi nagrinėjo suderinamumo hipotezės kriterijus, paremtus Vaseršteino metrika⁵. Ši metrika įvertina Vaseršteino atstumą (angl. *Wasserstein distance*) tarp sugeneruoto ir pradinio duomenų rinkinio. Autorė pristatė svorinių kvantilių koreliacijos kriterijų K_n . Siekiant palyginti pasiūlyto kriterijaus galią su kitais, atliktas modeliavimo tyrimas. K_n kriterijus palygintas su Kolmogorovo ir Smirnov, Kramerio ir von Mises, Andersono ir Darlingo bei Watsono kriterijais. Modeliavimo tyrimas atliktas generuojant 100 000 nepriklausomų imčių su 13 simetrinė ir asimetrinių skirstinių. Procentinė empirinė galia buvo skaičiuojama generuotiems imties ($n = 20, 50, 100$) dydžiams su reikšmingumo lygmenimis $\alpha = 0,1$ ir $\alpha = 0,05$. Pagal gautus modeliavimo rezultatus nustatyta, kad pasiūlytas kriterijus buvo galingiausias Koši, eksponentinio, lognormaliojo ir Chi-kvadrato skirstiniams, kai imties dydis $n \geq 50$ [47, 48].

M. Steele nagrinėjo diskrečiuosius Kramerio ir von Mises (W^2), Kolmogorovo ir Smirnov (S) bei Chi-kvadrato (χ^2) tipo suderinamumo hipotezės kriterijus. Ieškojo tinkamiausių kriterijų kategoriniams duomenims. Kategoriniams duomenims buvo parinkti Chi-kvadrato ir diskretieji Kramerio ir von Mises bei Kolmogorovo ir Smirnov kriterijai. M. Steele palygino χ^2 , S , W^2 , U^2 (diskretusis Watsono), A^2 (diskretusis Andersono ir Darlingo) ir NS (nominalusis Kolmogorovo

⁵ Vaseršteino metrika yra atstumo funkcija, apibrėžta tarp tam tikros metrinės erdvės tikimybių skirstinių.

ir Smirnovo) kriterijus įvairioms skirstinių grupėms. Kriterijų empirinė galia buvo skaičiuojama imties dydžiams $n = 10, 20, 30, 50, 100$ ir 200 su reikšmingumo lygmeniu $\alpha = 0,05$. Modeliavimo tyrimas atliktas generuojant 10 000 nepriklausomų imčių. Autoriaus gauti modeliavimo rezultatai pateikiami 1.2 (a–f) paveiksluose. Pagal šiuos paveikslus galima teigti, kad kriterijus A^2 yra galingiausias mažėjančios formos (angl. *consistent trend*) skirstiniams, o silpniausias – bimodaliniams skirstiniams. Kriterijus W^2 yra galingiausias laiptinės formos skirstiniams, o silpniausias – lėkštos formos ir bimodaliniams skirstiniams. Kriterijus U^2 yra galingiausias trikampio formos skirstiniams, o silpniausias – bimodaliniams skirstiniams. Kriterijus χ^2 yra galingiausias alternatyviems skirstiniams su lėkštos bei smailios formos tankio funkcijomis ir bimodaliniams, o silpniausias – trikampio formos skirstiniams [57].



1.2 pav. Kriterijų galios palyginimas mažėjančios (a), laiptinės (b), trikampio (c), bukos (d), smailios (e) formų ir bimodaliniams (f) skirstiniams

Autorius taip pat pateikia šešių kriterijų (χ^2 , S , W^2 , U^2 , A^2 ir NS) statistikų galios palyginimą šešioms skirtingos formos skirstinių grupėms, kuri pateikiama 1.1 lentelėje [21, 50, 51].

1.1 lentelė. Suderinamumo kriterijų statistikų galios palyginimas

Nr.	Pasiskirstymų grupės pagal tankio funkcijos formą	Kriterijų statistikų palyginimas
1.	Mažėjantys	$A^2 > W^2 > S > \chi^2 > U^2 > NS$
2.	Laiptiniai	$W^2 \approx S \approx A^2 > U^2 \gg \chi^2 \approx NS$

1.1 lentelė. Suderinamumo kriterijų statistikų galios palyginimas

Nr.	Pasiskirstymų grupės pagal tankio funkcijos formą	Kriterijų statistikų palyginimas
3.	Trikampiai (trikampio formos)	$U^2 \gg \chi^2 \approx NS > A^2 > W^2 \approx S$
4.	Lėkštieji (su neigiamuoju eksceso koeficientu)	$U^2 \approx \chi^2 \approx NS \gg A^2 > W^2 \approx S$
5.	Smalieji (su teigiamuoju eksceso koeficientu)	$U^2 \approx \chi^2 > NS \gg S > W^2 \approx A^2$
6.	Bimodaliniai	$\chi^2 \approx NS > U^2 > S \approx A^2 \approx W^2$

X. Romão su kolegomis empiriškai palygino vienmačių suderinamumo hipotezės kriterijų galią. Pasirinkti 33 kriterijai, kurie paremti empirine pasiskirstymo funkcija, asimetrijos bei eksceso koeficientais ir regresijos bei koreliacijos koeficientais. Šie 33 kriterijai pateikti 2.1 skyriuje. Kriterijai buvo lyginami trims imčių dydžiams ($n = 25, 50, 100$) atsižvelgiant į reikšmingumo lygmenis ($\alpha = \{0,1; 0,05; 0,025; 0,01\}$) ir simetriniams, asimetriniams bei modifikuotiems alternatyviesiems skirstiniams. Palyginimas buvo atliktas Monte Karlo modeliavimo metodu sugeneruojant 1 000 000 nepriklausomųjų imčių. Tyrimo metu nustatyta, kad galingiausi iš pasirinktų suderinamumo hipotezės kriterijų simetriškiems alternatyviems skirstiniams buvo Coino β_3^2 , Cheno ir Šapiro, Bonetto ir Seiero bei Gelo, Miao ir Gastwirtho kriterijai, asimetriškiems – Zhango ir Wu (Z_A), Cheno ir Šapiro bei Šapiro ir Vilko kriterijai, o Cheno ir Šapiro bei Barrio, Cuesta, Albertos, Matrános, Rodríguezo ir Rodríguezo bei Šapiro ir Vilko kriterijai buvo galingiausi modifikuotiems normaliesiems alternatyviesiems skirstiniams. Rekomenduojama naudoti Hoskingo, Bryso, Huberto ir Struyfo kriterijus, paremtus asimetrijos ir eksceso koeficientais, nes jie yra mažiau jautrūs išskirtims [28].

G. Brysas su kolegomis pasiūlė suderinamumo kriterijų, paremtą robastiniais⁶ asimetrijos koeficientais ir uodegų svoriais. Jo pasiūlytas kriterijus yra Jarques ir Beros kriterijaus modifikacija. Šis kriterijus plačiau apibrėžiamas 2.1.3 poskyryje. G. Bryso pristatytas kriterijus yra palygintas su Jarquės ir Beros bei Moorso kriterijais. Šiam tyrimui buvo naudojami du imties dydžiai ($n = 100$ ir 1000) įvairiems skirstiniams. Modeliavimo tyrimas buvo atliktas generuojant 1 000 nepriklausomųjų imčių su reikšmingumo lygmeniu $\alpha = 0,05$. Pagal gautus rezultatus nustatyta, kad pasiūlytas kriterijus buvo galingiausias palyginti su kitais, o ypač tinkamas duomenims su išsiskiriančiomis reikšmėmis [24].

F. Ahmadzas su R. A. K. Sherwani atliko suderinamumo hipotezės kriterijų palyginimą. Tyrimui pasirinkta 12 kriterijų. Šių kriterijų empirinė galia buvo lyginama imties dydžiams ($n = \{10, 20, 30, 40, 50, 100, 200, 500\}$), skirtingiems skirstiniams (Koši, eksponentinis, logistinis ir Veibulo) su reikšmingumo lygmenimis $\alpha = 0,05$ ir $0,01$. Modeliavimo tyrimas atliktas generuojant 10 000 nepriklausomųjų imčių. Pagal gautus rezultatus nustatyta, kad tinkamesnis reikšmingumo lygmuo $\alpha = 0,05$, nes visų nagrinėjamų kriterijų galia buvo

⁶ Robastinis yra nejautrus išskirtims.

didžiausia. Kolmogorovo ir Smirnovo, Kramerio ir von Mises kriterijai galingiausi visiems nagrinėjamiems alternatyviesiems skirstiniams. Apibendrinęs atliktą tyrimą autorius padarė išvadą, kad kriterijų galiai įtakos turėjo imties dydis, reikšmingumo lygmuo ir pasirinktas duomenų skirstinys. Didinant imties dydį ir reikšmingumo lygmenį, kriterijų galia didėja [22].

H. Hernandezas pristatė kriterijų, paremtą rangine statistika, ir atliko platų modeliavimo tyrimą, kurio metu palygino daugelio autorių 1990–2021 m. pristatytus suderinamumo kriterijus (tarp jų ir [P1]). Autorius palygino 55 kriterijus, iš kurių 13 buvo paremti empirine pasiskirstymo funkcija, 17 – rangine statistika (regresija ir koreliacija), 17 – momentais ir kiti 8. Kriterijų galia įvertinta taikant Monte Karlo modeliavimo metodą skirtingo dydžio ($n = 5, 10, 15, 20, 25, 30, 40, 50, 100, 200$) imtims su reikšmingumo lygmenimis ($\alpha = 0,05$ ir $0,08$) ir šešiams (standartinis normalusis, tolygusis, eksponentinis, Chi-kvadrato ir lognormalusis) skirstiniams. Modeliavimo tyrimas buvo atliktas taikant 1000 nepriklausomųjų imčių. Modeliavimui atlikti buvo naudojama R programinė įranga. Kriterijai suranguoti pagal geriausią įvertinimą (t. y. tyrimuose išrinktas kaip galingiausias), įvertinimų santykį (t. y. $\frac{\text{galingiausiai įvertintas}}{\text{tyrimų skaičius, kuriuose buvo įtrauktas kriterijus}}$) ir vidutinis rangas. Apibendrinti kriterijų rangų palyginimo rezultatai pagal jų tipą pateikiami 1.2 lentelėje. Pagal gautus rezultatus nustatyta, kad galingiausi yra kriterijai, paremti rangine statistika. Autorius nustatė, kad galingiausias kriterijus yra Šapiro ir Vilko, o silpniausias – Bakšajevio [10].

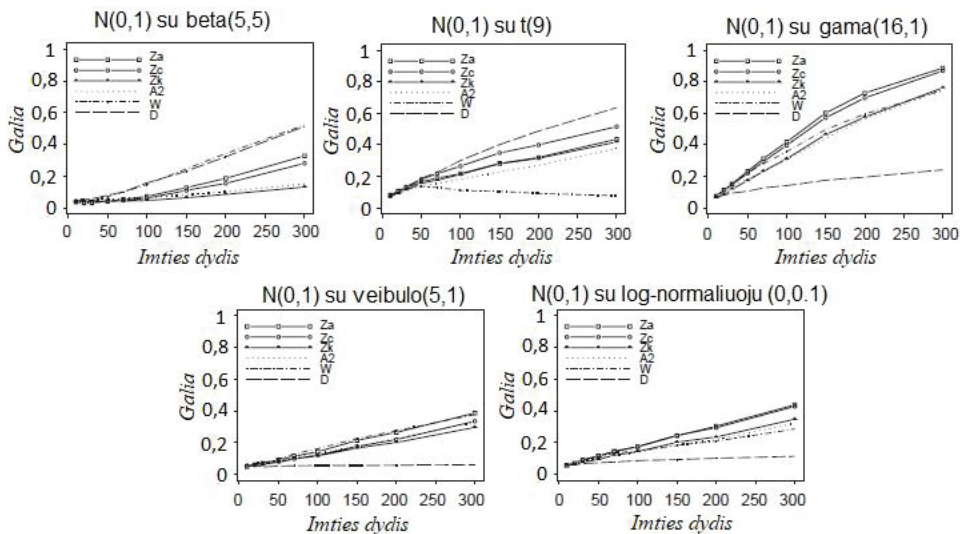
1.2 lentelė. Kriterijų rangų palyginimas pagal tipą

Kriterijų tipas	Empirinė pasiskirstymo funkcija	Ranginė statistika	Momentai	Kiti
Vertinimo kriterijus				
Geriausiai įvertinti	1	1	1	2
# Galingiausių skaičius	4	12	4	0
Įvertinimų santykis	21,1 %	60,0 %	22,2 %	0,0 %
Vidutinis rangas	8,1	6,5	11,3	11,8
# Tyrimų skaičius	19	20	18	12
# Kriterijų vertinimas	74	70	62	33

J. Adefisoye su B. Golam Kibria atliko 18 suderinamumo kriterijų empirinės galios palyginamąjį tyrimą. Kriterijų palyginimui buvo taikomas Monte Karlo modeliavimo metodas įvairiems simetriniams (beta, tolygusis, Stjudento ir Laplaso) ir asimetriniams (gama, chi-kvadrato, eksponentinis, lognormalusis, Veibulo ir Gompertzo) skirstiniams su skirtingais imties dydžiais ($n = 20, 50, 100, 500$). Modeliavimo tyrimas atliktas taikant 10 000 nepriklausomųjų imčių su reikšmingumo lygmeniu $\alpha = 0,05$. Šiam tyrimui atlikti buvo naudojama R programinė įranga. Pagal gautus modeliavimo rezultatus nustatyta, kad beta skirstinio atveju galingiausias eksceso kriterijus, o silpniausias – Bryso, Huberto ir Struyfo kriterijus. Tolygiojo skirstinio atveju galingiausias eksceso kriterijus, o silpniausias – asimetrijos. Stjudento skirstinio atveju galingiausias robustinis Jarque

ir Beros kriterijus, o silpniausias – Bryso, Huberto ir Struyfo kriterijus. Laplaso skirstinio atveju galingiausias Gelo, Miao ir Gastwirtho kriterijus, o silpniausias – Bryso, Huberto ir Struyfo kriterijus. Šapiro ir Vilko kriterijus galingiausias asimetrinių skirstinių (išskyrus Gompertzo) atvejais. Bryso, Huberto ir Struyfo kriterijus buvo silpniausias ir asimetrinių skirstinių atvejais. Gompertzo skirstinio atveju galingiausias buvo asimetrijos kriterijus [29].

Zhangas ir Wu pristatė suderinamumo kriterijų, paremtą tikėtinumo santykiu ir empirine pasiskirstymo funkcija. Autoriai pirmiausiai palygino visus tris Z_A , Z_C ir Z_K pasiūlytus kriterijus normaliojo skirstinio $N(\mu, \sigma^2)$ atveju. Visi šie trys kriterijai plačiau aptariami 2.1.2 poskyryje. Palyginimui buvo taikomas Monte Karlo metodas su 1 000 000 nepriklausomųjų imčių. Modeliavimui buvo taikomi imties dydžiai $n = 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 40, 50, 70, 100, 150, 200, 300, 500, 1000$. Autoriai nustatė, kad kriterijų galia priklauso nuo μ ir σ parametų. Šią išvadą pagrindė pavyzdžiais [34, 58–60]. Zhango kriterijai [58, 61] buvo palyginti su Šapiro ir Vilko (W), D’Agostino (D), Andersono ir Darlingo (A^2) kriterijais, normaliojo su beta, Stjudento, gama, Veibulo, lognormaliuoju, skirstinių mišinių atvejais. Šiam palyginimui pasirinkti imties dydžiai $n = 10, 20, 30, 50, 70, 100, 150, 200$ ir 300 su reikšmingumo lygmeniu $\alpha = 0,05$. Modeliavimas atliekamas su 10 000 nepriklausomųjų imčių. Autorių gauti modeliavimo rezultatai kiekvienam skirstinio atvejui pateikiami 1.3 paveiksle.



1.3 pav. Kriterijų palyginimas kiekvienam nagrinėtam skirstinių mišinio atvejui

Pagal gautus modeliavimo rezultatus padarytos išvados: Zhango pasiūlyti Z_A , Z_C ir Z_K kriterijai yra labai galingi ir patikimi įvairių nukrypimų nuo normaliojo skirstinio atvejais. Z_A ir Z_C kriterijai paprastai buvo galingesni už W , o galingiausi Z_K ir A^2 . D’Agostino kriterijus yra jautrus eksceso koeficiento kitimui. Bendras šešių kriterijų galios palyginimas apibrėžiamas [34]:

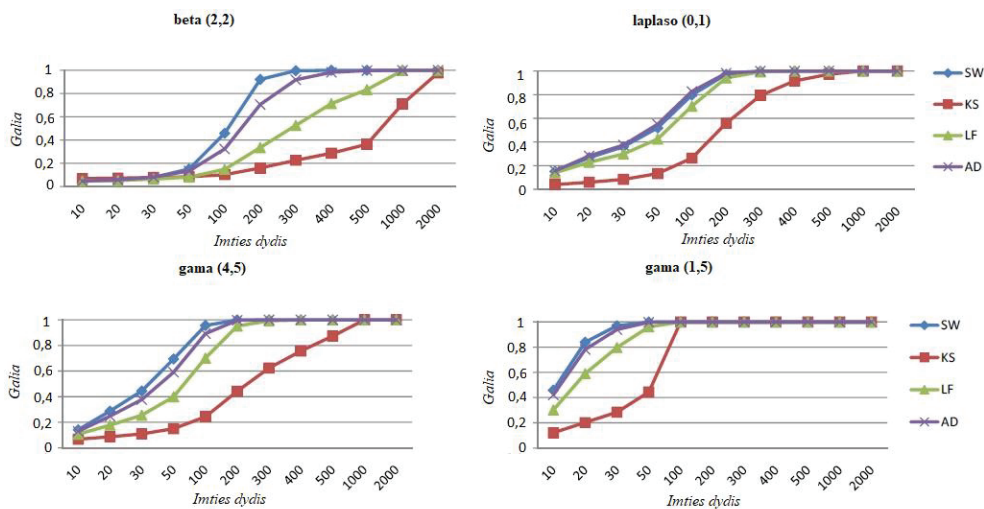
$$\bar{\beta}_{Z_A} > \bar{\beta}_{Z_C} > \bar{\beta}_W > \bar{\beta}_{Z_K} > \bar{\beta}_{A^2} > \bar{\beta}_D;$$

čia $\bar{\beta}_i = 1 - \beta$ yra i -ojo kriterijaus galia.

I. A. Friday su E. F. Runyi palygino chi-kvadrato bei Kolmogorovo ir Smirnovo kriterijus normalumo prielaidos tikrinimui. Šie kriterijai buvo palyginti 50 modeliuotų duomenų rinkinių atvejais, kai imties dydis $n = 100$ su reikšmingumo lygmeniu $\alpha = 0,05$. Tyrimo metu buvo laikoma, kad kai p -reikšmė $< 0,05$, nulinė hipotezė apie duomenų pasiskirstymą pagal normalųjį skirstinį yra atmetama, o priešingu atveju H_0 yra neatmetama. Pagal modeliuotų ir realių duomenų palygintų kriterijų p -reikšmes nustatyta, kad Kolmogorovo ir Smirnovo kriterijus buvo galingesnis už chi-kvadrato kriterijų, nes Kolmogorovo ir Smirnovo kriterijaus p -reikšmė yra didesnė už chi-kvadrato kriterijaus [26].

H. Torabi su kolegomis pasiūlė suderinamumo kriterijų H_n , paremtą dviejų tiesinių pasiskirstymo funkcijų matu. Šis kriterijus buvo palygintas remiantis teoriniais ir empiriniais pasiskirstymų tyrimais. Autoriai įrodė, kad naujas kriterijus pasižymi invariantiškumo ir suderinamumo savybėmis. Atliktas modeliavimo tyrimas siekiant palyginti pasiūlytą kriterijų su 40 kitų mokslininkų kriterijais. Šiam tyrimui buvo pasirinkti simetriniai (standartinis normalusis, Stjudento, dvimatis eksponentinis, beta, Koši), asimetriniai (Gumbelio, beta, chi-kvadrato), skirstiniai intervaluose $(0; \infty)$ ir $(0; 1)$ bei kiti (tolygusis, Johnsono, Turkio) skirstiniai taikant skirtingus imties dydžius ($n = 10, 20, 50, 100, 1000$). Modeliavimas atliekamas su 10 000 nepriklausomųjų imčių taikant reikšmingumo lygmenį $\alpha = 0,05$. Modeliavimui atlikti buvo naudojama R programinė įranga. Pagal gautus modeliavimo rezultatus autoriai nustatė, kad robastinis Jarque ir Beros (RJB) bei Gelio, Miao ir Gastwirtho (SJ) kriterijai galingiausi simetrinų skirstinių atvejais. Vasiceko (TV) bei Alizadeho ir Arghamio (TA) kriterijai galingiausi skirstiniams intervale $(0; 1)$. H_n ir LaRiccia (T_{1n}) kriterijai galingiausi asimetrinių skirstinių atvejais, esant mažiems imties dydžiams [33].

N. M. Razali su Y. B. Wah palygino Šapiro ir Vilko (SW), Kolmogorovo ir Smirnovo (KS), Lilliefors (LF) bei Andersono ir Darlingo (AD) kriterijus. Palygimas atliktas taikant Monte Karlo modeliavimo metodą su skirtingais imties dydžiais $n = 10, 15, 20, 25, 30, 40, 50, 100, 200, 300, 400, 500, 1000, 1500$ ir 2000. Kriterijų galia apskaičiuota su dviem skirtingais reikšmingumo lygmenimis ($\alpha = 0,1$ ir $0,05$). Modeliavimas atliekamas su 50 000 nepriklausomųjų imčių. Kriterijai palyginami 14 skirstinių atvejais iš jų 7 yra simetriniai (tolygusis, beta, 4 Stjudento, Laplaso) ir 7 asimetriniai (3 beta, 2 chi-kvadrato, 2 gama). Tyrimas buvo atliktas standartizuotiems duomenims. Vieni iš pagrindinių autorių rezultatų pateikiami 1.4 paveiksle.



1.4 pav. Kriterijų galios palyginimas beta, Laplaso ir gama skirstinių atvejais

Pagal gautus rezultatus autoriai nustatė, kad iš keturių nagrinėtų kriterijų galingiausias buvo Šapiro ir Vilko kriterijus visiems taikytiems skirstiniams ir imties dydžiams. Tačiau esant mažiems imties dydžiams Šapiro ir Vilko kriterijaus galia buvo gan maža. Silpniausias kriterijus (iš nagrinėtų) buvo Kolmogorovo ir Smirnovo. Andersono ir Darlingo kriterijaus galia buvo panaši į Šapiro ir Vilko kriterijaus [32]. Šiuos rezultatus taip pat patvirtina tyrimai [62–64].

Vienmačių suderinamumo hipotezės kriterijų tyrimai vis dar išlieka aktualūs. XXI a. tyrimai susitelkė ties kriterijais, paremtais N -metrika, empirine pasiskirstymo ir tankio funkcijomis, eksceso ir asimetrijos koeficientais, Vaseršteino metrika ir daugeliu kitų. Kriterijai buvo lyginami remiantis matematiniais įrodymais [5–7, 25, 47, 48] ir Monte Karlo modeliavimo metodu [10, 21, 22, 24–26, 28–30, 32–34, 50, 51, 57–59, 60–64].

1.3. Daugiamačių kriterijų tyrimų apžvalga ir analizė

XX a. pabaigoje pradėti kurti ir nagrinėti daugiamačiai suderinamumo hipotezės tikrinimo kriterijai [37]. Dar ir šiuo metu yra aktualu kurti ir palyginti kriterijus esant skirtingoms sąlygoms (imties dydžiu, sklaida ir kt.). Tačiau daugiamačiai kriterijai yra rečiau taikomi praktikoje dėl jų sudėtingumo. XXI a. šios tematikos tyrimai buvo atliekami [37–46, 52, 65–71]. Mokslininkai kūrė ir lygino daugiamačius kriterijus taikydami Monte Karlo modeliavimo metodą. Daugiamačius kriterijus lygino įvairiems simetriniams ir asimetriniams skirstiniams bei jų mišiniams, skirtingiems imties dydžiams (10–1000), kai dimensija buvo 2–10. Dažniausiai buvo pasirenkamas reikšmingumo lygmuo $\alpha = 0,05$. Tyrėjai dažnai naudojo R programinę įrangą su MVN ir $mvnTest$ bibliotekomis.

N . Henze pristatė daugiamačių kriterijų klasę, priklausančią nuo dimensijos ir glodumo parametro h . Glodumo parametras priklauso nuo dimensijos (d) ir imties dydžio (n), ir yra apibrėžiamas:

$$h = h_d(n) = \frac{1}{\sqrt{2}} \left(\frac{2d}{4} \right)^{\frac{1}{d+4}} \frac{1}{n^{d+4}}.$$

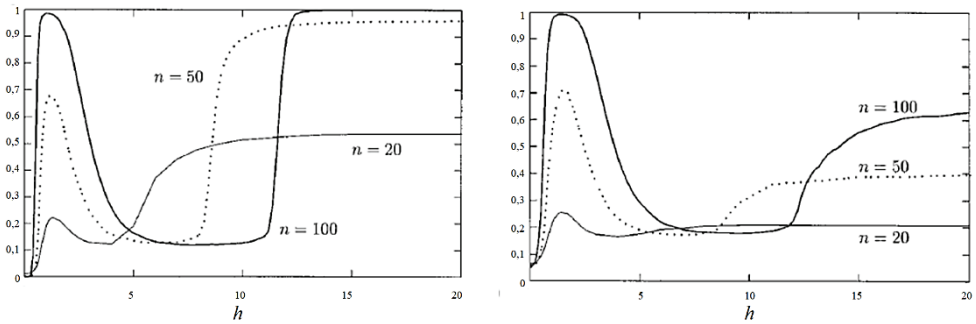
Ši klasė paremta neneigiamuoju funkciniu atstumu. Šiems kriterijams buvo patikrintos ir įrodytos invariantiškumo ir suderinamumo savybės. N. Henze pasiūlytas kriterijus yra apibrėžtas 2.2.3 poskyryje. Pasiūlyti kriterijai palyginti su 8 (tenkinantys invariantiškumo savybę) kitais populiariais daugiamačiais kriterijais, kurie buvo paremti eksceso ir asimetrijos koeficientais, poliniu kintamųjų skaidymu (t. y. $X = U \cdot R$; čia $U = X/\|X\|$ ir $R = \|X\|$), tankiu, apibendrintais kvantiliais ir kt. Taip pat buvo lyginama su BHEP (Baringhauso, Henzės, Eppsio ir Pulley) grupės kriterijais. Modeliavimui taikytos $h = 0,5, 1, 3$ ir $h_d(n)$ parametro reikšmės, $d = 2$ ir $d = 5$ dimensijos ir $n = 20$ ir 50 dydžio imtys. Tyrimui buvo generuojami 32 skirstiniai ir jų modifikacijos. Modeliavimo tyrimas buvo atliktas taikant 50 000 nepriklausomųjų imčių. Kriterijai buvo lyginami pagal galią. Modeliavimo rezultatai parodė, kad parinkus glodinimo parametą $h = 0,5$ pasiūlytas kriterijus buvo galingiausias sunkias uodegas turintiems skirstiniams [39, 40].

Henze nagrinėjo BHEP kriterijų klasę. Šiame tyrime buvo panaudojama tiesinių funkcijų silpna konvergavimo teorija Frecheto⁷ erdvėje $C(\mathbb{R}^d)$. Taikant šią teoriją buvo gauta:

- nauja ribinė BHEP kriterijaus statistikos $T_{n,h}$ pasiskirstymo išraiška $C(\mathbb{R}^d)$ erdvėje;
- jungtinis ribinis $T_{n,h}$ pasiskirstymas kelioms β reikšmėms;
- $T_{n,h}$ galios palyginimas su alternatyvomis.

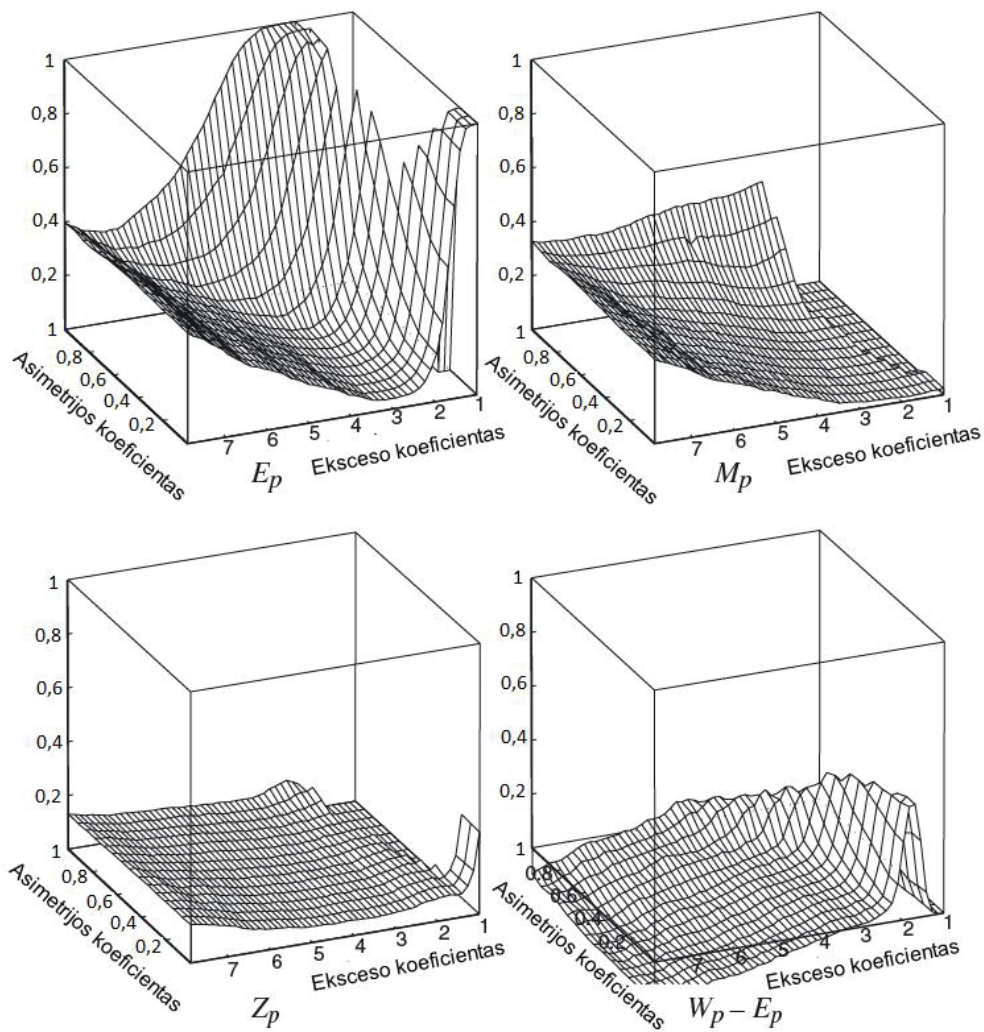
Atliktas modeliavimo tyrimas skaičiuojant $T_{n,h}$ galią, kai $h = 0,1; 0,5; 1, 3$, $n = 20, 50, 100$, $\alpha = 0,1$ ir $0,05$ bei $d = 2, 3, 5, 10$. Monte Karlo metodas taikytas 10 000 nepriklausomųjų imčių. Autoriai pateikia nagrinėto $T_{n,h}$ galios kitimą skirtingiems imties dydžiams didinant glodinimo parametro reikšmę (žr. 1.5 pav.). Iš 1.5 pav. pastebėta, kad tolygiojo skirstinio atveju $T_{n,h}$ galia didelė, kai h įgyja mažas reikšmes (tarp 0 ir 1) ir didesnes už 12. Normaliojo skirstinio mišinio atveju $T_{n,h}$ galia yra didelė, kai $0 < h < 2$ [65].

⁷ Frecheto erdvės yra specialios topologinės vektorinės erdvės. Tai Banacho erdvės apibendrinimas (normuotos vektorinės erdvės, kurios yra baigtinės metrikos atžvilgiu). Visos Banacho ir Hilberto erdvės yra Frecheto erdvės.



1.5 pav. $T_{n,h}$ galios palyginimas ($d = 5$) tolygiojo $[0,1]$ ir normaliojo skirstinio mišinio atvejais

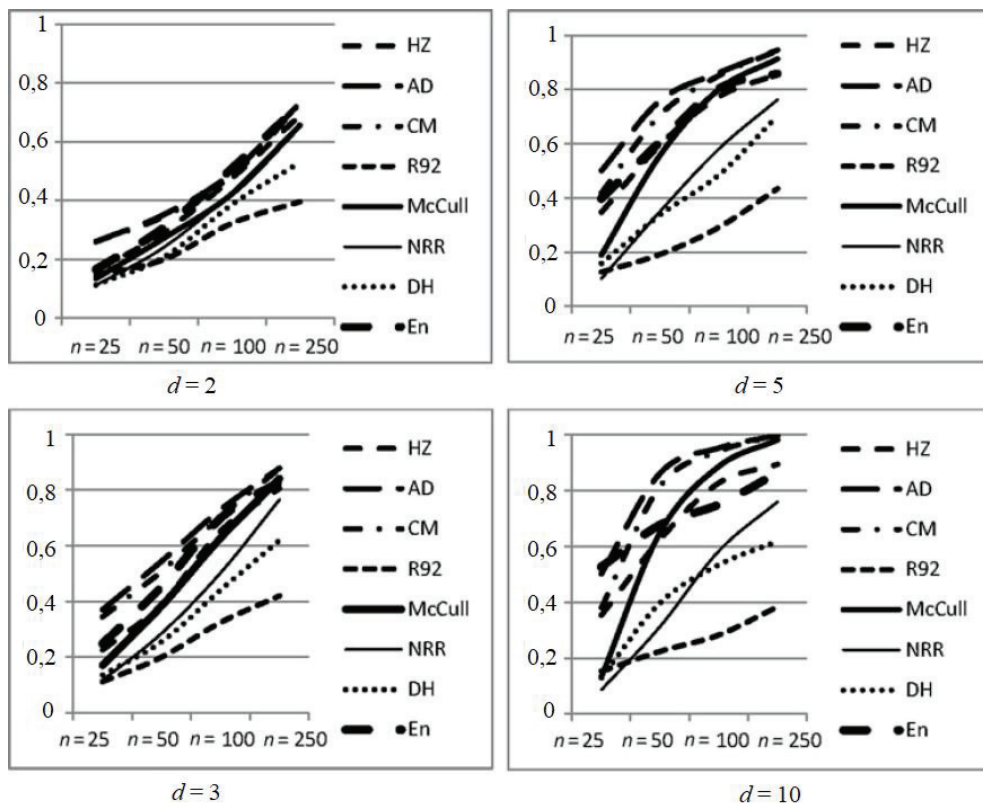
J. A. Doornik su H. Hansen pasiūlė naują *Omnibus* kriterijaus versiją, paremtą eksceso ir asimetrijos koeficientais [38]. Doorniko kriterijus (E_p) tinkamas naudoti vienmačiams ir daugiamačiams duomenims. A. Doorniko kriterijus yra apibrėžtas 2.1.3 poskyryje. Šį kriterijų palygino su Bowmano ir Shentono (Q_p), Mudholkaro, McDermotto ir Srivastavos (Z_p), Mardijos (M_p), Šapiro ir Vilko (W_p) daugiamačiais kriterijais taikydami imties dydį ($n = 50$), dimensiją ($d = 2$ ir 3) ir standartinį normalųjį skirstinį. Autorių gauti modeliavimo rezultatai trimačiais (3D) vaizdais pateikti 1.6 paveiksle. Iš 1.6 pav. nustatyta, kad E_p galia buvo didžiausia, o Z_p – mažiausia [38].



1.6 pav. Kriterijų galios palyginimas ($d = 5, n = 50$) standartinio normaliojo skirstinio atveju

O. Kesemen su kolegomis pasiūlė daugiamačių Kolmogorovo ir Smirnovo kriterijų, paremtą skirtumu tarp empirinės pasiskirstymo ir teorinės pasiskirstymo funkcijų. Modeliavimui taikytos $d = 2, 3, 5, 10$ dimensijos ir $n = 10, 30, 50, 100, 200, 500$ dydžio imtys. Modeliavimo tyrimas buvo atliktas taikant 100 000 nepriklausomųjų imčių. Tyrimui buvo generuojami simetriniai ir asimetriniai skirstiniai. Modeliavimo rezultatai parodė, kad pasiūlytas kriterijus buvo paprastesnis ir pranašesnis už kitus lygintus kriterijus. Kriterijai lyginti pagal galią. Pranašumas išryškėja dėl kur kas paprastesnio statistikos skaičiavimo bei itin paprasto pritaikymo įvairaus dydžio duomenims. Pasiūlytas kriterijus didžiausią galią turėjo asimetrinių skirstinių atvejais [41].

V. Voinov su kolegomis pristatė naują invariantišką ir suderintą daugiamačių kriterijų, kuris yra paremtas Karhunen ir Loève transformacija⁸ daugiamačiams duomenims. Pasiūlytas Nikulino, Rao ir Robsono (NRR) kriterijus buvo lyginamas su Doorniko ir Hanseno (DH) [38], Henzės ir Zirklero (HZ) [39], Andersono ir Darlingo (AD) [40], Kramerio von Mises (CM) [40], Roystono (R92) [45], McCullocho (McCull) [73] ir energijos (En) [74] kriterijais. Modeliavimui buvo taikytos $d = 2, 3, 5, 10$ dimensijos ir $n = 25, 50, 100, 250$ dydžio imtys. Modeliavimo tyrimas atliktas taikant 20 000 nepriklausomųjų imčių. Tyrėjai naudojo *R* programinę įrangą su *MVN*, *mvnTest* ir kitomis bibliotekomis. Autorių gauti modeliavimo rezultatai pateikti 1.7 paveiksle. Iš šio paveikslo nustatyta, kad naujo kriterijaus galia beveik nepriklauso nuo dimensijos dydžio. Iš 1.7 pav. taip pat nustatyta, kad visi kriterijai gana galingai veikia, išskyrus NRR, DH ir R92 kriterijus. Klasikiniai AD ir CM kriterijai buvo galingiausi [66].

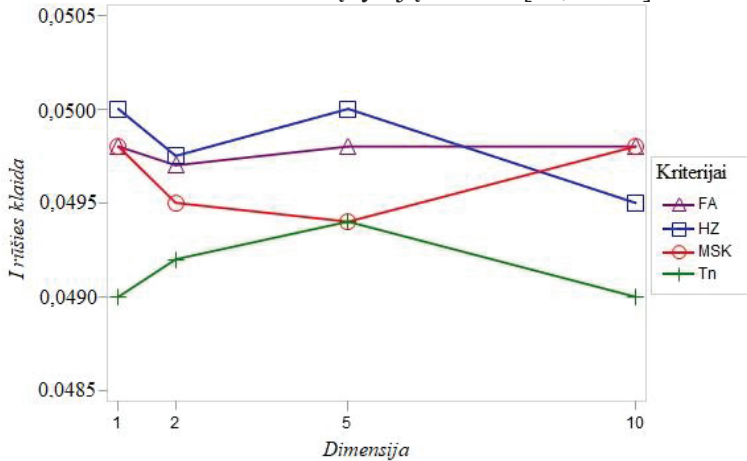


1.7 pav. Aštuonių daugiamačių kriterijų empirinės galios palyginimas

M. Zhou su Y. Shao pristatė naują daugiamačių kriterijų, kuris lengvai suprantamas biomedicinos tyrėjams. Pasiūlytas kriterijus buvo realizuotas *R*

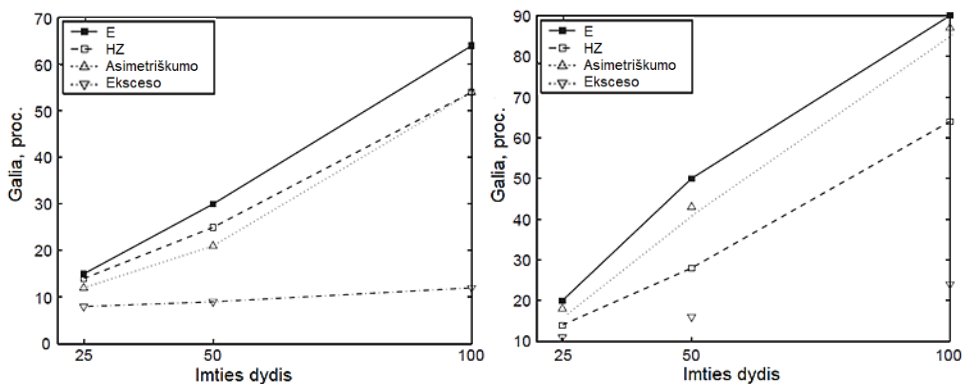
⁸ Karhunen ir Loève transformacija apibrėžiama kaip tiesinė transformacija, kurios baziniai vektoriai yra duomenų kovariacijos matricos tikriniai vektoriai.

funkcijoje „shapiro.test“. Šis kriterijus konkurencingas tiek vienmačiams, tiek ir daugiamačiams duomenims. Modeliavimo būdu pasiūlytas kriterijus buvo lygintas su Mardia (MSK) [75], Fattorini (FA) [76], Henzės ir Zirklero (HZ) [39] kriterijais. Modeliavimui taikytos $d = 1, 2, 5, 10$ dimensijos ir $n = 50$ dydžio imtis. Autoriai parinko labai daug skirstinių ir jų mišinių. Autorių gauti rezultatai pateikiami 1.8 paveiksle. Iš 1.8 pav. nustatyta, kad pasiūlyto kriterijaus statistika T_n pranašesnė už kitų kriterijų. Kriterijai taip pat pritaikyti realiems biomedicininiais duomenų rinkiniams. Gauti rezultatai atitiko kitų tyrėjų išvadas [46, 77, 78].



1.8 pav. Kriterijų palyginimas pagal I rūšies klaidos tikimybę, kai $n = 50$.

G. J. Szekely su M. L. Rizzo pasiūlė naują ciklinę daugiamačių kriterijų klasę, kuri tenkina invariantiškumo ir suderinamumo savybes. Šie kriterijai paremti Euklido atstumu ir imties charakteristikomis. Šis kriterijus yra apibrėžtas 2.2.3 poskyryje. Taikant Monte Karlo modeliavimo metodą atliktas aštuonių kriterijų palyginimas. Palyginimui parinktos dvi kriterijų kategorijos – kriterijai paremti asimetrijos bei eksceso koeficientais ir empirine pasiskirstymo funkcija. Modeliavimui taikytos $d = 2, 3, 5, 10$ dimensijos ir $n = 25, 50, 100$ dydžio imtys. Autorių gauti rezultatai normaliojo skirstinio mišiniams ($0,5N(0,1) + 0,5N(2,1)$ ir $0,9N(0,1) + 0,1N(2,1)$) pateikiami 1.9 paveiksle. Gauti rezultatai parodė, kad pasiūlytas kriterijus (E) yra esamų kriterijų konkurentas, tačiau labai jautrus sunkių uodegų skirstiniams [67].



1.9 pav. Kriterijų palyginimas ($d = 5, n = 50$)

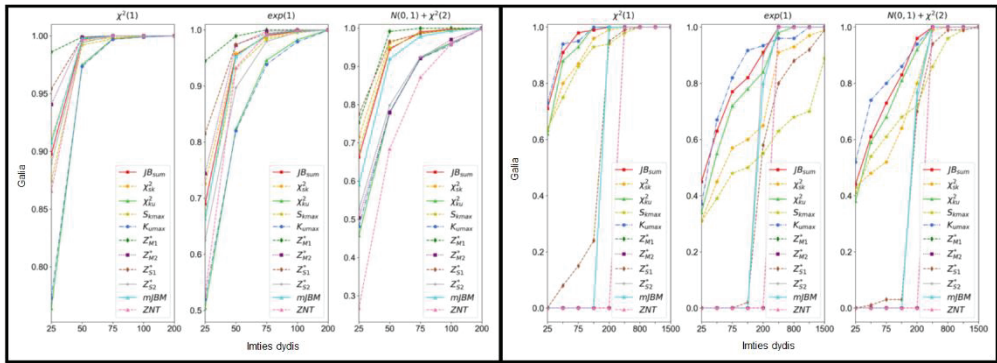
R. Enomoto su kolegomis pasiūlė duomenis normalizuojančią daugiamačio kriterijaus statistiką ZNT, kuri yra apibrėžiama [68]:

$$ZNT = \sqrt{\frac{n}{2\beta} \left[\frac{\exp[\delta\{k - \beta\}] - 1}{\delta} + \frac{2\beta(1 - 2\delta)}{n} \right]};$$

čia $\beta = d(d + 2)$, $\delta = -\frac{d+8}{3d(d+2)}$. ZNT statistika tenkina invariantiškumo savybę.

ZNT statistikos eksceso ir asimetrijos koeficientai yra artimi standartiniam normaliajam pasiskirstymui dideliems imčių ir dimensijos dydžiams. Ši statistika buvo palyginta su Mardia eksceso (ZM*) bei asimetrijos (ZM), ir transformuotu Wilsono bei Hilferto (WH) kriterijais. Kriterijai buvo lyginti taikant sunkių ir lengvų uodegų, simetriinių, asimetriinių ir dvimodalinius skirstinius. Pasirinktos imtys buvo $n = 20, 30, 40, 50, 100, 200, 400$ dydžio ir $d = 2, 3, 4, 5, 7, 10$ dimensijos. Modeliavimo tyrimas atliktas taikant 1 000 000 nepriklausomųjų imčių. Kriterijai lyginti pagal galią. Autoriai modeliavimo būdu įrodė, kad ZNT statistika galingiausia skirstiniams su lengvomis uodegomis. Taip pat nustatyta, kad ZNT statistika buvo galingiausia dideliems imčių dydžiams [68].

Y. Song su X. Zhao pasiūlė taikyti pagrindinių komponentių analizę duomenų dimensijos sumažinimui ir po to kiekvieną komponentę įvertinti modifikuota Jarques ir Beros (JB) kriterijaus statistika. Atliekant modeliavimo tyrimą autorių pasiūlytas kriterijus buvo lyginamas su 10 dažniausiai literatūroje aprašomų kriterijų. Naudoti duomenys turėjo nuo 5 iki 200 kintamųjų. Kriterijai buvo lyginami pagal empirinę galią (t. y. klaidingi atvejai padalijami iš teisingų). Autorių gauti rezultatai pateikiami 1.10 paveiksle. Iš 1.10 pav. nustatyta, kad JB kriterijus konkurencingas kitų kriterijų atžvilgiu, nes jo empirinė galia didesnė už kitų, o ir modeliavimo rezultatai buvo stabilesni. Taigi modifikuotą Jarques ir Beros kriterijų galima efektyviai taikyti mažos ir didelės dimensijos duomenims [69].



1.10 pav. Kriterijų palyginimas 5-mačiams ir 200-mačiams duomenims

P. Dörr su kolegomis pristatė ir ištyrė naują invariantiškų bei suderintų daugiamačių kriterijų klasę. Pasiūlyti kriterijai yra paremti daline diferencialine lygtimi [70, 71]:

$$U_{n,a} = n \int_{\mathbb{R}^d} |\Delta \psi_n(t) - (\|t\|^2 - d)\psi_n(t)|^2 w_a(t) dt \quad (1.3)$$

$$T_{n,a} = n \int |\Delta \psi_n(t) - \Delta \psi(t)|^2 w_a(t) dt \quad (1.4)$$

čia $\Delta \psi_n(t) = -\frac{1}{n} \sum_{j=1}^n \|Y_{n,j}\| \exp(it^T Y_{n,j})$, $t \in \mathbb{R}^d$ yra empirinė charakteristinė funkcija, $Y_{n,j} = S_n^{-\frac{1}{2}}(X_j - \bar{X}_n)$, $j = 1, \dots, n$, $a > 0$ yra fiksuota konstanta, Δ – Laplaso operatorius, $w_a(t) = \exp(-a\|t\|^2)$, $t \in \mathbb{R}^d$. Atlikus modeliavimo tyrimą ir palyginus $U_{n,a}$ ir $T_{n,a}$ kriterijus, rekomenduojama naudoti $T_{n,a}$, kuris buvo galingesnis atsižvelgiant į parametro a reikšmę. Tačiau $U_{n,a}$ kriterijus yra konkurencingas kitų tradicinių vienmačių ir daugiamačių suderinamumo kriterijų atžvilgiu [70].

Pagal apžvelgtus daugiamačių suderinamumo hipotezės kriterijų tyrimus nustatyta, kad vieni galingiausių kriterijų yra paremti eksceso ir asimetrijos koeficientais, neneigiamuoju funkciniu atstumu, empirine pasiskirstymo funkcija, tankiu, Euklido atstumu ir imties charakteristikomis. Pasiūlytos įvairios duomenų transformacijos padeda didinti kriterijų galią.

1.4. Literatūros apžvalgos bei analizės išvados ir apibendrinimas

Atlikus suderinamumo hipotezės kriterijų tyrimų apžvalgą galima padaryti išvadas.

1. Suderinamumo hipotezės apie duomenų pasiskirstymą pagal normalųjį skirstinį tikrinimas yra plačiai naudojamas duomenų moksle. Tai viena iš pagrindinių duomenų analizės metodų prielaidų. Pagal prielaidos rezultatą pasirenkamas analitinių metodų tipas.

2. Apžvelgus vienmačių suderinamumo kriterijų tyrimus nustatyta, kad Šapiro ir Vilko kriterijus bei jo modifikacijos yra galingiausi simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms. Modeliavimo tyrimui atrinktas 41 galingiausias vienmatis suderinamumo kriterijus (žr. 2.1 sk.), kurie paremti empirine pasiskirstymo funkcija, rangine statistika, momentais ir kt.
3. Atlikus daugiamačių suderinamumo kriterijų tyrimų apžvalgą nustatyta, kad klasikiniai Andersono ir Darlingo bei Kramerio ir von Mises kriterijai yra vieni galingiausių. Modeliavimo tyrimui atrinkta 12 galingiausių daugiamačių kriterijų, kurie paremti rangine statistika, momentais ir kt.

Apibendrinant apžvelgtus literatūros šaltinius, numatyta suderinamumo vienmačių ir daugiamačių kriterijų palyginimui naudoti:

- 12 (iš jų 6 simetriniai ir 6 asimetriniai) vienmačių;
- 12 (iš jų 5 simetriniai, 5 asimetriniai ir kt.) daugiamačių skirstinių (taikant įvairius parametrų rinkinius);
- modifikuotus normaliuosius skirstinius bei jų mišinius.

Taip pat atlikti generuotų duomenų standartizavimą. Kriterijų palyginimui numatyta vertinti empirinę galią imties dydžiams $n = 32, 64, 128, 256, 512$ ir 1024 naudojant reikšmingumo lygmenį $\alpha = 0,05$. Modeliavimo tyrimą atlikti taikant Monte Karlo metodą 1 000 000 nepriklausomųjų imčių [25, 28, 34, 68].

II. METODOLOGIJA

Šioje dalyje apibrėžiami R programiniame pakete realizuoti suderinamumo kriterijai, kurie pagal jų konstravimo ypatumus yra suskirstyti į grupes. Aprašyti pasiūlyti suderinamumo hipotezės tikrinimo vienmatis ir daugiamatis kriterijai, kurie buvo realizuoti SAS. Pristatomi tyrime naudojami duomenų pasiskirstymai ir kriterijų galios apskaičiavimo procedūra. Aprašomas modeliavimo procesas Monte Karlo metodu. Pateikiamos šią dalį apibendrinančios išvados.

2.1. Vienmačiai suderinamumo hipotezės tikrinimo kriterijai

Suderinamumo kriterijai, kurie skirti normalumo prielaidos tikrinimui, tikrina nulinę ir alternatyvią hipotezes:

$$\begin{cases} H_0: \text{duomenys yra pasiskirstę pagal normalųjį skirstinį,} \\ H_A: \text{duomenys nėra pasiskirstę pagal normalųjį skirstinį.} \end{cases}$$

Laikoma, kad $X = (X_1, X_2, \dots, X_n)$ žymi atsitiktinio dydžio n imtį, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ žymi ranginės statistikos imtį⁹, \bar{X} ir $\hat{\sigma}^2$ žymi imties vidurkį ir dispersiją, kurie apibrėžiami:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_{(i)}, \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{(i)} - \bar{X})^2. \quad (2.1)$$

2.1.1 Vienmačiai kriterijai, paremti empirine pasiskirstymo funkcija

Šiame poskyryje pristatomi vienmačiai kriterijai, paremti pasiskirstymo funkcija. Standartizuotas vidurkio ir sklaidos atžvilgiu dydis apibrėžiamas:

$$Z_{(i)} = \frac{X_{(i)} - \bar{X}}{\hat{\sigma}}; \quad (2.2)$$

čia \bar{X} ir $\hat{\sigma}$ – imties vidurkis ir standartinis nuokrypis.

- **Kolmogorovo ir Smirnovo (KS)**

1933 m. Kolmogorovas ir Smirnovas pasiūlė naują kriterijų, kuris paremtas pasiskirstymu, apibrėžtu žinomais parametrais. Kolmogorovo ir Smirnovo kriterijus taikomas tikrinti, ar du pasiskirstymai yra vienodi. Kolmogorovo ir Smirnovo kriterijus yra apibrėžiamas [2]:

$$D_n = \sup_{x \in X} |\hat{F}_1 - F_2|; \quad (2.3)$$

čia \hat{F}_1 – empirinė pasiskirstymo funkcija, F_2 – teorinė pasiskirstymo funkcija.

⁹ Ranginės statistikos imtys yra tokios, kurių reikšmės išdėstytos didėjimo tvarka.

- **Andersono ir Darlingo (AD)**

1952 m. Andersonas ir Darlingas sukūrė dar vieną Kolmogorovo ir Smirnovo kriterijaus atmainą. Šis kriterijus yra galingesnis už Kolmogorovo ir Smirnovo kriterijų [79]. Andersono ir Darlingo kriterijaus statistika apibrėžiama [3]:

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} (\ln(Z_{(i)}) + \ln(1 - Z_{(n+1-i)})). \quad (2.4)$$

- **Kramerio ir von Mises (CVM)**

1962 m. T. Andersonas pasiūlė Kramerio ir von Mises kriterijų. Kramerio ir von Mises kriterijus yra alternatyva Kolmogorovo ir Smirnovo kriterijui. Šio kriterijaus statistika apibrėžiama [4]:

$$CM = \frac{1}{12n} + \sum_{i=1}^n \left(Z_{(i)} - \frac{2i-1}{2n} \right)^2. \quad (2.5)$$

- **Lillieforso (LF)**

1967 m. Lillieforsas išplėtojo Kolmogorovo ir Smirnovo (KS) kriterijų. Lillieforso kriterijus tinka, kai pasiskirstymas nėra tiksliai nurodytas. Lillieforso kriterijaus statistika apibrėžiama [80]:

$$L = \sup_x |Z_{(i)} - F_1(X)|. \quad (2.6)$$

- **Gleno, Leemiso ir Barro (GLB)**

2001 m. Glenas, Leemisas ir Barras išplėtė Kolmogorovo ir Smirnovo bei Andersono ir Darlingo kriterijus ir taip sudarė Gleno, Leemiso ir Barro kriterijų. Šio kriterijaus statistika yra apibrėžiama tiesine logaritmų kombinacija [81]:

$$P_S = -n - \frac{1}{n} \cdot \sum_{i=1}^n [(2n+1-2i) \cdot \ln(p_{(i)}) + (2i-1) \cdot \ln(1-p_{(i)})]; \quad (2.7)$$

čia $p_{(i)}$ – vektoriaus p elementai, kuriuose yra didėjimo tvarka surikiuoti ranginės statistikos kvantiliai. Vektoriaus p elementus galima gauti apibrėžiant vektorių u , kurio $u_{(i)} = \Phi(z_{(i)})$. Vektoriaus u elementai išdėstyti didėjimo tvarka. Vektoriaus p elementus galima nustatyti žinant, kad $u_{(i)}$ seka yra pasiskirsčiusi pagal $Beta(i; n-i+1)$ skirstinį [82]. Šios statistikos reikšmė yra didelė, kai bent viena $p_{(i)}$ reikšmė yra arti 0 arba 1. Laikoma, kad $p_{(1)}$ – mažiausias vektoriaus elementas, o $p_{(n)}$ – didžiausias vektoriaus elementas. Didėjant statistikos reikšmei P_S , didėja tikimybė atmesti nulinę hipotezę.

- **Zhang ir Wu (ZW1 ir ZW2)**

2005 m. Zhangas ir Wu pristatė Z_C ir Z_A suderinamumo hipotezės tikrinimo kriterijus. Z_C ir Z_A statistikos yra panašios į Kramerio ir von Mises bei Andersono ir

Darlingo kriterijų statistikas. Zhang ir Wu pasiūlyti kriterijai yra galingesni [58]. Z_C ir Z_A kriterijų statistikos apibrėžiamos [34]:

$$Z_C = \sum_{i=1}^n \left[\log \frac{F_0(X_{(i)})^{-1} - 1}{\frac{n-0,5}{i-0,75} - 1} \right]^2 \quad Z_C = \sum_{i=1}^n \left[\ln \frac{\Phi(Z_{(i)})^{-1} - 1}{\frac{n-0,5}{i-0,75} - 1} \right]^2, \quad (2.8)$$

$$Z_A = - \sum_{i=1}^n \left[\frac{\log F_0(X_{(i)})}{n-i+0,5} + \frac{\log[1-F_0(X_{(i)})]}{i-0,5} \right];$$

čia $F_0(X_{(i)}) = (i - 0,5)/n$, $\Phi(X_i; \bar{X}; \hat{\sigma}^2)$ yra normaliojo skirstinio funkcija.

2.1.2 Vienmačiai kriterijai, paremti regresija ir koreliacija (rangine statistika)

Šiame poskyryje apibrėžiami vienmačiai kriterijai, paremti c_i konstanta, medianomis ir jų transformacijomis, svorio w_i parametru, dviejų standartinių nuokrypio įverčių $\hat{\sigma}_1, \hat{\sigma}_2$ santykio logaritmais ir kt. Konstantų vektorius $c_{(i)}$ gaunamas [36]:

$$c_{(i)} = (c_1, c_2, \dots, c_n) = \frac{z^T \Sigma^{-1}}{((z^T \Sigma^{-1})(\Sigma^{-1} z))^{1/2}}; \quad (2.9)$$

čia $z = (z_1, \dots, z_n)^T$, z_1, \dots, z_n – tikėtinosios nepriklausomųjų vienodai pasiskirsčiusių atsitiktinių santykių skalės dydžių standartinio normaliojo pasiskirstymo statistikos reikšmės, Σ – statistikų kovariacinės matricos.

Laikoma, kad $w_{(1)}, \dots, w_{(n)}$ yra statistikos įvertintos medianos, kai $w_{(i)}$ yra gaunama [83]:

$$w_{(i)} = \begin{cases} 1 - 0.5\left(\frac{1}{n}\right) & i = 1, \\ \frac{(i - 0.3175)}{(n + 0.365)} & 1 < i < n, \\ 0.5\left(\frac{1}{n}\right) & i = n. \end{cases}$$

- **Šapiro ir Vilko (SW)**

1965 m. Šapiras ir Vilkas sudarė originalų kriterijų. Šis kriterijus – pirmasis, kuriuo galima nustatyti nukrypimus nuo normaliojo skirstinio taikant eksceso arba asimetryjos koeficientus. Šapiro ir Vilko kriterijaus statistika apibrėžiama [36]:

$$W = \frac{(\sum_{i=1}^n c_i X_{(i)})^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}; \quad (2.10)$$

čia $X_{(i)}$ – variacinės eilutės i -tasis imties narys, konstantos c_i apibrėžiamos (2.9) formulėje.

- **D'Agostino (DA)**

1971 m. D'Agostinas pristatė Šapiro ir Vilko kriterijaus praplėtimą. D'Agostino kriterijus yra paprastesnis, nes jame nebereikia apibrėžti svorio vektoriaus a . Šio kriterijaus statistika apibrėžiama [84]:

$$D = \frac{\sum_{i=1}^n (i - (n + 1)/2) \cdot X_{(i)}}{n^2 \cdot \hat{\sigma}}. \quad (2.11)$$

- **Šapiro ir Francijos (SF)**

1972 m. Šapiras ir Francija supaprastino Šapiro ir Vilko kriterijų, kurio skaičiavimo laikas kur kas trumpesnis. Kriterijaus statistika apibrėžiama [85]:

$$W_{SF} = \frac{\sum_{i=1}^n (X_{(i)} - \bar{X})(z_i - \bar{z})}{\sqrt{(\sum_{i=1}^n (X_{(i)} - \bar{X})^2) (\sum_{i=1}^n (z_i - \bar{z})^2)}}. \quad (2.12)$$

Šio kriterijaus statistikos reikšmės yra nuo 0 iki 1. Kai reikšmė artima 1, tai parodo, kad duomenys yra pasiskirstę pagal normalųjį skirstinį, o reikšmė, artima 0, parodo, kad duomenys nėra pasiskirstę pagal normalųjį skirstinį.

- **Fillibeno (Filli)**

1975 m. Fillibenas apibrėžė tikimybinį koreliacijos koeficientą r kaip suderinamumo hipotezės kriterijų. Koreliacijos koeficientas yra apibrėžtas tarp imties statistikos ir teorinės statistikos įvertintų medianų. Pagal teorinės statistikos įvertintas medianas gaunama transformacija $\mathcal{M}_{(i)} = \Phi^{-1}(\varpi_{(i)})$. Tada koreliacijos koeficientas r apibrėžiamas [83]:

$$r = \frac{\sum_{i=1}^n X_{(i)} \cdot \mathcal{M}_{(i)}}{\sqrt{\sum_{i=1}^n \mathcal{M}_{(i)}^2} \cdot \sqrt{(n-1) \cdot \hat{\sigma}^2}}. \quad (2.13)$$

Didėjant statistikos reikšmei r , didėja tikimybė atmesti nulinę hipotezę.

- **Cheno ir Šapiro (ChenS)**

1995 m. Chenas ir Šapiras pristatė alternatyvų kriterijų, paremtą normalizuotais atstumais ir apibrėžiamą [86]:

$$CS = \frac{1}{(n-1) \cdot s} \sum_{i=1}^{n-1} \frac{X_{(i+1)} - X_{(i)}}{\varpi_{i+1} - \varpi_i}; \quad (2.14)$$

čia $\varpi_i = \Phi^{-1} \left[\frac{i-0,375}{n+0,25} \right]$. Didėjant statistikos reikšmei CS , didėja tikimybė atmesti nulinę hipotezę.

- **Modifikuotas Šapiro ir Vilko (SWRG)**

1997 m. Rahmanas ir Govindarajulas pasiūlė modifikuotą Šapiro ir Vilko kriterijų (W_{RG}). Šis kriterijus yra paprasčiau apskaičiuojamas ir remiasi nauju svorio w apibrėžimu. Modifikuoto Šapiro ir Vilko kriterijaus statistikoje didelėms

reikšmėms priskiriami didesni svoriai nei Šapiro ir Vilko kriterijaus atveju. Taigi kiekvieno naujo svorio elementas w_i tampa [87]:

$$w_i = -(n+1)(n+2)\phi(t_i)[t_{i-1}\phi(t_{i-1}) - 2t_i\phi(t_i) + t_{i+1}\phi(t_{i+1})];$$

čia laikoma, kad $t_0\phi(t_0) = t_{n+1}\phi(t_{n+1}) = 0$. Mažėjant statistikos reikšmei W_{RG} , didėja tikimybė atmesti nulinę hipotezę [87].

- **Zhango (ZQ, ZQstar, ZQQstar)**

1999 m. Zhangas pristatė Q kriterijų, paremtą dviejų standartinių nuokrypio įverčių $\hat{\sigma}_1, \hat{\sigma}_2$ santykio logaritmais $Q = \ln(\hat{\sigma}_1/\hat{\sigma}_2)$. Įverčiai $\hat{\sigma}_1$ ir $\hat{\sigma}_2$ yra gaunami: $\hat{\sigma}_1 = \sum_{i=1}^n a_i X_{(i)}$ ir $\hat{\sigma}_2 = \sum_{i=1}^n b_i X_{(i)}$; čia i -osios eilės tiesiniai koeficientai a_i ir b_i , apskaičiuojami [88]:

$$a_i = [(u_i - u_1)(n-1)]^{-1}, \text{ kai } i \neq 1, a_1 = \sum_{i=2}^n a_i,$$

$$b_i = \begin{cases} -b_{n-i+1} = [(u_i - u_{i+4})(n-4)]^{-1}, & i = 1, \dots, 4 \\ (n-4)^{-1} \cdot [(u_i - u_{i+4})^{-1} - (u_{i-4} - u_i)^{-1}], & i = 5, \dots, n-4 \end{cases};$$

čia i -oji tikėtina standartinio normaliojo paskirstymo pozicinės statistikos reikšmė, $u_i = \Phi^{-1}[(i - 0,375)/(n + 0,25)]$.

Zhangas taip pat pristatė alternatyvią statistiką Q^* sukeičiančią i -ąją pozicinę statistiką $X_{(i)} = -X_{(n-i+1)}$. Tada $Q^* = \ln(\hat{\sigma}_1^*/\hat{\sigma}_2^*)$, kai $\hat{\sigma}_1^* = \sum_{i=1}^n a_i X_{(n-i+1)}$ ir $\hat{\sigma}_2^* = \sum_{i=1}^n b_i X_{(n-i+1)}$. Pagal Q ir Q^* kriterijus nulinė hipotezė yra atmetama, kai statistikų reikšmės yra mažos arba didelės, kai kriterijaus kritinė sritis yra dvipusė [88].

Be jau minėtų kriterijų, Zhangas dar pasiūlė $Q - Q^*$ kriterijų, kai Q ir Q^* yra nepriklausomieji. Taikant šį kriterijų nulinė hipotezė yra atmetama su reikšmingumo lygmeniu α , kai bent vieno kriterijaus atveju yra nustatoma, kad nulinę hipotezę reikia atmesti su reikšmingumo lygmeniu $\alpha/2$ [88].

Q ir Q^* apytiksliai yra pasiskirstę pagal normalųjį skirstinį. 2007 m. Hwangas ir Weinas įrodė, kad šie kriterijai galingesni, kai yra paremti empiriniu pasiskirstymu [89].

- **Barrio, Cuestos, Albertos, Matrano ir Rodriguezo (BCMR)**

1999 m. Barrio, Cuestos, Albertos, Matrano ir Rodríguezas pasiūlė naują kriterijų $BCMR$, kuris yra paremtas L_2 -Wassersteino atstumu. Šio kriterijaus statistika apibrėžiama [90]:

$$BCMR = \frac{m_2 - \left[\sum_{i=1}^n X_{(i)} \cdot \int_{(i-1)/n}^{i/n} \Phi^{-1}(t) dt \right]^2}{m_2}; \quad (2.15)$$

čia m_2 – imties standartizuotas antrasis momentas, o skaitiklis žymi L_2 – Wassersteino atstumą. Didėjant statistikos reikšmei $BCMR$, didėja tikimybė atmesti nulinę hipotezę.

- **Coino (Coin)**

2008 m. D. Coinas pasiūlė kriterijų, paremtą polinomine regresija ir skirtą nustatyti simetrinių skirstinių grupės pasiskirstymus. Šio kriterijaus modelio tipas [91]:

$$Z_{(i)} = \eta_1 \cdot u_i + \eta_3 \cdot u_i^3; \quad (2.16)$$

čia η_1 ir η_3 yra regresijos lygties (angl. *fitting*) parametrai, u_i – tikėtinioji standartinio normaliojo pasiskirstymo pozicinės statistikos reikšmė. Tai lemia, kad η_3 reikšmė skiriasi nuo nulio, kai nagrinėjamas pasiskirstymas yra simetriškas, bet nėra normalusis. Coinas pasiūlė naudoti η_3^2 statistiką, tikrinant normalųjį pasiskirstymą. Tada nulinė hipotezė yra atmetama, kai η_3^2 yra didelės reikšmės.

2.1.3 Momentais paremti vienmačiai kriterijai

Šiame poskyryje apžvelgiami kriterijai, paremti eksceso ir asimetrijos koeficientais. Šie koeficientai yra apibrėžiami:

$$s = \frac{m_3}{\sqrt{m_2^3}}, \quad k = \frac{m_4}{m_2^2}; \quad (2.17)$$

čia $m_i = \frac{1}{n} \sum_{j=1}^n (X_{(i)} - \bar{X})^j$ yra i -asis centrinis momentas.

- **D'Agostino ir Pearsono (DAP)**

1973 m. D'Agostinas ir Pearsonas pasiūlė naują suderinamumo kriterijų. D'Agostino ir Pearsono kriterijaus statistika apibrėžiama [92]:

$$DP = Z^2(s) + Z^2(k); \quad (2.18)$$

$Z(s)$ ir $Z(k)$ yra pasiskirstę pagal normalųjį skirstinį. DP statistika yra asimptotiškai artima chi-kvadrato pasiskirstymui su dviem laisvės laipsniais, kai imtis turi normalųjį pasiskirstymą.

- **Jarqués ir Beros (JB)**

1987 m. Jarque ir Bera pasiūlė kriterijų, paremtą normuotais asimetrijos ir eksceso koeficientais. Šio kriterijaus statistika apibrėžiama [93]:

$$JB = \frac{n}{6} \left(s^2 + \frac{(k-3)^2}{4} \right). \quad (2.19)$$

Jarque ir Beros kriterijaus statistika yra pasiskirsčiusi pagal chi-kvadrato pasiskirstymą su dviem laisvės laipsniais. Šis kriterijus dažnai taikomas ekonomikos srityje.

- **Hoskingo (H1-H4)**

1990 m. Hoskingas ir Wallis pasiūlė kriterijų, paremtą L -momentais ir rangine statistika. L -momentai yra mažiau jautrūs imties dydžiui ir duomenų sklaidai. Remdamasis antruoju, trečiuoju ir ketvirtuoju L -momentais, kurie atitinka antrąjį,

trečiąjį ir ketvirtąjį centrinius momentus, Hoskingas įvedė naujus asimetrijos ir eksceso koeficientus. Šie nauji koeficientai apibrėžiami [94]:

$$s_3^* = \frac{\mathbb{I}_3}{\mathbb{I}_2}, \quad k_4^* = \frac{\mathbb{I}_4}{\mathbb{I}_2};$$

čia \mathbb{I}_r – ranginės statistikos L -momentas, kurį galima įvertinti:

$$\mathbb{I}_r = \sum_{i=0}^{r-1} \mathbb{P}_{r-1,i}^* \mathbb{I}_i;$$

čia

$$\mathbb{P}_{r-1,i}^* = (-1)^{r-i} \binom{r}{i} \binom{r+i}{i}, \quad \mathbb{I}_i = \frac{1}{n} \sum_{j=1}^n \frac{(j-1)(j-2) \dots (j-i)}{(n-1)(n-2) \dots (n-i)}.$$

Hoskingo (H1–H4) kriterijų statistikos apibrėžiamos [94]:

$$\begin{aligned} H1_{Lmom} &= \frac{s_3^* - \mu_{s_3^*}}{\sigma_{s_3^*}^2} + \frac{k_4^* - \mu_{k_4^*}}{\sigma_{k_4^*}^2}, \\ H2_{Lmom}^{(1)} &= \frac{s_3^{*(1)} - \mu_{s_3^*}^{(1)}}{\sigma_{s_3^*}^{2(1)}} + \frac{k_4^{*(1)} - \mu_{k_4^*}^{(1)}}{\sigma_{k_4^*}^{2(1)}}, \\ H3_{Lmom}^{(2)} &= \frac{s_3^{*(2)} - \mu_{s_3^*}^{(2)}}{\sigma_{s_3^*}^{2(2)}} + \frac{k_4^{*(2)} - \mu_{k_4^*}^{(2)}}{\sigma_{k_4^*}^{2(2)}}, \\ H4_{Lmom}^{(3)} &= \frac{s_3^{*(3)} - \mu_{s_3^*}^{(3)}}{\sigma_{s_3^*}^{2(3)}} + \frac{k_4^{*(3)} - \mu_{k_4^*}^{(3)}}{\sigma_{k_4^*}^{2(3)}}, \end{aligned} \tag{2.20}$$

čia $(\mu_{s_3^*}, \mu_{k_4^*})$ ir $(\sigma_{s_3^*}^2, \sigma_{k_4^*}^2)$ – vidurkių ir dispersijos reikšmės, kurios gaunamos modeliavimo būdu. H2–H4 kriterijų statistikoms panaudotos Elamiro ir Seheulto apibrėžtos naujos asimetrijos ir eksceso koeficientų modifikacijos [116]:

$$s_3^{*(t)} = \frac{\mathbb{I}_3^{(t)}}{\mathbb{I}_2^{(t)}}, \quad k_4^{*(t)} = \frac{\mathbb{I}_4^{(t)}}{\mathbb{I}_2^{(t)}};$$

čia t – simetriškumo pritaikymo lygis (angl. *symmetric trimming level*), o

$$\mathbb{I}_r^{(t)} = \frac{1}{r} \sum_{i=t+1}^{n-t} \left\{ \frac{\sum_{j=0}^{r-1} \left[(-1)^j \binom{r-1}{j} \binom{i-1}{r+t-1-j} \binom{n-i}{t+j} \right]}{\binom{n}{2t+r}} \right\} X_{(i)}.$$

Didėjant statistikų $H1_{Lmom}$, $H2_{Lmom}^{(1)}$, $H3_{Lmom}^{(2)}$ ir $H4_{Lmom}^{(3)}$ reikšmėms, didėja tikimybė atmesti nulinę hipotezę.

- **Kabaños ir Kabaños (CC1, CC2)**

1994 m. Kabaña ir Kabaña pasiūlė $T_{S,d}$ ir $T_{K,d}$ kriterijus [95]. 2003 m. jie pristatė keturias kriterijų grupes, skirtas normaliojo pasiskirstymo nustatymui. Jos paremtos transformuotais empiriniais procesais. Dvi kriterijų grupės yra Kolmogorovo ir Smirnovo tipo, o likusios dvi – Kramerio ir von Mises tipo. Viena iš šių grupių pagrįsta asimetrijos koeficiento pokyčiais, o kita yra jautri eksceso koeficiento pokyčiams. Pasiūlytas $T_{S,d}$ kriterijus yra apibrėžiamas [96]:

$$w_{S,d} = \Phi(x) \cdot \bar{H}_3 - \phi(x) \cdot \sum_{j=1}^d \frac{1}{\sqrt{j}} H_{j-1}(x) \cdot \bar{H}_{j+3};$$

čia d – dimensija, $\phi(x)$ – standartinio normaliojo skirstinio tikimybės tankio funkcija, $H_j(\cdot)$ žymi j -osios eilės normalizuotą Hermito polinomą, \bar{H}_j – j -osios eilės normalizuoto Hermito polinomo vidurkis, kuris yra apibrėžiamas [96]:

$$\bar{H}_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n H_j(X_{(i)}).$$

Pasiūlyti apytiksliai transformuoti empiriniai procesai, jautrūs eksceso koeficiento pokyčiams, yra apibrėžiami pagal formulę [96]:

$$w_{K,d} = -\phi(X) \cdot \bar{H}_3 + [\Phi(X) - X \cdot \phi(X)] \cdot \bar{H}_4 - \phi(X) \cdot \sum_{j=2}^d \left(\sqrt{\frac{j}{j-1}} H_{j-2}(X) \cdot H_j(X) \right) \cdot \bar{H}_{j+3}.$$

Kolmogorovo ir Smirnovo tipo kriterijaus statistika yra jautri asimetrijos ir eksceso koeficientų pokyčiams. Kriterijų $T_{S,d}$ ir $T_{K,d}$ statistikos apibrėžiamos [96]:

$$T_{S,d} = \max |w_{S,d}(X)|, \quad T_{K,d} = \max |w_{K,d}(X)|. \quad (2.21)$$

Abiem atvejais, didėjant statistikos reikšmei, didėja tikimybė atmesti nulinę hipotezę.

- **Bonetto ir Seiero (BS)**

2002 m. Bonettas ir Seieras pristatė T_w kriterijų. Šio kriterijaus statistika yra apibrėžiama [97]:

$$T_w = \frac{\sqrt{n+2} \cdot (\hat{\omega} - 3)}{3,54}; \quad (2.22)$$

čia $\hat{\omega} = 13,29 \left[\ln \left(\sqrt{n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2} \right) - \log(n^{-1} \sum_{i=1}^n |X_i - \bar{X}|) \right]$. Ši statistika turi standartinį normalųjį pasiskirstymą pagal nulinę hipotezę.

- **Bontempo ir Meddahio (BM1, BM2)**

2005 m. Bontempsas ir Meddahis pasiūlė kriterijų, paremtą momento sąlygomis, kurias nusako normaliojo skirstinio charakteristikos. Šio kriterijaus statistika apibrėžiama [98]:

$$\begin{aligned}
 BM_{3-4} &= \sum_{j=3}^4 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n H_j(Z_{(i)}) \right)^2, \\
 BM_{3-6} &= \sum_{j=3}^6 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n H_j(Z_{(i)}) \right)^2;
 \end{aligned} \tag{2.23}$$

čia $H_j(\cdot)$ žymi j -osios eilės Hermito polinomą.

- **Doorniko ir Hanseno (DH)**

2008 m. Doornikas ir Hansenas pristatė kriterijų, tinkantį vienmačiams ir daugiamačiams duomenims. Jis yra paremtas duomenų eksceso (k) ir asimetrijos (s) koeficientų transformacijomis siekiant užtikrinti duomenų nepriklausomumą. Asimetrijos koeficiento transformacija apibrėžiama [38]:

$$Z(s) = \frac{\ln\left(\frac{y}{c} + \sqrt{\left(\frac{y}{c}\right)^2 + 1}\right)}{\sqrt{\ln(\mathcal{W})}};$$

čia y , c ir \mathcal{W} yra gaunami pagal lygtis:

$$\begin{aligned}
 y &= s \sqrt{\frac{(n+1)(n+3)}{6(n-2)}}, \quad c = \sqrt{\frac{2}{\mathcal{W}^2 - 1}}, \\
 \mathcal{W}^2 &= -1 + \sqrt{2(\mathfrak{X}_2 - 1)}, \quad \mathfrak{X}_2 = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}.
 \end{aligned}$$

Bowmanas ir Šentonas pasiūlė eksceso koeficiento transformaciją, kuri yra apibrėžiama [72]:

$$\mathbf{z}_2 = \left[\sqrt[3]{\frac{\xi}{2a}} - 1 + \frac{1}{9a} \right] \sqrt{9a},$$

čia ξ ir a yra gaunami pagal lygtis:

$$\xi = 2l(k - 1 - s^2), \quad l = \frac{(n+5)(n+7)(n^3 + 37n^2 + 11n - 313)}{12(n-3)(n+1)(n^2 + 15n - 4)},$$

$$\alpha = \frac{(n+5)(n+7)[(n-2)(n^2+27n-70) + s^2(n-7)(n^2+2n-5)]}{6(n-3)(n+1)(n^2+15n-4)}.$$

Doorniko ir Hanseno kriterijaus statistika apibrėžiama [38]:

$$DH = [Z(s)]^2 + [z_2]^2 \sim \chi^2(2). \quad (2.24)$$

Šio kriterijaus statistika DH yra pasiskirsčiusi pagal chi-kvadrato pasiskirstymą su dviem laisvės laipsniais. Didėjant statistikos DH reikšmei, didėja tikimybė atmesti nulinę hipotezę.

- **Robastinis Jarquès ir Beros (GG)**

2007 m. Gelas ir Gastwirthas modifikavo Jarque ir Beros kriterijų ir pristatė galingesnę Jarquès ir Beros kriterijų. Jis galingiausias pasiskirstymams su sunkiomis uodegomis. Šio kriterijaus statistika apibrėžiama [99]:

$$RJB = \frac{n}{6} \left(\frac{m_3}{J_n^3} \right)^2 + \frac{n}{64} \left(\frac{m_4}{J_n^4} - 3 \right)^2; \quad (2.25)$$

čia J_n yra robastinis sklaidos matas (angl. *robust measure of dispersion*), kuris apibrėžiamas $J_n = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n |X_{(i)} - \mathfrak{w}|$, ir \mathfrak{w} – imties mediana, kuri apibrėžta 2.1.2 skyrelio pradžioje. Šio kriterijaus statistika turi asimptotinę chi-kvadrato pasiskirstymą su dviem laisvės laipsniais.

- **Bryso, Huberto ir Struyfo (BHS)**

2008 m. Brysas, Hubertas ir Struyfas pristatė T_{MC-LR} kriterijų. Šis kriterijus paremtas robastiniu asimetrijos koeficientu ir yra tinkamiausias pasiskirstymams su sunkiomis uodegomis. Robastinis asimetrijos matas yra laikomas (angl. *medcouple*) [24]:

$$MC = \underset{X_{(i)} \leq \mathfrak{w} \leq X_{(j)}}{\text{med}} K(X_{(i)}, X_{(j)});$$

čia med – imties \mathfrak{w} mediana, K – branduolio funkcija, kuri apibrėžiama [24]:

$$K(X_{(i)}, X_{(j)}) = \frac{(X_{(j)} - \mathfrak{w})(\mathfrak{w} - X_{(i)})}{X_{(i)} - X_{(j)}}.$$

Kairioji vidurinė jungtis (LMC) ir dešinioji vidinė jungtis (RMC) yra laikomos patikimais kairiosios ir dešinėsios uodegos svorio matais ir yra apibrėžiamos [24]:

$$LMC = -MC(X < \mathfrak{w}), \quad RMC = MC(X > \mathfrak{w}).$$

Kriterijaus T_{MC-LR} statistika yra apibrėžiama:

$$T_{MC-LR} = n(w - \omega)^T V^{-1} (w - \omega); \quad (2.26)$$

čia w yra nustatoma pagal $[MC, LMC, RMC]^T$, o ω ir V yra gaunamos pagal įverčių poveikio funkciją. Šio kriterijaus statistika turi asimptotinę chi-kvadrato

pasiskirstymą su trimis laisvės laipsniais. Normaliojo pasiskirstymo atveju ω ir V reikšmės [24]:

$$\omega = [0; 0,199; 0,199]^T, \quad V = \begin{bmatrix} 1,25 & 0,323 & -0,323 \\ 0,323 & 2,62 & -0,0123 \\ -0,323 & -0,0123 & 2,62 \end{bmatrix}.$$

- **Bryso, Huberto, Struyfo, Bonetto ir Seiero (BHSBS)**

2008 m. Brysas, Hubertas, Struyfas, Bonettas ir Seieras pristatė jungtinį T_{MC-LR} & T_W kriterijų. Taikant šį kriterijų nulinė hipotezė yra atmetama, kai bent vieno kriterijaus reikšmingumo lygmuo yra $\alpha/2$. Šio kriterijaus statistika apibrėžiama [24]:

$$T_{MC-LR} \& T_W = n(w - \omega)'V^{-1}(w - \omega) \& \frac{\sqrt{n+2} \cdot (\hat{\omega} - 3)}{3,54}.$$

- **Desgagné Lafaye de Micheaux Leblanco (DLDMRLn)**

2009 m. Desgagné Lafaye de Micheaux Leblancas pristatė R_n kriterijų. Šio kriterijaus R_n statistika apibrėžiama [25]:

$$R_n = nr_n(\bar{X}_n, \sigma)^T \left(J_0 - \frac{1}{2} \mathbf{v}_0 \mathbf{v}_0^T \right)^{-1} \cdot r_n(\bar{X}_n, S_n); \quad (2.27)$$

kai $X \sim N(\mu, \sigma^2)$,

$$r_n(\bar{X}_n, \sigma) = r_n(\mu, \sigma^2) - \frac{1}{2}(1 - T_n)\mathbf{v}_0 + \mathbf{o}_P\left(n^{-\frac{1}{2}}\right)\mathbf{1}_3,$$

$$T_n = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_{(i)} - \mu}{\sigma^2} \right)^2;$$

čia $\mathbf{1}_3 = (1 \ 1 \ 1)$ ir

$$J_0 = \begin{pmatrix} 0,42423099 & 0,13285410 & 0,05401580 \\ 0,13285410 & 0,10080860 & 0,04056528 \\ 0,05401580 & 0,04056528 & 0,01632673 \end{pmatrix},$$

ir vektorius

$$\mathbf{v}_0 = (0,86481850 \ 0,38512925 \ 0,15594892)^T,$$

$$J_0 - \frac{1}{2} \mathbf{v}_0 \mathbf{v}_0^T = \begin{pmatrix} 0,0502754623 & -0,0336793487 & -0,0134179540 \\ -0,0336793487 & 0,0266463308 & 0,0105350321 \\ -0,0134179540 & 0,0105350321 & 0,00416669944 \end{pmatrix}.$$

Šio kriterijaus statistika R_n yra asimptotiškai pasiskirsčiusi pagal chi-kvadrato pasiskirstymą su trimis laisvės laipsniais.

- **Desgagné ir Lafaye de Micheaux Leblanco (DLDMXAPD)**

2009 m. A. Desgagné ir P. Lafaye de Micheaux Leblancas pristatė X_{APD}^a kriterijų, paremtą modifikuotais asimetrijos (s_2) ir eksceso (k_2) koeficientais. X_{APD}^a

kriterijų tinka naudoti, kai imties dydis didesni už 10. Šio kriterijaus statistika yra apibrėžiama [25]:

$$X_{APD}^a = \frac{ns_2^2}{\left(3 - \frac{8}{\pi}\right)\left(1 - \frac{1,9}{n}\right)} + \frac{n \left[\sqrt[3]{k_2 - s_2^2} - \left(\sqrt[3]{\frac{2 - \log 2 - \gamma}{2}} \cdot \frac{1 - 1,026}{n} \right) \right]^2}{\frac{1}{72} \left(\frac{2 - \log 2 - \gamma}{2} \right)^{-\frac{4}{3}} (3\pi^2 - 28) \left(\frac{1 - 2,25}{n^{0,8}} \right)}; \quad (2.28)$$

čia $\gamma = 0,577215665$ yra Eulerio ir Mascheronio konstanta. Modifikuoti asimetrijos (s_2) ir eksceso (k_2) koeficientai apibrėžiami:

$$s_2 = \frac{1}{n} \sum_{i=1}^n Z_{(i)}^2 \text{sign}(Z_{(i)}), \quad k_2 = \frac{1}{n} \sum_{i=1}^n Z_{(i)}^2 \log|Z_{(i)}|.$$

Šio kriterijaus statistika X_{APD}^a yra asimptotiškai pasiskirsčiusi pagal chi-kvadrato pasiskirstymą su dviem laisvės laipsniais, kai $n \geq 10$. Kai X_{APD}^a reikšmė yra didesnė už chi-kvadrato kvantilį $\chi_{2,\alpha}^2$ su reikšmingumo lygmeniu α , tada didėja tikimybė atmesti nulinę hipotezę.

- **Desgagné ir Lafaye de Micheaux Leblanco (DLDMZEPD)**

2016 m. A. Desgagné ir P. Lafaye de Micheaux Leblancas pristatė Z_{EPD}^a kriterijų, paremtą modifikuotu eksceso (k_2) koeficientu. Šio kriterijaus statistika apibrėžiama [25]:

$$Z_{EPD}^a = \frac{n^{1/2} \left[\frac{(2k_2)^{\mathcal{G}_n} - 1}{\mathcal{G}_n} + \frac{((2 - \log 2 - \gamma)^{-0,06} - 1)}{0,06} + \frac{1,32}{n^{0,95}} \right]}{\left[\frac{(2 - \log 2 - \gamma)^{-2,12} (3\pi^2 - 28)}{2} - \frac{3,78}{n^{0,733}} \right]^{1/2}}; \quad (2.29)$$

čia $\mathcal{G}_n = -0,06 + 2,1/n^{0,67}$. Šio kriterijaus statistika Z_{EPD}^a yra asimptotiškai pasiskirsčiusi pagal $N(0,1)$, kai $n \geq 10$. Kai $|Z_{EPD}^a|$ reikšmė yra didesnė už normaliojo pasiskirstymo kvantilį $z_{\alpha/2}$ su reikšmingumo lygmeniu α , tada didėja tikimybė atmesti nulinę hipotezę.

2.1.4 Kiti vienmačiai kriterijai

Šiame poskyryje apibrėžiami vienmačiai kriterijai, paremti grupavimu, dviejų dispersijos įverčių santykiu, kai vienas iš įverčių yra stiprusis dvisvoris sklaidos įvertis \hat{S}_b^2 ir robustinis sklaidos matas J_n . Sklaidos įvertis S_b^2 apibrėžiamas [100]:

$$\hat{S}_b^2 = \frac{n \cdot \sum_{|\tilde{z}_i| < 1} (X_{(i)} - \mathfrak{w})^2 (1 - \tilde{z}_i^2)^4}{\left[\sum_{|\tilde{z}_i| < 1} (1 - \tilde{z}_i^2)(1 - 5\tilde{z}_i^2) \right]^2};$$

čia $\tilde{z}_i = (X_{(i)} - \omega)/(9A)$, $A = |X_{(i)} - \omega|$.

J_n yra apibrėžiamas [99]:

$$J_n = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n |X_{(i)} - \omega|.$$

- **Chi-kvadrato (CHI2)**

1900 m. Karlas Pearsonas sukonstravo suderinamumo hipotezės tikrinimo kriterijų, kuris iki šiol yra vienas iš populiariausių neparametrinių kriterijų. Chi-kvadrato kriterijus tinkamas dideliems imčių dydžiams. Nagrinėjami atsitiktiniai dydžiai turi būti nepriklausomi. Šis kriterijus parodo, ar empirinio ir teorinio skirstinių skirtumas yra reikšmingas, t. y. tikrinama, ar turimas empirinis skirstinys yra suderintas su teoriniu modeliu. Chi-kvadrato kriterijaus statistika apibrėžiama [1]:

$$\chi^2 = \sum_{i=1}^t \frac{(O_{(i)} - E_{(i)})^2}{E_{(i)}}; \quad (2.30)$$

čia $O_{(i)}$ – stebėjimų i -tasis dažnis, $E_{(i)}$ – tikėtinas (teorinis) i -tasis dažnis, t – grupių skaičius.

- **Martinezas ir Iglewiczas (MI)**

1981 m. Martinezas ir Iglewiczas pasiūlė suderinamumo hipotezės tikrinimo kriterijų, paremtą dviejų dispersijos įverčių santykiu. Martinezo ir Iglewiczas pasiūlyto kriterijaus statistika I_n apskaičiuojama [100]:

$$I_n = \frac{\sum_{i=1}^n (X_{(i)} - \omega)^2}{(n-1) \cdot \hat{S}_b^2}. \quad (2.31)$$

Didėjant statistikos I_n reikšmei, didėja tikimybė atmesti nulinę hipotezę. Martinezas ir Iglewiczas parodė, kad šis kriterijus yra galingas sunkių uodegų simetriniams duomenų rinkiniams.

- **Eppso ir Pulley (EP)**

1983 m. Epps ir Pulley pasiūlė kriterijaus statistiką T_{EP} , paremtą svertiniu integralu [101]:

$$T_{EP} = \int_{-\infty}^{\infty} |\varphi_n(t) - \hat{\varphi}_0(t)|^2 dG(t);$$

čia $\varphi_n(t) = n^{-1} \sum_{j=1}^n \exp(itx_j)$ yra empirinė charakteristinė funkcija, $\hat{\varphi}_0(t) = \exp(it\bar{x} - 0.5m_2t^2)$ – normaliojo skirstinio charakteristinės funkcijos imties įvertis ir $G(t)$ – tam tikra funkcija, pasirinkta atsižvelgiant į keletą savybių. Nustatant $dG(t) = g(t)dt$ ir parenkant $g(t) = \sqrt{m_2/2\pi} \cdot \exp(-0.5m_2t^2)$, tada T_{EP} statistiką galima apibrėžti [101]:

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} \exp\left(\frac{-(X_j - X_i)^2}{2m_2}\right) - \sqrt{2} \sum_{j=1}^n \exp\left(\frac{-(X_j - \bar{X})^2}{4m_2}\right). \quad (2.32)$$

Didėjant statistikos T_{EP} reikšmei, didėja tikimybė atmesti nulinę hipotezę.

- **Gelo, Miao ir Gastwirtho (GMG)**

2007 m. Gelas, Miao ir Gastwirthas pasiūlė R_{SJ} kriterijų, kuris paremtas santykiu tarp standartinio nuokrypio ir J_n . R_{SJ} kriterijaus statistika apibrėžiama pagal formulę [99]:

$$R_{SJ} = \frac{\hat{\sigma}}{J_n}. \quad (2.33)$$

Didėjant statistikos R_{SJ} reikšmei, didėja tikimybė atmesti nulinę hipotezę. Statistika $\sqrt{n}(R_{SJ} - 1)$ asimptotiškai pasiskirsto pagal normalųjį skirstinį.

- **Bakšajevo (Baksh)**

2009 m. A. Bakšajevas pristatė suderinamumo kriterijų, paremtą N -metrika. Laikoma, kad $(\mathfrak{X}, \mathfrak{U})$ yra mačioji erdvė, \mathfrak{B} yra tikimybinių matų ϱ , kurie yra apibrėžti mačiojoje erdvėje, aibė. Tegul K yra reali ir tolydi simetrinė funkcija, \mathfrak{B}_K yra aibė tikimybinių matų ϱ iš \mathfrak{B} , kurie tenkina sąlygą:

$$\int_{\mathfrak{X}} \int_{\mathfrak{Y}} K(x, y) d\varrho(x) d\varrho(y) < \infty.$$

Svarbiausia N -metrikų teorijos sąvoka yra stipriai neigiamai apibrėžtas branduolys. Jeigu $\forall n \in \mathbb{N}, \forall x_1, \dots, x_n \in \mathfrak{X}$ ir $\forall c_1, \dots, c_n \in \mathbb{C}$ su sąlyga, kad $\sum_{i=1}^n c_i = 0$, funkcija $K(x, y)$ tenkina nelygybę

$$\sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) c_i \bar{c}_j \leq 0,$$

tada K yra neigiamai apibrėžtas branduolys.

Klebanovas parodė [7], kad jeigu $K(x, y) = K(y, x)$ ir $K(x, x) = 0, \forall x, y \in \mathfrak{X}$, tai nelygybė

$$\begin{cases} N(\varrho, \nu) > 0, & \varrho \neq \nu, \\ N(\varrho, \nu) = 0, & \varrho = \nu; \end{cases}$$

čia $N(\varrho, \nu) = 2 \int_{\mathfrak{X}} \int_{\mathfrak{X}} K(x, y) d\varrho(x) d\nu(y) - \int_{\mathfrak{X}} \int_{\mathfrak{X}} K(x, y) d\varrho(x) d\varrho(y) - \int_{\mathfrak{X}} \int_{\mathfrak{X}} K(x, y) d\nu(x) d\nu(y), \varrho, \nu \in \mathfrak{B}_K$. Ši nelygybė yra teisinga visiems matams $\varrho, \nu \in \mathfrak{B}_K$ tada ir tik tada, kai $K(x, y)$ yra stipriai neigiamai apibrėžtas branduolys. Šis teiginys leidžia konstruoti suderinamumo kriterijus.

Pasiūlytas kriterijus apibrėžiamas tokiu būdu: esant nulinei hipotezei,

$$T_n = -n \int_0^1 \int_0^1 K(x) d(F_n^*(x) - x), \quad (2.34)$$

skirstinys bus kvadratinės formos [5, 6]:

$$T_n = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{a_{kj}}{\pi^2 kj} \xi_k \xi_j;$$

čia ξ_k yra standartinio normaliojo skirstinio nepriklausomieji atsitiktiniai dydžiai ir

$$a_{kj} = -2 \int_0^1 \int_0^1 K(x) d\sin(\pi kx).$$

A. Bakšajevs pritaikė branduolio funkciją [5, 6]:

$$K(x, y) = |x - y|^r, \quad (0 < r < 2). \quad (2.35)$$

2.1.5 Naujas vienmatis kriterijus (N1)

Sukonstruotas naujas suderinamumo hipotezės tikrinimo kriterijus, paremtas N -metrika, su pritaikyta nauja branduolio funkcija. Jis yra realizuotas naudojant SAS programinę įrangą. Šiame kriterijuje A. Bakšajevs pristatytai kriterijaus statistikai (2.34) buvo pritaikyta (2.36) branduolio funkcija. Branduolio sąvoka statistikoje turi dvi prasmes. Jis taikomas Bajeso statistikoje ir atliekant neparametrinių tankio įvertinių skaičiavimus. Bajeso statistikoje tikimybės tankio funkcijos branduolys yra tam tikra išraiška, kurioje yra tik kintamieji ir veiksniai, priklausantys nuo branduolio funkcijos. Tai, kas nepriklauso nuo funkcijos kintamųjų, yra neįtraukiama. Neparametrinėje statistikoje branduolys yra svorio funkcija. Ji dažnai taikoma neparametriniuose duomenų analizės metoduose. Branduoliai dažniausiai naudojami norint nustatyti atsitiktinių dydžių pasiskirstymą. Branduolys gali įgyti neneigiamąsias realiąsias reikšmes, jam galioja kelios savybės [5, 6]:

- $\int_{-\infty}^{+\infty} K(x) dx = 1$;
- $K(-x) = K(x), \forall x \in \mathbb{R}$.

Dažniausiai taikomos branduolio funkcijos [5, 6, **P3**, **K2**, **K3**]:

- Standartinio normaliojo:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-0,5 \cdot x^2}.$$

- Jepanečnikovo:

$$K(x) = \frac{3}{4} (1 - x^2), \text{ kai } x \in [-1; 1].$$

- Dvisvorė:

$$K(x) = \frac{15}{16}(1 - x^2)^2, \text{ kai } x \in [-1; 1].$$

- Trisvorė:

$$K(x) = \frac{35}{32}(1 - x^2)^3, \text{ kai } x \in [-1; 1].$$

- Trikubė:

$$K(x) = \frac{70}{81}(1 - |x|^3)^3, \text{ kai } x \in [-1; 1].$$

Disertaciniame tyrime pasiūlytas A. Bakšajevo suderinamumo hipotezės tikrinimo kriterijaus, paremto N -metrika, patobulinimas. A. Bakšajevo pristatytai statistikai (2.34) buvo pasiūlyta pritaikyti naują branduolio funkciją (žr. 2.1 pav.), kuri apibrėžiama:

$$K(x) = \varphi(\bar{g}(x))\bar{g}'(x); \quad (2.36)$$

čia $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

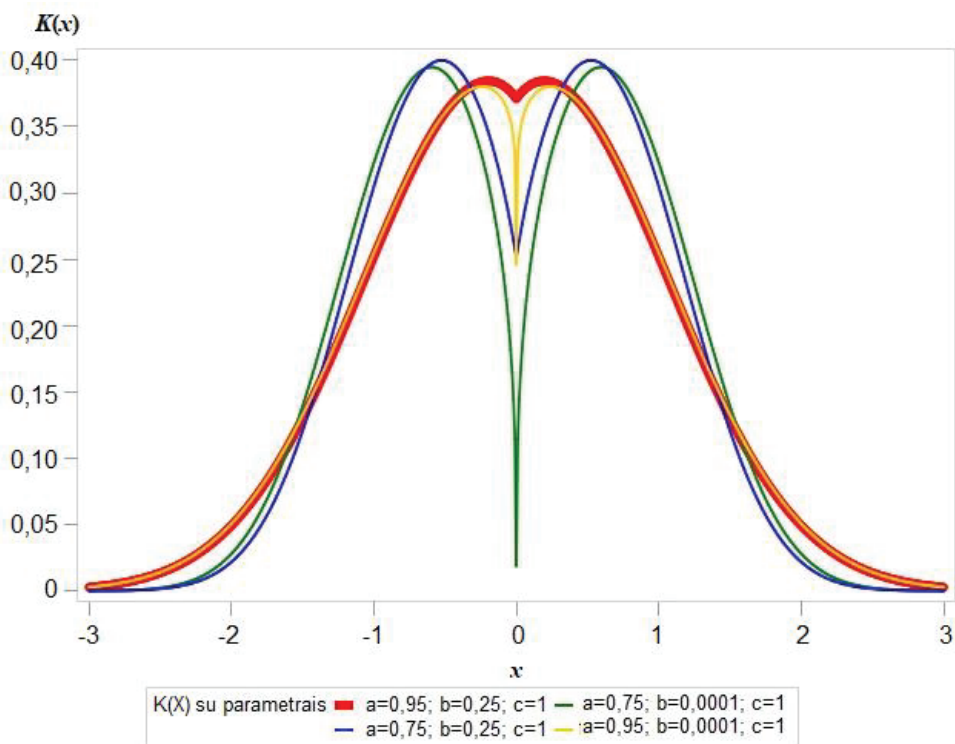
Pasiūlyta branduolio funkcija apskaičiuojama pagal imties reikšmes, t. y. $X = X(t)$. Šios branduolio funkcijos pagrindinė savybė: jos forma parinkta taip, kad sumažinamas poslinkis, atsirandantis kriterijaus sudarymo metu panaudojant imties reikšmes. Branduolio konstrukcija parinkta taip, kad naudojamo imties taško įtaka konstruojamam įvertiniui būtų mažesnė už to taško aplinką.

Tegul Ξ ir Ψ yra pasiskirstymo ir tankio funkcijos, $g: \mathbf{R} \rightarrow \mathbf{R}$ yra nelyginė monotoniškai didėjanti funkcija. Tada atsitiktinio kintamojo $Y = g(x)$ pasiskirstymo funkcija $F_y = \Xi(\bar{g}(x))$; čia \bar{g} yra funkcijos g atvirkštinė. Atsitiktinio kintamojo y pasiskirstymo tankis $f_y = \Psi(\bar{g}(x))\bar{g}'(x)$. Pasirinktos parametrinės funkcijos \bar{g} klasė priklauso nuo trijų parametrų [**P1**, **K1**]:

$$\bar{g} = x(c + |x|^b)^a, \quad a, b, c > 0,$$

$$\bar{g}' = (c + |x|^b)^a + a|x|(c + |x|^b)^{a-1}b|x|^{b-1};$$

čia a – sklaidos, b ir c – įdubos ir viršūnės formos parametrai.



2.1 pav. Pasiūlytos branduolio funkcijos forma su keletu nagrinėtų parametru rinkinių

2.2. Daugiamačiai suderinamumo hipotezės tikrinimo kriterijai

Laikoma, kad d -matis kintamasis yra pasiskirstęs pagal normalųjį skirstinį $N_d(\mu, \Sigma)$; čia μ yra tikėtinasis vektorius $\mu = (\mu_1, \dots, \mu_d)^T$. N_d žymi visų galimų d -mačių kintamųjų normaliųjų skirstinių rinkinį.

Tegul X_1, X_2, \dots, X_n ; čia $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$, $i = 1, 2, \dots, n$ yra baigtinis d -matis vektorius (stulpelis) X su pasiskirstymo funkcija F_X . Vektoriaus vidurkis \bar{X} yra gaunamas:

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j.$$

Inties kovariacinė matrica apibrėžiama:

$$\Sigma = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})^T.$$

Daugiamačiam X normalumui (rementis stebėta X_1, X_2, \dots, X_n imtimi) įvertinti yra sukurta daug statistinių kriterijų. Jiems svarbiausios dvi savybės: invariantiškumo ir suderinamumo.

N_d yra uždara panašių transformacijų atžvilgiu, t. y.

$$F_{AX+b} \in N_d \Leftrightarrow F_X \in N_d;$$

visiems $b \in \mathbb{R}^d$ ir visoms nesinguliarosioms matricoms $A \in \mathbb{R}^{d \times d}$. Taigi daugiamačio kriterijaus statistikos T_n reikšmė turėtų būti vienoda pradinei ir tiesiškai transformuotai imtims

$$T_n(AX_1 + b, \dots, AX_n + b) = T_n(X_1, \dots, X_n); \quad (2.37)$$

visiems $b \in \mathbb{R}^d$ ir visoms nesinguliarosioms matricoms $A \in \mathbb{R}^{d \times d}$.

Invariantiškojo kriterijaus statistika turi tenkinti (2.37) sąlygą. Kriterijai taikomi standartizuotosioms imtims $Y_j = \Sigma^{-\frac{1}{2}}(X_j - \bar{X})$, kai $j = 1, \dots, n$ tenkina invariantiškumo savybę. Tačiau ne visada. Praktikoje dažnai nėra žinomas imties X_1, X_2, \dots, X_n pasiskirstymas. Tokiu atveju svarbu taikyti suderintąjį (angl. *consistent*) kriterijų, kurio nulinės hipotezės teisingo atmetimo tikimybė artėja prie 1, didėjant imties dydžiui [39, 40]. Išsamiau daugiamačių kriterijų invariantiškumo ir suderinamumo savybės aptariamos [39, 40].

2.2.1. Daugiamačiai kriterijai, paremti regresija ir koreliacija (rangine statistika)

Šiame poskyryje pateikti kriterijai, paremti kvadratine forma (angl. *squared radii*). Ši forma apibrėžiama:

$$D_{n,j} = (X_j - \bar{X})^T \Sigma^{-1} (X_j - \bar{X}); \quad j = 1, 2, \dots, n.$$

$D_{n,j}$ yra asimptotiškai pasiskirsčiusi pagal chi-kvadrato pasiskirstymą su k laisvės laipsniais.

- **Chi-kvadrato (CHI2)**

1981 m. Moore ir Stubblebine pristatė daugiamatį chi-kvadrato suderinamumo hipotezės tikrinimo kriterijų, paremtą rangine statistika. Šio kriterijaus statistika apibrėžiama [37]:

$$M_{n,k} = \frac{k}{n} \sum_{l=1}^k \left(N_{n,l} - \frac{n}{k} \right)^2; \quad (2.38)$$

čia $N_{n,l} = \sum_{j=1}^n 1\{a_{l-1} < D_{n,j} \leq a_l\}$, ($l = 1, 2, \dots, k$; $a_0 = 0, a_k = +\infty$). $M_{n,k}$ gali įgyti ekvivalenčią formą [37]:

$$M_{n,k} = k \sum_{l=1}^k (\mathcal{G}_n(a_l) - \mathcal{G}_n(a_{l-1}))^2;$$

čia $\mathcal{G}(\cdot)$ yra $\chi^2(d)$ tikimybinė pasiskirstymo funkcija, $\mathcal{G}_d(a_l) - \mathcal{G}_d(a_{l-1}) = k^{-1}$ ($l = 1, 2, \dots, k$; $\mathcal{G}_d(+\infty) = 1$).

- **Kramerio ir von Mises (CVM)**

1982 m. Koziolis pasiūlė naudoti Kramerio ir von Mises tipo daugiamačių suderinamumo hipotezės kriterijų, paremtą rangine statistika. Šio kriterijaus statistika apibrėžiama [40]:

$$CM = \frac{1}{12n} + \sum_{j=1}^n \left(\mathcal{G}_d(D_{(n,j)}) - \frac{2j-1}{n} \right)^2. \quad (2.39)$$

- **Andersono ir Darlingo (AD)**

1987 m. Paulson, Roohan ir Sullo pasiūlė Andersono ir Darlingo tipo daugiamačių suderinamumo hipotezės kriterijų, paremtą ranginę statistika. Šio kriterijaus statistika apibrėžiama [40]:

$$AD = -n - \sum_{j=1}^n \frac{2j-1}{n} \left(\log \mathcal{G}_d(D_{(n,j)}) + \log \left(1 - \mathcal{G}_d(D_{(n+1-j)}) \right) \right). \quad (2.40)$$

- **Nikulino, Rao ir Robsono (NRR) bei Dzhaparidzės ir Nikulino (DN)**

1981 m. Moore ir Stubblebine pasiūlė daugiamačio suderinamumo hipotezės tikrinimo kriterijaus Nikulino, Rao ir Robsono statistiką, kuri bet kokios dimensijos kovariacijos matricai yra apibrėžiama [37]:

$$Y_n^2 = \sum V_i^2 + \frac{4dr(\sum V_i d_i)^2}{1 - 4dr \sum d_i^2}; \quad (2.41)$$

čia V_i – standartizuotų reikšmių dažnių vektorius su komponentais

$$V_i = V_{in}(\hat{\theta}_n) = \frac{(N_{in} - n/r)}{\sqrt{n/r}}, \quad i = 1, \dots, r;$$

čia $\theta = (\mu_1, \dots, \mu_d, \sigma_{11}, \sigma_{12}, \sigma_{22}, \dots, \sigma_{1j}, \sigma_{2j}, \dots, \sigma_{jj}, \dots, \sigma_{dd})$ yra vektorius su nežinomais parametrais, o $\hat{\theta}_n$ yra parametro θ maksimalaus tikėtinumo įvertis, N_{in} – stebėtų atsitiktinių vektorių X_1, \dots, X_n , patenkančių į $E_{in}(\hat{\theta}_n) = \{X \in \mathbb{R}^d: c_{i-1} \leq D_{n,j} < c_i = i/r\}, i = 1, \dots, r$, dažnis. Tada ribinė standartizuotų dažnių kovariacijos matrica $V_n(\hat{\theta}_n) = \tau_l = I - qq^T - BJ^{-1}B^T$, kai B yra $r \times \sigma$ dydžio matrica

$$B_{ij} = \frac{1}{\sqrt{p_i(\theta)}} \frac{\partial p_i(\theta)}{\partial \theta_j}, i = 1, \dots, r, j = 1, \dots, \sigma;$$

čia q – r -axis vektorius su reikšmėmis $1/\sqrt{r}$, $J = J(\theta)$ yra Fišerio informacijos matrica vienam stebiniui, kuri apibrėžiama [37]:

$$J(\theta) = \begin{bmatrix} \tau^{-1} & 0 \\ 0 & \Sigma^{-1} \end{bmatrix};$$

čia Q yra $d(d+1)/2 \times d(d+1)/2$ dydžio w kovariacinė matrica (\sqrt{nS} vektorius, išdėstytas stulpeliais, paimant viršutinius trikampius elementus) [37]:

$$w = (s_{11}, s_{12}, s_{22}, s_{13}, s_{23}, s_{33}, \dots, s_{pp})^T.$$

Antrasis Y_n^2 narys (2.41) atkuria informaciją, kuri buvo prarasta dėl duomenų grupavimo. Kita naudinga Y_n^2 išraiška yra

$$Y_n^2 = U_n^2 + S_n^2;$$

čia U_n^2 – 1974 m. Dzhaparidzės ir Nikulino statistika, kuri yra apibrėžiama [37]:

$$U_n^2 = V_n^T(\hat{\theta}_n)[I - B_n(B_n^T B_n)^{-1} B_n^T] V_n(\hat{\theta}_n); \quad (2.42)$$

ir 1985 m. pasiūlyta McCulloch statistika [37, 66]:

$$S_n^2 = Y_n^2 - U_n^2 = V_n^T(\hat{\theta}_n) B_n [(J_n - B_n^T B_n)^{-1} + (B_n^T B_n)^{-1}] B_n^T V_n(\hat{\theta}_n);$$

čia B_n ir J_n yra parametru B ir J maksimalaus tikėtimumo įverčiai.

Jeigu $B = s$, tada statistikos U_n^2 ir S_n^2 yra asimptotiškai nepriklausomai pasiskirsčiusios pagal χ_{r-s-1}^2 ir χ_s^2 .

2.2.2. Momentais paremti daugiamačiai kriterijai

Šiame poskyryje apžvelgiami daugiamačiai kriterijai, paremti eksceso ir asimetrijos koeficientais. Šie koeficientai yra apibrėžti 2.1.3 poskyryje.

- **Mardia (Mar)**

1970 m. K. V. Mardia pristatė naują daugiamatį suderinamumo hipotezės tikrinimo kriterijų, parentą eksceso ir asimetrijos koeficientais. Šio kriterijaus statistika pagal eksceso ir asimetrijos koeficientus apibrėžiama [75]:

$$M_S(s) = \frac{n \cdot s \cdot d}{6} \rightarrow \chi^2 \left(\frac{d(d+1)(d+2)}{6} \right), \quad (2.43)$$

$$M_k(k) = \frac{n(k - d(d+2))^2 \cdot d}{8d(d+2)} \rightarrow \chi^2(1). \quad (2.44)$$

- **Roystono (Roy)**

1982 m. Roystonas pristatė kriterijų, kuris naudoja Šapiro ir Vilko arba Šapiro ir Francijos statistiką tikrinant normalumo prielaidą daugiamačiu atveju. Jei duomenų eksceso koeficientas yra didesnis nei 3, tada taikoma Šapiro ir Francijos statistika, o priešingu atveju taikoma Šapiro ir Vilko statistika [42, 45].

Tegul \mathcal{W}_j yra Šapiro ir Vilko arba Šapiro ir Francijos statistika j -ajam kintamajam ($j = 1, 2, \dots, d$), o Z_j reikšmės yra gautos taikant Roystono pasiūlytas normalumo transformacijas [42, 45]:

$$\begin{aligned} \text{jei } 4 \leq n \leq 11, & \quad y = n & \quad \text{ir} & \quad w_j = -\log[\gamma - \log(1 - \mathcal{W}_j)], \\ \text{jei } 12 \leq n \leq 2000, & \quad y = \log(n) & \quad \text{ir} & \quad w_j = \log(1 - \mathcal{W}_j). \end{aligned}$$

Iš šių transformacijų matyti, kad y ir w_j keičiasi atsižvelgiant į imties dydį n . Kiekvieno atsitiktinio kintamojo transformuotos reikšmės yra gaunamos [42, 45]:

$$Z_j = \frac{w_j - \ell}{\sigma};$$

čia Y , ℓ ir σ yra gaunami iš Roystono pasiūlytų polinomų aproksimacijos skirtingiems imties dydžiams [45, 42]:

$$Y = a_{0Y} + a_{1Y}Y + a_{2Y}Y^2 + \dots + a_{dY}Y^d,$$

$$\ell = a_{0\ell} + a_{1\ell}Y + a_{2\ell}Y^2 + \dots + a_{d\ell}Y^d,$$

$$\log(\sigma) = a_{0\sigma} + a_{1\sigma}Y + a_{2\sigma}Y^2 + \dots + a_{d\sigma}Y^d.$$

Tada Roystono daugiamatnio kriterijaus statistiką galima apibrėžti [45]:

$$H = \frac{e \sum_{j=1}^d \psi_j}{d} \sim \chi_e^2; \quad (2.45)$$

čia e – laisvės laipsnių skaičius, $\Phi(\cdot)$ – standartinio normaliojo pasiskirstymo funkcija, kai

$$e = \frac{d}{[1 + (d-1)\bar{c}]},$$

$$\psi_j = \left\{ \Phi^{-1} \left[\frac{\Phi(-Z_j)}{2} \right] \right\}^2, \quad j = 1, 2, \dots, d.$$

Tegul R yra koreliacijos matrica, o r_{ij} – koreliacija tarp i -ojo ir j -ojo kintamųjų. Tada papildomą narį \bar{c} galima apskaičiuoti:

$$\bar{c} = \sum_i \sum_j \frac{c_{ij}}{d(d-1)}, \quad \{c_{ij}\}_{i \neq j},$$

čia

$$c_{ij} = \begin{cases} g(r_{ij}, n) & \text{kai } i \neq j, \\ 1 & \text{kai } i = j. \end{cases}$$

Kai $g(0, n) = 0$ ir $g(1, n) = 1$, tada $g(\cdot)$ galima apibrėžti:

$$g(r, n) = r^e \left[1 - \frac{\ell}{v} (1-r)^\ell \right].$$

1980 m. Rossas nežinomus parametrus ℓ , ρ ir v įvertino modeliavimo būdu [102]. Jis nustatė, kad $\mu = 0,715$ ir $\rho = 5$ imties dydžiui $10 \leq n \leq 2000$, o v yra kubinė funkcija, kurią galima gauti:

$$v(n) = 0,21364 + 0,015124(\log(n))^2 - 0,0018034(\log(n))^3.$$

2.2.3. Kiti daugiamačiai kriterijai

Šiame poskyryje apžvelgiami kriterijai, paremti Mahalanobio atstumu. $\mathcal{D}_i = (X_i - \bar{X})^T \Sigma^{-1} (X_i - \bar{X}) = m_{ii}$ pažymi kvadratinį Mahalanobio i -ojo stebinio atstumą iki centroido, o $\mathcal{D}_{ij} = (X_i - X_j)^T \Sigma^{-1} (X_i - X_j)$ pateikia Mahalanobio atstumą tarp i -ojo ir j -ojo stebinių.

- **Henzės ir Zirklerio (HZ)**

1990 m. Henze ir Zirkleris pristatė kriterijų, paremtą neneigiamuoju funkciniu atstumu, kuris matuoja atstumą tarp dviejų paskirstymo funkcijų. Jei duomenys yra pasiskirstę pagal daugiamatį normalųjį skirstinį, tada kriterijaus statistika turi log-normalųjį pasiskirstymą. Pirmiausia yra apskaičiuojamas vidurkis, dispersija ir glodinimo parametras (h). Pagal imtis normalizuojama panaudojant vidurkį ir dispersiją ir apskaičiuojama p -reikšmė. Henzės ir Zirklerio daugiamačio kriterijaus statistika apibrėžiama [39]:

$$HZ = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{h^2}{2} \mathcal{D}_{ij}} - 2(1 + h^2)^{-\frac{d}{2}} \sum_{i=1}^n e^{-\frac{h^2}{2(1+h^2)} \mathcal{D}_i} + n(1 + 2h^2)^{-\frac{d}{2}}; \quad (2.46)$$

čia $h = \frac{1}{\sqrt{2}} \left(\frac{n(2d+1)}{4} \right)^{\frac{1}{d+4}}$.

Jei duomenys turi daugiamatį normalųjį pasiskirstymą, tai šio kriterijaus statistika turi lognormalųjį pasiskirstymą su vidurkiu:

$$1 - \frac{a^{-\frac{d}{2}} \left(1 + dh^{\frac{2}{d}} + (d(d+2)h^4) \right)}{2a^2},$$

ir dispersija [39]:

$$2(1 + 4h^2)^{-\frac{d}{2}} + \frac{2a^{-d}(1 + 2dh^4)}{a^2} + \frac{3d(d+2)h^8}{4a^4} - 4w_h^{-\frac{d}{2}} \left(1 + \frac{3dh^4}{2w_h} + \frac{d(d+2)h^8}{2w_h^2} \right);$$

čia $a = 1 + 2h^2$ ir $w_h = (1 + h^2)(1 + 3h^2)$.

- **Lobato ir Velasco (LV)**

2004 m. I. Lobato ir C. Velasco patobulino Jarquès ir Beros kriterijų ir pritaikė jį stacionariems procesams. Šio kriterijaus statistika apibrėžiama [103]:

$$\mathcal{G} = \frac{n\hat{\mu}_3^2}{6\hat{F}^{(3)}} + \frac{n(\hat{\mu}_4 - 3\hat{\mu}_2)^2}{24\hat{F}^{(4)}}; \quad (2.47)$$

čia $\hat{F}^{(k)} = \sum_{t=-1}^{n-1} \hat{\psi}(t) [\hat{\psi}(t) + \hat{\psi}(n - |t|)]^{k-1}$ yra auto-kovariacinė funkcija.

- **Energijos (En)**

2013 m. G. Szekely ir M. Rizzo pristatė naują daugiamatį suderinamumo hipotezės kriterijų, paremtą “energijos“ atstumu tarp daugiamatčių skirstinių. Šio kriterijaus statistika apibrėžiama [104]:

$$\mathcal{E}_n = n \left(\frac{2}{n} \sum_{j=1}^n \mathbb{E} \|\tilde{Y}_{n,j} - N_1\| - \mathbb{E} \|N_1 - N_2\| - \frac{1}{n^2} \sum_{j,k=1}^n \|\tilde{Y}_{n,j} - \tilde{Y}_{n,k}\| \right); \quad (2.48)$$

čia $\tilde{Y}_{n,j} = \sqrt{n/(n-1)} Y_{n,j}$, $Y_{n,j} = \sum_{i=1}^d (X_j - \bar{X}_n)_i$, $j = 1, \dots, n$ yra vadinama sklaidos liekanomis. N_1 ir N_2 yra nepriklausomieji pagal normalųjį skirstinį atsitiktinai pasiskirstę vektoriai. $\mathbb{E} \|N_1 - N_2\| = 2\Gamma\left(\frac{d+1}{2}\right) / \Gamma\left(\frac{d}{2}\right)$; čia $\Gamma(\cdot)$ yra gama funkcija.

Nulinė hipotezė yra atmetama, kai \mathcal{E}_n įgyja dideles reikšmes.

2.2.4. Naujas daugiamatis kriterijus (N2)

Sukonstruotas naujas pasiskirstymo tankių skirtumo vertinimu ir apvertimo formulės taikymu paremtas daugiamatis kriterijus. Jis realizuotas naudojant SAS programinę įrangą. Tankio įvertinimo metodas, paremtas charakteristinės funkcijos ir apvertimo formulės taikymu, pasižymi geromis savybėmis (mažesnis jautrumas nuo žvalgomojo tyrimo taškų). Šio kriterijaus pradinės idėjos buvo pristatytos [9]. Tankio įvertinio palyginimas su adaptuotu branduoliniu ir kt. buvo atliktas naudojant Gauso skirstinių mišinius, pasiūlytus Marrono ir Wano [105]. Šis palyginimas parodė, jog toks tankio įvertinys, paremtas charakteristinės funkcijos ir apvertimo formulės taikymu, yra geresnis už kitus lygintus, kai naudojami žymiai nuo Gauso skirstinio besiskiriantys duomenys [**P2**, **K4**, **K5**].

Laikoma, kad atsitiktinis vektorius $X \in \mathbb{R}^d$ tenkina skirstinių mišinio modelį, jeigu jo pasiskirstymo tankis $f(X)$ tenkina lygybę

$$f(X) = f(X, \theta_i) = \sum_{i=1}^q p_i f_i(X). \quad (2.49)$$

Parametras q vadinamas mišinio klasterių (komponentų, klasių) skaičiumi, o p_i – apriorinėmis tikimybėmis. Jos tenkina sąlygas:

$$p_i > 0, \sum_{i=1}^q p_i = 1. \quad (2.50)$$

(2.49) formulėje funkcija $f_i(X)$ yra pasiskirstymo tankio funkcija, o θ – daugiamatis modelio parametras. Tariaama, kad X yra d -matis atsitiktinis vektorius, kurio pasiskirstymo tankis $f(X)$, ir gaunama iš nepriklausomų X dydžių sudaryta imtis $X = (X_1, \dots, X_n)$. Laikoma, kad imtis tenkina mišinio modelį, jeigu $X(t)$ tenkina (2.49).

Apvertimo formulės tankio įvertinys. Nagrinėjant parametrinių metodų aproksimacijas, reikia pabrėžti, jog, didėjant duomenų dimensijai, modelio parametrų skaičius sparčiai auga, todėl sunkiau rasti tikslius parametrų įverčius. Vienamatių duomenų projekcijų

$$x_\tau = \tau^T x \quad (2.51)$$

tankį f_τ rasti daug lengviau negu daugiamačių duomenų tankį f . Kadangi egzistuoja abipus vienareikšmė atitiktis

$$f \leftrightarrow \{f_\tau, \tau \in \mathbb{R}^d\}, \quad (2.52)$$

tai visai natūralu bandyti rasti daugiamačią tankį f naudojant vienamačių stebinių projekcijų tankių įverčius \hat{f}_τ [9].

Pažymėtina, kad mišinio (2.49) atveju, kai skirstiniai yra Gauso, stebinių projekcijos (2.51) taip pat pasiskirsčiusios pagal (vienamačią) Gauso skirstinių mišinio modelį:

$$f_\tau(x) = f_\tau(x, \theta_\tau) = \sum_{i=1}^q p_{i,\tau} \varphi_{i,\tau}(x); \quad (2.53)$$

čia $\varphi_{i,\tau}(x) = \varphi(x; \kappa_{i,\tau}, \mathcal{E}_{i,\tau}^2)$ – vienamatis Gauso tankis. Daugiamačio mišinio parametrai $\theta_\tau = (p_{i,\tau}, \kappa_{i,\tau}, \mathcal{E}_{i,\tau}^2)$, $i = 1, \dots, q$ sieja lygybės:

$$\begin{aligned} p_{j,\tau} &= p_j, \\ \kappa_{j,\tau} &= \tau^T M_j, \\ \mathcal{E}_{j,\tau}^2 &= \tau^T \mathcal{R}_j \tau. \end{aligned} \quad (2.54)$$

Taikoma *apvertimo formulė*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-it^T x} \psi(t) dt; \quad (2.55)$$

čia

$$\psi(t) = E e^{it^T x} \quad (2.56)$$

žymi atsitiktinio dydžio x charakteristinę funkciją. Pažymėjus $u = |t|$, $\tau = t/|t|$ ir pakeitus kintamuosius į sferinę koordinačių sistemą, gaunama:

$$f(x) = \frac{1}{(2\pi)^d} \int_{\tau: |\tau|=1} ds \int_0^\infty e^{-i u \tau^T x} \psi(u\tau) u^{d-1} du; \quad (2.57)$$

čia pirmasis integralas suprantamas kaip vienetinės sferos paviršinis integralas. Pažymėjus stebimo atsitiktinio dydžio projekcijos charakteristinę funkciją

$$\psi_\tau(u) = E e^{i u \tau^T x}, \quad (2.58)$$

galioja lygybė

$$\psi(u\tau) = \psi_\tau(u). \quad (2.59)$$

Projektavimo kryptys buvo parenkamos taip: tolygiai generuota hiperkubė, kurio kraštinė lygi 2. Tikrinama, ar sugeneruotas taškas, skirtas krypties nurodymui, patenka į vienetinio spindulio hipersferą. Jeigu taškas buvo už sferos, tada buvo atliekamas pergeneravimas. Jeigu taškas buvo sferos viduje, tada juo buvo apibrėžta kryptis. Pasirinkus ($\tau = 10\,000$) projektavimo krypčių, tolygiai išsidėsčiusių ant

sferos, aibę T ir charakteristinę funkciją keičiant jos įvertiniu, gaunama formulė įverčiui apskaičiuoti [8, 9]:

$$\hat{f}(x) = \frac{A(d)}{\#T} \sum_{\tau \in T} \int_0^\infty e^{-iu\tau^T x} \hat{\psi}_\tau(u) u^{d-1} e^{-hu^2} du; \quad (2.60)$$

čia ir toliau $\#T$ žymi aibės T elementų skaičių. Pasinaudojus d -mačio rutulio tūrio formule

$$V_d(\mathcal{R}) = \frac{\frac{d}{\pi^2} \mathcal{R}^d}{\Gamma(\frac{d}{2}+1)} = \begin{cases} \frac{\frac{d}{\pi^2} \mathcal{R}^d}{(\frac{d}{2})!}, & \text{kai } d \bmod 2 \equiv 0, \\ \frac{2^{\frac{d+1}{2}} \pi^{\frac{d-1}{2}} \mathcal{R}^d}{d!!}, & \text{kai } d \bmod 2 \equiv 1, \end{cases} \quad (2.61)$$

galima apskaičiuoti konstantą $A(d)$, priklausančią nuo duomenų dimensijos:

$$A(d) = \frac{(V_d(1))'_\mathcal{R}}{(2\pi)^d} = \frac{d2^{-d}\pi^{-\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}. \quad (2.62)$$

Kompiuterinio modeliavimo tyrimai parodė, jog taikant apvertimo formulę gauti tankio įverčiai yra neglodūs. Todėl (2.60) formulėje po integralo ženklų naudojamas papildomas daugiklis e^{-hu^2} . Šis daugiklis įvertį $\hat{f}(x)$ papildomai glodina su Gauso branduolio funkcija. Maža to, tokia daugiklio forma leidžia analitiškai apskaičiuoti integralo reikšmę, o modeliavimas parodė, jog jį naudojant gerokai sumažėja įverčių paklaidos. (2.60) formulė gali būti taikoma esant įvairiems projektuotų duomenų charakteristinės funkcijos įvertiniams. Aptariami du būdai. Pirmasis iš jų paremtas tankio aproksimacija Gauso skirstinių mišinio modeliu. Nagrinėjamu atveju naudojamas parametrinis charakteristinės funkcijos įvertinys:

$$\hat{\psi}_\tau(u) = \sum_{i=1}^{\hat{q}_\tau} \hat{p}_{i,\tau} e^{iu\hat{x}_{i,\tau} - u^2 \hat{\varepsilon}_{i,\tau}^2 / 2}. \quad (2.63)$$

Į (2.60) įstačius (2.63), gaunama:

$$\begin{aligned} \hat{f}(x) &= \frac{A(d)}{\#T} \sum_{\tau \in T} \sum_{i=1}^{\hat{q}_\tau} \hat{p}_{i,\tau} \int_0^\infty e^{iu(\hat{x}_{i,\tau} - \tau^T x) - u^2(h + \hat{\varepsilon}_{i,\tau}^2/2)} u^{d-1} du \\ &= \frac{A(d)}{\#T} \sum_{\tau \in T} \sum_{i=1}^{\hat{q}_\tau} \hat{p}_{i,\tau} I_{d-1} \left(\frac{\hat{x}_{i,\tau} - \tau^T x}{\sqrt{\hat{\varepsilon}_{i,\tau}^2 + 2h}} \right) \left(\sqrt{\hat{\varepsilon}_{i,\tau}^2 + 2h} \right)^{-d}; \end{aligned} \quad (2.64)$$

čia

$$I_j(y) = \operatorname{Re} \left[\int_0^\infty e^{iyt - t^2/2} t^j dt \right]. \quad (2.65)$$

Pažymėtina, jog čia galima nagrinėti tik realią išraiškos dalį (menamųjų dalių suma turi būti lygi nuliui), nes tankio įvertis $\hat{f}(x)$ gali įgyti tik realiąsias reikšmes. Pasirinkta glodinimo daugiklio forma e^{-hu^2} leidžia susieti glodinimo parametą h su projekcijų klasterių dispersijomis – skaičiavimuose dispersijos tiesiog padidinamos dydžiu $2h$. Apskaičiuojama (2.65) išraiška. Pažymėjus

$$C_j(y) = \int_0^\infty \cos(yt) \cdot e^{-t^2/2} \cdot t^j dt, \quad (2.66)$$

$$S_j(y) = \int_0^\infty \sin(yt) \cdot e^{-t^2/2} \cdot t^j dt, \quad (2.67)$$

galioja lygybė

$$\int_0^\infty e^{-iyt-t^2/2} t^j dt = C_j(y) + iS_j(y). \quad (2.68)$$

Integruojant dalimis, gaunama:

$$C_j(y) = e^{-\frac{t^2}{2}} t^{j-1} \cos(yt) \Big|_0^\infty + \int_0^\infty e^{-\frac{t^2}{2}} \left((j-1)t^{j-2} \cos(yt) - yt^{j-1} \sin(yt) \right) dt = \mathbb{1}_{\{j=1\}} + (j-1)C_{j-2}(y) - yS_{j-1}(y), j \geq 1. \quad (2.69)$$

Analogiškai išreiškus $S_j(y)$ bei atsižvelgus į j indekso apribojimus, gaunamos rekurentinės lygtys:

$$C_j(y) = (j-1)C_{j-2}(y) - yS_{j-1}(y), j \geq 2, \quad (2.70)$$

$$C_1(y) = 1 - yS_0(y), \quad (2.71)$$

$$S_j(y) = (j-1)S_{j-2}(y) - yC_{j-1}(y), j \geq 2, \quad (2.72)$$

$$S_1(y) = yC_0(y). \quad (2.73)$$

Funkcijoms $C_0(y)$ bei $S_0(y)$ apskaičiuoti pasinaudojama tuo, kad

$$(S_0(y))'_y = \int_0^\infty t \cos(yt) \cdot e^{-t^2/2} dt = C_1(y). \quad (2.74)$$

Iš (2.71) ir (2.74) gaunama, kad S_0 tenkina diferencialinę lygtį

$$S'_0(y) = 1 - yS_0(y), \quad S_0(0) = 0. \quad (2.75)$$

Ši lygtis sprendžiama S_0 skleidžiant Teiloro eilute:

$$S'_0(y) = \sum_{l=0}^\infty c_{l+1} (l+1) y^{l+1} = 1 - \sum_{l=2}^\infty c_{l-1} y^l. \quad (2.76)$$

Sulyginus koeficientus prie panašių narių, randamos jų reikšmės:

$$\begin{aligned} c_0 &= 0, c_1 = 1, \\ c_l &= -c_{l-2}/l, l \geq 2. \end{aligned} \quad (2.77)$$

Taigi,

$$S_0(y) = \sum_{l=0}^{\infty} \frac{(-1)^l y^{2l+1}}{(2l+1)!!} = y - \frac{y^3}{3!!} + \frac{y^5}{5!!} - \frac{y^7}{7!!} + \dots \quad (2.78)$$

C_0 randama iš (2.66) išraiškos:

$$\begin{aligned} C_0(y) &= \int_0^{\infty} \cos(yt) \cdot e^{-t^2/2} dt = \frac{1}{2} \int_{-\infty}^{\infty} \cos(yt) \cdot e^{-t^2/2} dt \\ &= \frac{1}{2} \int_{-\infty}^{\infty} (\cos(yt) - i \sin(yt)) \cdot e^{-t^2/2} dt = \sqrt{\frac{\pi}{2}} e^{-y^2/2}. \end{aligned} \quad (2.79)$$

Ieškomo integralo (2.60) reikšmė

$$I_j(y) = C_j(y). \quad (2.80)$$

Apvertimo formulės modifikuotas tankio įvertinys. Vienas iš apvertimo formulės metodo, apibrėžto (2.60), trūkumų yra tas, kad šiuo įvertiniu aprašomas Gauso skirstinių mišinio modelis (2.49) (kai $f_i = \varphi_i$) gerai vertina tik jam artimo pasiskirstymo stebinių tankį. Aproximuojant tiriamą tankį Gauso skirstinių mišiniu dažnai apvertimo formulės tankio įvertinys tampa sudėtingas dėl didelio komponentų su mažomis *apriorinėmis* tikimybėmis skaičiaus. Jų skaičių galima sumažinti įvedant triukšmo klasterį.

Nagrinėjamas modifikuotas algoritmas, sukurtas remiantis daugiamačių Gauso skirstinių mišinio modelio naudojimu. Pritaikius apvertimo formulę (2.55) apibrėžiamas parametrinis tolygaus skirstinio tankio charakteristinės funkcijos parametrinis įvertinys:

$$\hat{\psi}(u) = \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}. \quad (2.81)$$

Tankio įverčio skaičiavimo formulėje (2.60) charakteristinės funkcijos įvertinys konstruojamas kaip Gauso skirstinių mišinio ir tolygaus skirstinio charakteristinių funkcijų sąjunga su atitinkamomis *apriorinėmis* tikimybėmis:

$$\hat{\psi}_{\tau}(u) = \sum_{i=1}^{\hat{q}_{\tau}} \hat{p}_{i,\tau} e^{iu\hat{\kappa}_{i,\tau} - u^2 \hat{\epsilon}_{i,\tau}^2/2} + \hat{p}_{0,\tau} \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}; \quad (2.82)$$

čia antrasis narys aprašo tolygaus pasiskirstymo triukšmo klasterį, \hat{p}_0 – triukšmo klasterio svoris, $a = a(\tau)$, $b = b(\tau)$. Remdamiesi nustatytais tolygaus skirstinio parametru įverčiais ir projektuotais duomenimis, galima apibrėžti:

$$a = (\tau^T x)_{\min} - \frac{(\tau^T x)_{\max} - (\tau^T x)_{\min}}{2(n-1)}, \quad (2.83)$$

$$b = (\tau^T x)_{\max} + \frac{(\tau^T x)_{\max} - (\tau^T x)_{\min}}{2(n-1)}. \quad (2.84)$$

Įrašius (2.82) į (2.60), gaunama:

$$\begin{aligned} \hat{f}(x) = & \frac{A(d)}{\#T} \sum_{\tau \in T} \left[\sum_{i=1}^{\hat{q}_\tau} \hat{p}_{i,\tau} \int_0^\infty e^{iu(\hat{x}_{i,\tau} - \tau^T x) - u^2(h + \hat{\xi}_{i,\tau}^2/2)} u^{d-1} du \right. \\ & \left. + \frac{2\hat{p}_{0,\tau}}{b-a} \int_0^\infty e^{iu\left(\frac{a+b}{2} - \tau^T x\right) - u^2 h} \cdot \sin\left(\frac{(b-a)u}{2}\right) \cdot u^{d-2} du \right]. \end{aligned} \quad (2.85)$$

Naudojant žymėjimus kaip ir (2.64), galima užrašyti:

$$\begin{aligned} \hat{f}(x) = & \frac{A(d)}{\#T} \sum_{\tau \in T} \left[\sum_{i=1}^{\hat{q}_\tau} \hat{p}_{i,\tau} I_{d-1} \left(\frac{\hat{x}_{i,\tau} - \tau^T x}{\sqrt{\hat{\xi}_{i,\tau}^2 + 2h}} \right) (\hat{\xi}_{i,\tau}^2 + 2h)^{\frac{d}{2}} \right. \\ & \left. + \frac{2\hat{p}_{0,\tau}}{b-a} J_{d-2} \left(\frac{a+b-2\tau^T x}{2\sqrt{2h}}, \frac{b-a}{2\sqrt{2h}} \right) \cdot (2h)^{-\frac{d-1}{2}} \right]; \end{aligned} \quad (2.86)$$

čia $I_j(y)$ išraiška yra kaip ir (2.65), o jo reikšmė – (2.80) bei

$$J_j(y, t) = \mathcal{R}e \left[\int_0^\infty e^{iyu - u^2/2} \cdot \sin(tu) \cdot u^j du \right]. \quad (2.87)$$

Integruojant gaunama:

$$\begin{aligned} \int_0^\infty e^{iyu - \frac{u^2}{2}} \cdot \sin(tu) \cdot u^j du &= \int_0^\infty (\cos(yu) + i \sin(yu)) \cdot \sin(tu) \cdot \\ e^{-\frac{u^2}{2}} \cdot u^j du &= \int_0^\infty \left(\frac{\sin(y+t)u + \sin(t-y)u}{2} + i \frac{\cos(y-t)u - \cos(y+t)u}{2} \right) \cdot e^{-\frac{u^2}{2}} \cdot \\ u^j du &= \frac{1}{2} S_j(y+t) + \frac{1}{2} S_j(t-y) + i \frac{1}{2} C_j(y-t) - i \frac{1}{2} C_j(y+t), \end{aligned} \quad (2.88)$$

čia $S_j(y)$ ir $C_j(y)$ apibrėžti (2.66) ir (2.67). Ieškomo integralo (2.87) reikšmė

$$J_j(y, t) = \frac{1}{2} S_j(y+t) + \frac{1}{2} S_j(t-y). \quad (2.89)$$

Anksčiau minėta procedūra vadinama *modifikuotu apvertimo formulės tankio įvertiniu*.

Remiantis panašių statistikų išvedimais [101] sukonstruota NI kriterijaus statistika yra apibrėžiama:

$$\mathcal{T} = n^{-1} \sum_{i=1}^n |f(z_{(i)}) - \hat{f}(z_{(i)})|. \quad (2.90)$$

\mathcal{T} nepriklauso nuo pakankamo imties tūrio (≥ 32), tačiau priklauso nuo duomenų dimensijos. Pasirinkta naudoti tokios formos kriterijaus statistiką:

$$\mathcal{T}^* = -\log(\mathcal{T}) = -\log \left(n^{-1} \sum_{i=1}^n |f(z_{(i)}) - \hat{f}(z_{(i)})| \right). \quad (2.91)$$

Ši statistikos forma (2.91) pasižymėjo mažiausiu jautrumu nuo žvalgomojo tyrimo taškų. Eksperimentiškai nustatyta, kad statistika \mathcal{T}^* turi Johnsono S_U pasiskirstymą, kurio tankio funkcija apibrėžiama [106]:

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}} g' \left(\frac{x-\xi}{\lambda} \right) \exp \left(-0,5 \left[\gamma + \delta g \left(\frac{x-\xi}{\lambda} \right) \right]^2 \right);$$

čia $g(y) = \ln[y + \sqrt{y^2 + 1}]$, $g'(y) = \frac{1}{\sqrt{y^2 + 1}}$, kai $x \in (-\infty, +\infty)$. Parametrai: δ – formos, ($\delta > 0$), γ – formos, ξ – padėties, λ – sklaidos, ($\lambda > 0$). Šio skirstinio pasiskirstymo funkcija:

$$F(x) = \int_{-\infty}^x f(u) du.$$

Plačiau Johnsono ir Pearsono skirstinių tipai aptariami [106, 107]. \mathcal{T}^* statistikos Johnsono S_U pasiskirstymo parametų įverčiai skirtingoms dimensijoms yra pateikiami 2.1 lentelėje.

2.1 lentelė. Statistikos \mathcal{T}^* Johnsono S_U pasiskirstymo parametų įverčiai

Parametras	Žymėjimas	Įvertis
d=2		
Padėties	$\hat{\xi}$	4,342807
Sklaidos	$\hat{\lambda}$	0,585038
Formos	$\hat{\delta}$	1,498293
Formos	$\hat{\gamma}$	0,764906
d=5		
Padėties	$\hat{\xi}$	7,025845
Sklaidos	$\hat{\lambda}$	0,088023
Formos	$\hat{\delta}$	0,895003
Formos	$\hat{\gamma}$	0,400035
d=10		
Padėties	$\hat{\xi}$	5,195174
Sklaidos	$\hat{\lambda}$	1,578613
Formos	$\hat{\delta}$	2,24856
Formos	$\hat{\gamma}$	-1,83037

Statistikai \mathcal{T}^* buvo patikrintos invariantiškumo ir suderinamumo savybės. Invariantiškumo savybė yra patvirtinama, nes pačioje statistikoje naudojami standartizuoti dydžiai. Suderinamumo savybė patvirtinama eksperimentiniu būdu (žr. 3.2 skyriuje).

2.3. Modeliavimo tyrimui taikyti statistiniai skirstiniai

Šiame skyriuje aprašomi vienmačiai ir daugiamačiai skirstiniai. Jie pateikiami grupėmis su parinktais parametų rinkiniais. Smulkesnė pasiskirstymų informacija pateikiama 1 priede.

2.3.1. Modeliavimo tyrimui taikyti vienmačiai skirstiniai

Modeliavimo tyrimui buvo naudojama 12 skirstinių, kuriuos taikant vertinamas vienmačių kriterijų galingumas. Parinkta naudoti X. Romao [28] panaudotus įvairius parametrų rinkinius, kurie atspindi dažniausiai praktikoje sutinkamas problemas. Skirstiniai suskirstyti į tris grupes [28, P1]:

1. Simetrinių skirstinių grupė, kurią sudaro:
 - trys beta ($Beta(a, b)$) skirstinio atvejai – $Beta(0,5; 0,5)$, $Beta(1; 1)$ ir $Beta(2; 2)$; čia a ir b – formos parametrai;
 - trys Koši ($Cauchy(t, s)$) skirstinio atvejai – $Cauchy(0; 0,5)$, $Cauchy(0; 1)$ ir $Cauchy(0; 2)$; čia t ir s – padėties ir sklaidos parametrai;
 - vienas Laplaso ($Laplace(t, s)$) skirstinio atvejis – $Laplace(0; 1)$; čia t ir s – padėties ir sklaidos parametrai;
 - vienas Logistinio ($Logistic(t, s)$) skirstinio atvejis – $Logistic(2; 2)$; čia t ir s – padėties ir sklaidos parametrai;
 - keturi Stjudento ($t(v)$) skirstinio atvejai – $t(1)$, $t(2)$, $t(4)$ ir $t(10)$, čia v – laisvės laipsnių skaičius;
 - standartinis normalusis $N(0; 1)$ skirstinys.
2. Asimetrinių skirstinių grupė, kurią sudaro:
 - keturi $Beta(a, b)$ skirstinio atvejai – $Beta(2; 1)$, $Beta(2; 5)$, $Beta(4; 0,5)$ ir $Beta(5; 1)$;
 - penki chi-kvadrato ($Chi-squared(v)$) skirstinio atvejai – $\chi^2(1)$, $\chi^2(2)$, $\chi^2(4)$ ir $\chi^2(10)$; čia v – laisvės laipsnių skaičius;
 - šeši gama ($Gamma(a, b)$) skirstinio atvejai – $Gamma(2; 2)$, $Gamma(3; 2)$, $Gamma(5; 1)$, $Gamma(9; 1)$, $Gamma(15; 1)$ ir $Gamma(100; 1)$; čia a ir b yra formos ir sklaidos parametrai;
 - vienas Gumbelio ($Gumbel(t, s)$) skirstinio atvejis – $Gumbel(1; 2)$; čia t ir s – padėties ir sklaidos parametrai;
 - vienas lognormaliojo ($LN(t, s)$) skirstinio atvejis – $LN(0; 1)$; čia t ir s – padėties ir sklaidos parametrai;
 - keturi Veibulo ($Weibull(a, b)$) skirstinio atvejai – $Weibull(0,5; 1)$, $Weibull(1; 2)$, $Weibull(2; 3,4)$ ir $Weibull(3; 4)$; čia a ir b – padėties ir sklaidos parametrai.
3. Modifikuotų normaliųjų skirstinių grupė, kurią sudaro:
 - Standartinis normalusis skirstinys, sutrumpintas taškuose a ir b ($Trunc(a, b)$), kurio parametrai: $a = -1, b = 1$; $a = -2, b = 2$; $a = -3, b = 3$; $a = -2, b = 2$; $a = -3, b = 1$; $a = -3, b = 2$. Toliau tai pažymima „NORMAL1“.
 - Standartinis normalusis skirstinys su padėties triukšmo komponentu ($LoConN(p, a)$), kurio parametrai: $p = 0,3, a = 1$; $p = 0,4, a = 1$; $p = 0,5, a = 1$; $p = 0,3, a = 3$; $p = 0,4, a = 3$; $p = 0,5, a = 3$; $p = 0,3, a = 5$; $p = 0,4, a = 5$; $p = 0,5, a = 5$. Tikimybė p

normaliojo pasiskirstymo su vidurkiu a ir standartiniu nuokrypiu 1. Toliau tai pažymima „NORMAL2“.

- Standartinis normalusis skirstinys su sklaidos triukšmo komponentu ($ScConN(p, b)$), kurio parametrai: $p = 0,05, b = 0,25$; $p = 0,1, b = 0,25$; $p = 0,2, b = 0,25$; $p = 0,05, b = 2$; $p = 0,1, b = 2$; $p = 0,2, b = 2$; $p = 0,05, b = 4$; $p = 0,1, b = 4$; $p = 0,2, b = 4$.

Tikimybė p normaliojo pasiskirstymo su vidurkiu 0 ir standartiniu nuokrypiu b . Toliau tai pažymima „NORMAL3“.

- Standartinių normaliųjų skirstinių mišinys $MixN(p, a, b) = N(1 - p, 0,1) + N(p, a, b)$, kurio parametrai: $p = 0,3, a = 1, b = 0,25$; $p = 0,4, a = 1, b = 0,25$; $p = 0,2, a = 1, b = 0,25$; $p = 0,5, a = 1, b = 0,25$; $p = 0,3, a = 3, b = 0,25$; $p = 0,4, a = 3, b = 0,25$; $p = 0,5, a = 3, b = 0,25$; $p = 0,3, a = 1, b = 4$; $p = 0,4, a = 1, b = 4$; $p = 0,5, a = 1, b = 4$; $p = 0,3, a = 3, b = 4$; $p = 0,4, a = 3, b = 4$; $p = 0,5, a = 3, b = 4$. Šis mišinys susideda iš dviejų komponentių $1 - p$ ir p . Tikimybė p normaliojo pasiskirstymo su vidurkiu a ir standartiniu nuokrypiu b . Toliau tai pažymima „NORMAL4“.

2.3.2. Modeliavimo tyrimui taikyti daugiamačiai skirstiniai

Modeliavimo tyrimui buvo naudojama 12 skirstinių, kuriuos taikant vertinamas daugiamačių kriterijų galingumas. Parinkta naudoti S. Korkmazo [42] panaudotus įvairius parametrų rinkinius, kurie atspindi dažniausiai praktikoje sutinkamas problemas. Skirstiniai suskirstyti į keturias grupes [42, 46, P2]:

1. Simetrinių skirstinių grupė, kurią sudaro:
 - trys $Beta(a, b)$ skirstinio atvejai – $Beta(1; 1)$, $Beta(1; 2)$ ir $Beta(2; 2)$; čia a ir b – formos daugiamačiai parametrų vektoriai;
 - vienas $Cauchy(t, s)$ skirstinio atvejis – $Cauchy(0; 1)$; čia t ir s – padėties ir sklaidos daugiamačiai parametrų vektoriai;
 - vienas $Laplace(t, s)$ skirstinio atvejis – $Laplace(0; 1)$; čia t ir s – padėties ir sklaidos daugiamačiai parametrų vektoriai;
 - vienas $Logistic(t, s)$ skirstinio atvejis – $Logistic(0; 1)$; čia t ir s – padėties ir sklaidos daugiamačiai parametrų vektoriai;
 - keturi $t - Student(v)$ skirstinio atvejai – $t(2)$ ir $t(5)$; čia v – laisvės laipsnių skaičius.
2. Asimetrinių skirstinių grupė, kurią sudaro:
 - penki $Chi-squared(v)$ skirstinio atvejai – $\chi^2(1)$, $\chi^2(2)$, $\chi^2(5)$, $\chi^2(10)$ ir $\chi^2(15)$; čia v – laisvės laipsnių skaičius;
 - du $Gamma(a, b)$ skirstinio atvejai – $Gamma(0,5; 1)$ ir $Gamma(5; 1)$; čia a ir b – formos ir sklaidos daugiamačiai parametrų vektoriai;
 - vienas $Gumbel(t, s)$ skirstinio atvejis – $Gumbel(1; 2)$; čia t ir s – padėties ir sklaidos daugiamačiai parametrų vektoriai;
 - du $LN(t, s)$ skirstinio atvejai – $LN(0; 1)$ ir $LN(0; 0,25)$; čia t ir s – padėties ir sklaidos daugiamačiai parametrų vektoriai;

- keturi *Weibull*(β) skirstinio atvejai – *Weibull*(0.8), *Weibull*(1) ir *Weibull*(1,5); čia β – formos daugiatis parametru vektorius.
3. Normaliojo skirstinio mišinio grupė, kurią sudaro:
- Daugiatis normalusis skirstinys *MVNMIX*(a, b, c, d) su tankiu $aN(0, \Sigma_1) + (1 - a)N(b\mathbf{1}, \Sigma_2)$, čia 0 yra nulių vektorius stulpelis, $\mathbf{1}$ yra vektorius stulpelis su $\Sigma_1 = (1 - c)I + c\mathbf{1}\mathbf{1}^T$ ir $\Sigma_2 = (1 - d)I + d\mathbf{1}\mathbf{1}^T$. Parinkti parametrai: *MVNMIX* (0,5; 2; 0; 0), *MVNMIX* (0,5; 4; 0; 0), *MVNMIX* (0,5; 2; 0,9; 0), *MVNMIX* (0,5; 0,5; 0,9; 0), *MVNMIX* (0,5,0,5,0,9, 0,1), *MVNMIX*(0,5; 0,5; 0,9; 0,9), *MVNMIX*(0,7; 2; 0,9; 0,3), *MVNMIX*(0,3; 1; 0,9; 0,1).
4. Mišriųjų skirstinių grupė, kurią sudaro standartinis normalusis skirstinys su:
- Trimis atvejais *Laplace*(0, 1) skirstiniu. Pirmuoju atveju ($d - 1$) matmens standartinis normalusis skirstinys su 1 matmeniu *Laplace*(0, 1). Antruoju atveju $\lfloor d/2 \rfloor$ matmens standartinis normalusis skirstinys su $\lfloor d/2 \rfloor$ Laplaso skirstiniu. Trečias atvejis priešingas pirmajam.
 - Trimis atvejais $\chi^2(5)$ skirstiniu. Pirmuoju atveju ($d - 1$) matmens standartinis normalusis skirstinys su 1 matmeniu $\chi^2(5)$. Antruoju atveju $\lfloor d/2 \rfloor$ matmens standartinis normalusis skirstinys su $\lfloor d/2 \rfloor$ chi-kvadrato skirstiniu. Trečias atvejis priešingas pirmajam.
 - Trimis atvejais $t(5)$ skirstiniu. Pirmuoju atveju ($d - 1$) matmens standartinis normalusis skirstinys su 1 matmeniu $t(5)$. Antruoju atveju $\lfloor d/2 \rfloor$ matmens standartinis normalusis skirstinys su $\lfloor d/2 \rfloor$ Stjudento skirstiniu. Trečias atvejis priešingas pirmajam.
 - Devyniais atvejais *Beta*(α, β) skirstiniu. Pirmuoju atveju ($d - 1$) matmens standartinis normalusis skirstinys su 1 matmeniu *Beta*(a, b). Antruoju atveju $\lfloor d/2 \rfloor$ matmens standartinis normalusis skirstinys su $\lfloor d/2 \rfloor$ Beta skirstiniu. Trečias atvejis priešingas pirmajam. Parametrai $a = 1, b = 1$; $a = 1, b = 2$; $a = 2, b = 2$.

2.4. Kriterijų galios skaičiavimo procedūra

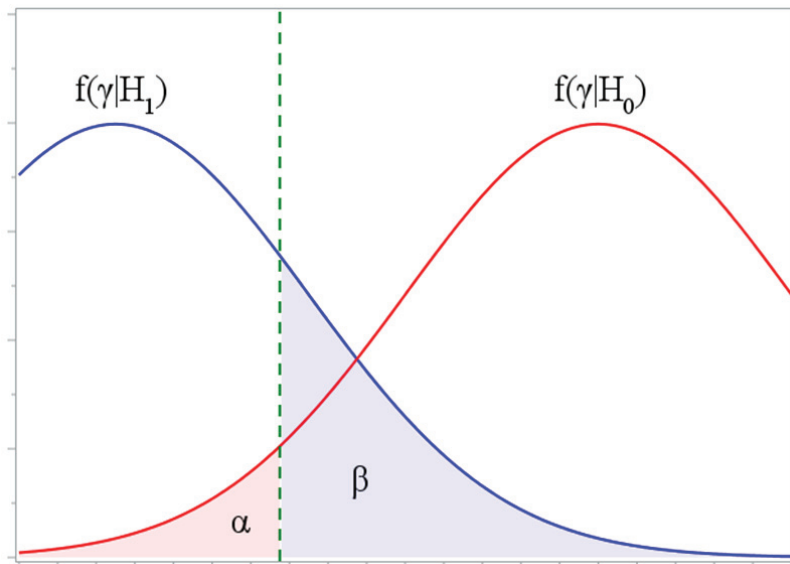
Kriterijaus galios apskaičiavimo procedūra gali būti atliekama trimis skirtingais būdais:

1. Kai kriterijaus statistikos pasiskirstymas (tikrasis arba kuriuo galima aproksimuoti) yra žinomas, tada taikoma 2.2 pav. galios apskaičiavimo schema.

Suderinamumo hipotezių tikrinimo kriterijų (pvz., AD, H1–H4) programinės realizacijos, kuriose yra kriterijaus galios skaičiavimo funkcionalumas, pateiktos [56]. 2.2 paveiksle iliustruojamas sąryšis tarp hipotezių tikrinimo klaidų tikimybių ir kriterijaus galios. Remiantis 2.2 pav. galima atlikti tokius žingsnius, kuriuos pasitelkus įvertinama kriterijaus galia.

1. Fiksuojama hipotezės tikrinimo I rūšies klaidos tikimybė α bei ją atitinkanti kritinė reikšmė pagal kriterijaus statistikos sąlyginį pasiskirstymą (teorinį arba empirinį), kai teisinga H_0 hipotezė ($f(\gamma|H_0)$).
2. Konkrečios imties atveju įvertinama kriterijaus statistikos reikšmė.
3. Konstruojamas kriterijaus statistikos pasiskirstymas, kai teisinga H_1 hipotezė ($f(\gamma|H_1)$). Įprastai, daroma prielaida, kad $f(\gamma|H_1)$ yra iš tokios pačios pasiskirstymų šeimos, kaip ir $f(\gamma|H_0)$ tik „paslinktas“ įvertintos kriterijaus statistikos reikšmės atžvilgiu. Pavyzdžiui, kai kriterijaus statistikos pasiskirstymas yra standartinis normalusis, tai poslinkis atliekamas vidurkio atžvilgiu. Kai kriterijaus statistikos pasiskirstymas yra chi-kvadrato, tai poslinkis atliekamas necentriškumo parametru.
4. Įvertinama II rūšies klaidos tikimybė β , kurią apriboja kritinė reikšmė ir $f(\gamma|H_1)$.
5. Apskaičiuojama kriterijaus galia $1 - \beta$.

Panagrinėjamas paprasčiausias atvejis. Taikomas chi-kvadrato kriterijus, kurio pasiskirstymo tankio funkcija yra gerai žinoma ($f(\gamma|H_0)$). Pasirenkama reikšmė $\alpha = 0,05$, kurią atitinka kritinė reikšmė χ_{krit}^2 . Pagal turimos imties tūrį apibrėžiamas dalinių intervalų kiekis ir laisvės laipsnių skaičius (dalinių intervalų kiekis – 1). Iš imties įvertinama kriterijaus statistikos reikšmė, kurios reikšmė prilyginama necentriškumo parametru. Sudaromas alternatyvios hipotezės tankis $f(\gamma|H_1)$. χ_{krit}^2 „atkerta“ dalį ploto po $f(\gamma|H_1)$ kreive, kurią vadiname II rūšies klaida (β). Tada įvertinama kriterijaus galia ($1 - \beta$).



2.2 pav. Kriterijų galios įvertinimo schema

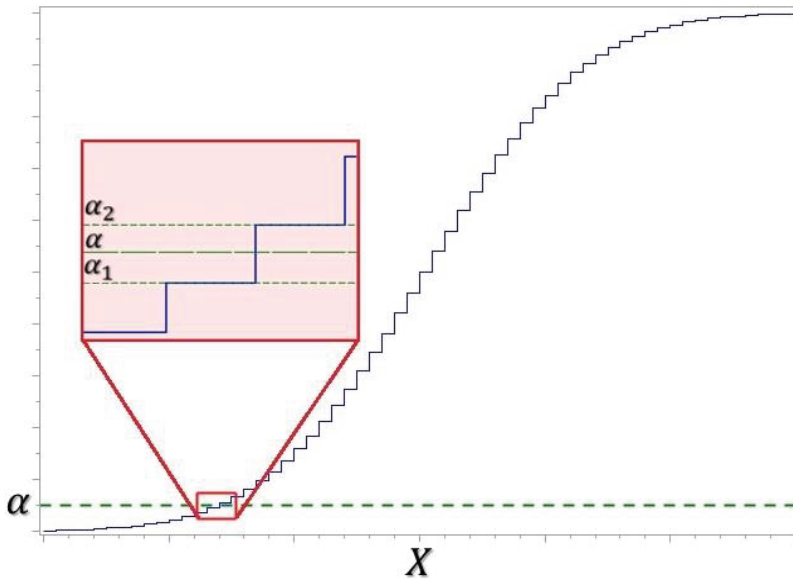
2. Kai kriterijaus statistikos pasiskirstymas nėra žinomas arba nėra galimybės jo aproksimuoti jokių kitų žinomu skirstiniu, tada atliekamas modeliavimo tyrimas, kurio metu formuojama laiptinė funkcija. Tada kriterijaus galia apskaičiuojama pagal (2.92) formulę ir 2.3 pav. Kitas empirinis kriterijaus galios apskaičiavimo būdas yra sudaryti tikimybinę lentelę, kurios pagrindu formuojama laiptinė funkcija, ir tada kriterijaus galia apskaičiuojama pagal (2.92) formulę ir 2.3 paveikslą.

Kai tikslus suderinamumo kriterijaus statistikos pasiskirstymas yra apibrėžiamas laiptine funkcija, kaip 2.3 pav., ši funkcija apskaičiuojama susumavus kriterijaus tikimybes. Taip galima gauti tą pačią kritinę reikšmę daugeliui gretimų reikšmingumo lygmenų α . Tiesinis kriterijų statistikos galios interpoliavimas naudojant reikšmingumo lygmenis (žr. 2.3 pav.), mažesnius (α_1) ir didesnius (α_2). Daugelis mokslininkų stengiasi parinkti tinkamiausius reikšmingumo lygmenis (pvz., [110]). Tokiu atveju (pagal 2.3 pav.) kriterijaus galią galima įvertinti atliekant tokius žingsnius [21].

1. Fiksuojamas reikšmingumo lygmuo α .
2. Parenkamos reikšmingumo lygmens reikšmės atitinkančios empirinės laiptinės pasiskirstymo funkcijos kritinės reikšmės $\gamma_1(\alpha)$ ir $\gamma_2(\alpha)$.
3. Apskaičiuojamos $\alpha_1 = P(T \geq \gamma_1(\alpha) | H_0)$ ir $\alpha_2 = P(T \geq \gamma_2(\alpha) | H_0)$ reikšmės.
4. Apskaičiuojama kriterijaus galia pagal formulę:

$$\mathbf{Gal\i{a}} = \frac{(\alpha - \alpha_1)P(T \geq \gamma_2(\alpha) | H_1) + (\alpha_2 - \alpha)P(T \geq \gamma_1(\alpha) | H_1)}{\alpha_2 - \alpha_1}; \quad (2.92)$$

čia $\gamma_1(\alpha)$ ir $\gamma_2(\alpha)$ kritinės reikšmės mažesnės už reikšmingumo lygmenį α . $\alpha_1 = P(T \geq \gamma_1(\alpha) | H_0)$ ir $\alpha_2 = P(T \geq \gamma_2(\alpha) | H_0)$ yra reikšmingumo lygmenys atitinkamai $\gamma_1(\alpha)$ ir $\gamma_2(\alpha)$.



2.3 pav. Statistikos, paremtos laiptine funkcija, reikšmingumo lygmenys

3. Nepriklausantis nuo kriterijaus statistikos pasiskirstymo, yra modeliavimas teisingai atmestų hipotezių skaičių padalinant iš pakartojimų kiekio.

Kai kriterijaus statistikos pasiskirstymas nėra žinomas, tada galios apskaičiavimo procedūra atliekama pagal šiuos žingsnius [21].

1. Sugeneruojamas duomenų rinkinys X_1, X_2, \dots, X_n pagal pasirinktą skirstinį.
2. Apskaičiuojama kriterijaus statistikos reikšmė. Jei statistikos reikšmė yra didesnė už taikomo kriterijaus kritinę reikšmę (kai $\alpha = 0,05$), tai H_0 hipotezė yra atmetama.
3. 1 ir 2 žingsniai kartojami K kartų ($K = 1\ 000\ 000$).
4. Apskaičiuojama kriterijaus galia kiekis/K , čia *kiekis* yra teisingai atmestų hipotezių skaičius.

2.5. Monte Karlo metodo taikymas skaičiuojant galią

1949 m. matematikai Džonas fon Neimanas ir Stanislavas Ulamas sukūrė ir išplėtojo Monte Karlo metodą kartu su pirmaisiais kompiuteriais [117]. Monte Karlo metodu vadinamas skaičiavimo algoritmas, paremtas statistiniu modeliavimu ir gautų rezultatų apdorojimu statistiniais metodais (dažnai dar vadinamas statistiniu modeliavimu). Šiuo metodu siekiama išvengti sudėtingų matematinių skaičiavimų ir taip sutrumpinti tyrimo trukmę. Monte Karlo metodas naudoja atsitiktinius skaičius (beveik pseudoatsitiktinius). Vilniaus universiteto Matematikos instituto profesorius Gediminas Stepanauskas įvardijo, kad modeliavimo procesą galima suskirstyti į keturias dalis [118]:

- tolygiai pasiskirsčiusių atsitiktinių dydžių generavimas intervale $[0; 1]$;
- gaunamos atsitiktinių dydžių reikšmės su parinktu pasiskirstymu;
- įvertinamos tiriamo objekto reikšmės;

- atliekama gautų rezultatų statistinė analizė.

Šis metodas buvo taikomas siekiant įvertinti suderinamumo hipotezės tikrinimo kriterijų galią įvairaus dydžio imtims ir pasiskirstymams. Norint padidinti šio metodo tikslumą buvo parinkta naudoti 1 000 000 replikų. Monte Karlo metodas taikytas pagal šiuos žingsnius.

1. Generuojami duomenys pagal pasirinktą pasiskirstymą su skirtingais parametru rinkiniais (žr. 2.3 skyrius) įvairaus dydžio imtims. Tam naudojamas *Mersennes Twister* atsitiktinių dydžių generatorius. Šis generavimo metodas taikomas daugelyje programinių įrangų *SAS*, *R*, *MATLAB* ir kt. Darbe naudotas *Mersennes Twister* generatorius, kuris yra realizuotas *SAS* programinėje įrangoje. Viena iš pagrindinių šio generatoriaus sudedamųjų dalių yra Mersenne skaičius, $M_p = 2^p - 1$; jį užrašant dvejetainė sistema susidaro p ilgio bitų eilutė, kurios visi bitai yra lygūs 1. Šiuo atveju naudoti $p = 32$ bitai. *Mersennes Twister* vykdomas [111]:

- Parenkamas r w -bitų skaičius $(x_i, i = 1, 2, \dots, r)$.
- Laikoma, kad

$$A = \begin{pmatrix} 0 & I_{w-1} \\ a_w & a_{w-1}, \dots, a_1 \end{pmatrix};$$

čia I_{w-1} yra $(w-1) \times (w-1)$ dydžio tapatybės matrica ir $a_i, i = 1, \dots, w$ įgyja 0 arba 1 reikšmes. Apibrėžiama:

$$x_{i+r} = \left(x_{i+s} \oplus \left(x_i^{(w:(s+1))} \middle| x_{i+1}^{(s:1)} \right) A \right);$$

čia $\left(x_i^{(w:(s+1))} \middle| x_{i+1}^{(s:1)} \right)$ parodo x_i reikšmingiausių (viršūnių) $w-s$ bitų ir x_{i+1} mažiausiai reikšmingų s bitų sujungimą.

- Iš eilės reikia atlikti keletą operacijų:

$$\begin{aligned} z &= x_{i+r} \oplus (x_{i+r} \gg t_1), \\ z &= z \oplus ((z \gg t_2) \& q_1), \\ z &= z \oplus ((z \gg t_3) \& q_2), \\ z &= z \oplus (x \gg t_4), \\ u_{i+r} &= \frac{z}{2^w - 1}; \end{aligned}$$

čia t_1, t_2, t_3 ir t_4 yra sveikieji skaičiai, q_1 ir q_2 – bitų kaukės¹⁰, o \oplus ir $\&$ yra „XOR“ ir „AND“ loginės bitų operacijos, $u_{i+r}, i = 1, 2, \dots$ sudaro atsitiktinių dydžių seką intervale (0; 1).

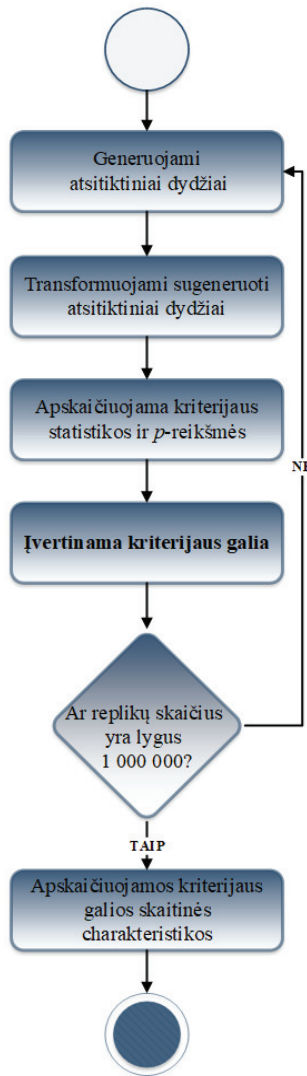
¹⁰ Kompiuterių moksle bitų kaukė yra duomenys, naudojami bitinėms operacijoms.

2. Atliekamas sugeneruotų duomenų transformavimas (normavimo ar kita kriterijui reikalinga transformacija). Sugeneruotą $\psi_j \in (0; 1)$ prilyginame pasiskirstymo funkcijai

$$\psi_j = F(X_j) \rightarrow X_j = F^{-1}(\psi_j).$$

3. Fiksuojamas reikšmingumo lygmuo α . Apskaičiuojama suderinamumo hipotezės tikrinimo kriterijaus statistikos ir p -reikšmės.
4. Kriterijaus galia įvertinama pirmuoju arba antruoju būdu. Tai priklauso nuo to, ar pasirinkto kriterijaus statistikos pasiskirstymas yra žinomas (žr. 2.4 skyrių).
5. 1–4 žingsniai kartojami 1 000 000 kartų.
6. Apskaičiuojamos kriterijaus galios skaitinės charakteristikos (vidurkis, standartinis nuokrypis, mediana, didžiausioji ir mažiausioji reikšmės). Nustatomi galingiausi suderinamumo hipotezės tikrinimo kriterijai pagal pasiskirstymą, imties dydį ir kitas sąlygas.

Modeliavimo procesą Monte Karlo metodu atspinti struktūrinė schema, kuri pateikiama 2.5 paveiksle.



2.4 pav. Kriterijaus galios įvertinimo Monte Karlo metodu schema

2.6. Paskirstytųjų skaičiavimų taikymas modeliavimo tyrimui

Šiame darbe yra atliktas platus modeliavimo tyrimas, kuris susideda iš dviejų dalių.

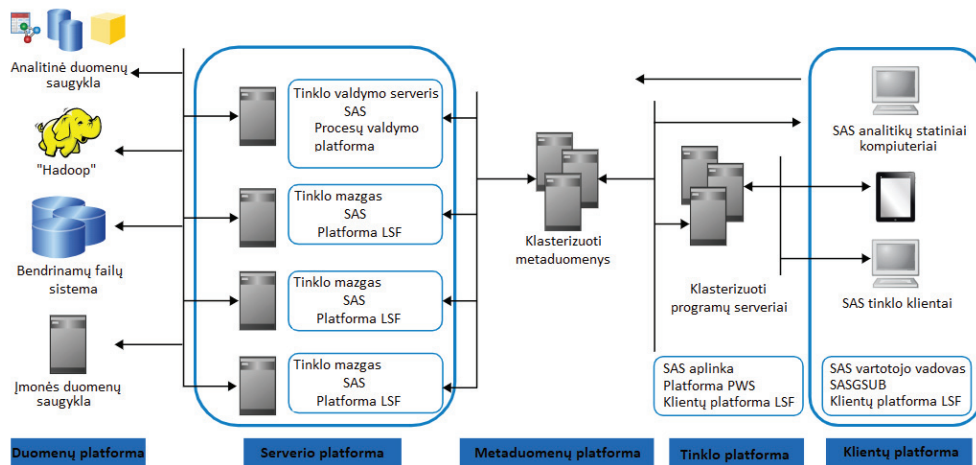
1. Vienmačių suderinamumo kriterijų modeliavimas. Tai sudaro: 42 kriterijai, 6 imčių tūriai, 67 duomenų variantai ir 1 000 000 nepriklausomųjų imčių.
2. Daugiamačių suderinamumo kriterijų modeliavimas. Tai sudaro: 13 kriterijų, 6 imčių tūriai, 48 duomenų variantai, 3 duomenų dimensijos ir 1 000 000 nepriklausomųjų imčių.

Šis tyrimas yra labai imlus laikui, todėl siekiant sutrumpinti modeliavimo proceso trukmę pasirinkta naudoti paskirstytuosius skaičiavimus. Paskirstytieji skaičiavimai yra keleto kompiuterių, kurie tarpusavyje komunikuoja per tinklą, naudojimas. Šie skaičiavimai sprendžiamą problemą suskirsto į daugybę užduočių, kurias sprendžia vienas ar keli kompiuteriai. Taip yra subalansuojamas darbo krūvis ir sutrumpinama tyrimo trukmė. Kiekvienas kompiuteris paskirstytoje skaičiavimo aplinkoje yra vadinamas mazgu [119].

Paskirstytiesiems skaičiavimams atlikti buvo taikytas *SAS Grid*[®] programinės įrangos technologinis sprendimas, kuriame skaičiavimo užduotys paskirstomos keliems kompiuteriams, esantiems tinkle. Taikant šį technologinį sprendimą darbo krūviai paskirstomi kompiuterių tinklo klasteryje. Toks darbo krūvio paskirstymas leidžia:

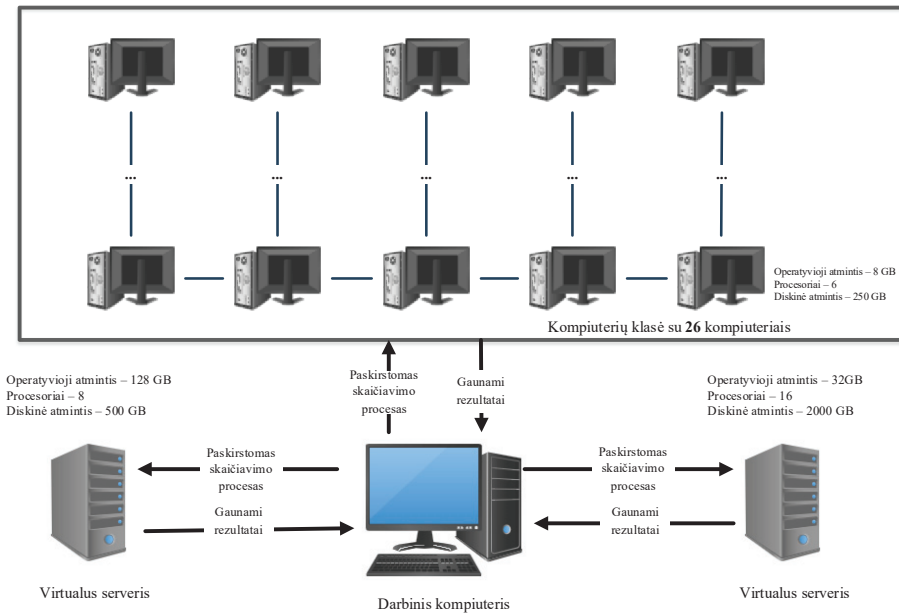
1. Subalansuoti darbo krūvį – keli vartotojai *SAS* aplinkoje gali paskirstyti darbo krūvius.
2. Pagreitinti apdorojimą – vartotojai gali paskirstyti padalintas *SAS* užduotis ir jas vykdyti lygiagrečiai.
3. Suplanuoti darbą – vartotojai gali planuoti užduotis, kurios tinkamu laiku automatiškai pradėdamos vykdyti.

SAS Grid[®] veikimo architektūra pateikiama 2.5 pav. su naudojamomis *SAS Workload Orchestrator* ir *SAS Job Flow Scheduler* aplinkomis.



2.5 pav. SAS tinklo skaičiavimo procedūros veikimo schema

Šiam modeliavimo tyrimui visiškai *SAS Grid*[®] programinės įrangos technologinio sprendimo funkcionalumas nėra būtinas, todėl naudota architektūra, apimanti tik darbą su nutolusiais tinklo mazgais (kompiuteriais). Tada 2.5 pav. schema tampa paprastesnė, ji pateikiama 2.6 paveiksle.



2.6 pav. Paskirstytojo modeliavimo proceso schema

2.7. Metodinės dalies išvados ir apibendrinimas

Apibendrinant suderinamumo su normaliuoju skirstiniu hipotezės kriterijų skaičiavimo metodiką galima padaryti išvadas.

1. Pagal atliktą vienmačių kriterijų analizę nustatyta, kad dauguma iš jų yra paremti asimetrijos bei eksceso koeficientais ir empirine pasiskirstymo funkcija. Pasirinkta naudoti R programiniame pakete realizuotus vienmačius kriterijus. Pasiūlyta naują branduolio funkciją taikyti N -metrika paremtam kriterijui. Pasiūlytos branduolio funkcijos privalumas: jos forma parenkama taip, kad kriterijaus sudarymo metu dėl imties reikšmių atsirandantis poslinkis būtų minimizuotas. Naujas kriterijus buvo realizuotas *SAS*.
2. Pagal atliktą daugiamačių kriterijų analizę nustatyta, kad dauguma iš jų yra paremti rangine statistika ir kovariacine matrica. Pasirinkta naudoti R programiniame pakete realizuotus daugiamačius kriterijus. Pasiūlytas naujas daugiamačias kriterijus, paremtas empirinio pasiskirstymo tankio vidutiniu absoliučiuoju nuokrypiu nuo teorinio pasiskirstymo tankio. Naudojant šį kriterijų, tankio įvertinys yra paremtas duomenų projektavimo ir apvertimo formulės (2.55) taikymu. Naujas kriterijus buvo realizuotas *SAS*.
3. Atlikti paskirstymai atspindi daugybę skirtingų praktikoje sutinkamų duomenų rinkinių, kurie dėl savo specifiškumo tyrėjams sukelia nemažai iššūkių. Kai kurias situacijas galima apibrėžti to paties pasiskirstymo įvairiais parametru rinkiniais.

III. TYRIMŲ REZULTATAI IR JŲ APTARIMAS

Šioje dalyje pristatomi suderinamumo hipotezės vienmačių ir daugiamačių kriterijų palyginimo modeliavimo būdu rezultatai. Pateikiamas kriterijų pritaikymas realiems duomenims. Pateikiamos tyrimus apibendrinančios išvados.

3.1. Vienmačių suderinamumo hipotezės tikrinimo kriterijų galios lyginamoji analizė

Šitame skyriuje pateikiamas išsamus vienmačių suderinamumo hipotezės tikrinimo kriterijų lyginamasis modeliavimo tyrimas. Šis tyrimas atliktas siekiant įvertinti pasirinktų kriterijų, skirtų patikrinti normalumo prielaidą, galią. Modeliavimo tyrimas atliekamas taikant Monte Karlo metodą. Šio tyrimo metu atsižvelgiama į imties dydį, reikšmingumo lygmenį ($\alpha = 0,05$) ir duomenų pasiskirstymą (beta, Koši, Laplaso, logistinis, Stjudento, chi-kvadrato, gama, Gumbelio, lognormalusis, Veibulo ir modifikuoti standartiniai normalieji). Modeliavimo tyrimas buvo atliktas taikant 42 kriterijus (įskaitant ir naujai pasiūlytą) 32, 64, 128, 256, 512, 1024 baitų dydžio sugeneruotoms ir standartizuotoms duomenų imtims. Šis tyrimas buvo atliktas naudojant 1 000 000 nepriklausomųjų imčių.

Pasiūlytas suderinamumo hipotezės tikrinimo kriterijus *NI* buvo modeliuojamas su įvairiais skirtingais parametrų rinkiniais. Kriterijaus parametrų rinkiniai buvo atrinkti eksperimentiniu būdu: sklaidos parametro *a* reikšmės naudotos nuo 0,001 iki 0,99, įdubos parametro *b* reikšmės naudotos nuo 0,01 iki 10, viršūnės formos parametro *c* reikšmės naudotos nuo 0,5 iki 50 taikant įvairų žingsnelį. Parametriui *c* taško 1 aplinkoje taikytas 0,001 žingsnelis ir nustatyta, kad tolstant nuo 1 tikslumas (kriterijaus galia) mažėja. Dėl šios priežasties visuose kriterijaus nagrinėtuose variantuose parametras $c = 1$. *NI* kriterijus buvo galingiausias su parametrų reikšmėmis $a = 0,95$; $b = 0,25$ ir $c = 1$ [P1]. *NI* kriterijaus aprašymas pateiktas 2.1.5 poskyryje.

Atliekant modeliavimo tyrimą taikyti lygiagretieji ir paskirstytieji skaičiavimai. 42 suderinamumo hipotezės tikrinimo kriterijų galia buvo skaičiuojama įvairiems imties dydžiams bei skirtingiems skirstiniams. Gauta daug modeliavimo rezultatų. Pasiskirstymai buvo sugrupuoti į simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupes. Šių grupių galios ir standartinio nuokrypio reikšmės kiekvienam kriterijui ir visiems nagrinėtiems imties dydžiams yra pateikiamos 3.1–3.3 lentelėse. Palyginus šiose lentelėse pateiktas reikšmes galima nustatyti, kad galingiausi kriterijai simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms yra Hoskingo (*H1*) tik mažiems imties dydžiams ir *NI* dideliems imties dydžiams. Visais nagrinėtais atvejais *NI* sklaida buvo mažesnė už *H1*. Palyginus *NI* kriterijų su Bakšajevio (Baksh) kriterijumi, nustatyta, kad visais nagrinėtais atvejais *NI* kriterijus buvo galingesnis. Pagal 3.1–3.2 lenteles galima pastebėti, kad, esant dideliems imties dydžiams, daugumos nagrinėtų suderinamumo hipotezės tikrinimo kriterijų galia yra artima 1. Išskyrus D'Agostino (*DA*) kriterijų, kurio galia žymiai mažesnė už kitų kriterijų.

3.1 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis simetrinių skirstinių grupei su skirtingais imties dydžiais

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
AD	0,714	0,169	0,799	0,123	0,863	0,089	0,909	0,062	0,939	0,041	0,957	0,031
Baksh	0,211	0,032	0,472	0,061	0,866	0,053	0,995	0,005	0,999	83E-8	0,999	83E-8
BCMR	0,718	0,171	0,809	0,123	0,875	0,089	0,92	0,059	0,947	0,038	0,962	0,023
BHS	0,431	0,189	0,551	0,194	0,663	0,175	0,752	0,139	0,818	0,108	0,868	0,084
BHSBS	0,68	0,191	0,778	0,14	0,85	0,105	0,903	0,07	0,938	0,044	0,959	0,03
BM1	0,628	0,173	0,754	0,13	0,853	0,096	0,917	0,064	0,949	0,038	0,962	0,022
BM2	0,726	0,166	0,835	0,118	0,905	0,08	0,945	0,049	0,965	0,029	0,974	0,017
BS	0,717	0,181	0,81	0,129	0,877	0,09	0,92	0,06	0,947	0,039	0,961	0,024
CC1	0,658	0,177	0,743	0,135	0,816	0,103	0,877	0,075	0,921	0,051	0,948	0,033
CC2	0,712	0,176	0,805	0,121	0,873	0,085	0,92	0,056	0,949	0,035	0,962	0,02
CHI2	0,663	0,21	0,778	0,177	0,842	0,153	0,884	0,144	0,941	0,089	0,945	0,096
CVM	0,591	0,135	0,733	0,093	0,805	0,069	0,855	0,049	0,919	0,028	0,949	0,017
ChenS	0,717	0,174	0,806	0,126	0,871	0,091	0,915	0,062	0,943	0,042	0,96	0,026
Coin	0,735	0,171	0,83	0,117	0,891	0,081	0,93	0,053	0,952	0,034	0,963	0,021
DA	0,266	0,121	0,295	0,095	0,314	0,066	0,319	0,043	0,315	0,028	0,311	0,021
DAP	0,723	0,185	0,82	0,13	0,883	0,09	0,924	0,059	0,948	0,037	0,962	0,023
DH	0,709	0,174	0,805	0,126	0,877	0,09	0,925	0,059	0,95	0,036	0,963	0,022
DLDMLRn	0,705	0,18	0,8	0,128	0,869	0,09	0,918	0,06	0,947	0,038	0,962	0,023
DLDMXAPD	0,725	0,174	0,817	0,126	0,883	0,088	0,925	0,057	0,95	0,036	0,962	0,022
DLDMZEPD	0,73	0,178	0,826	0,122	0,889	0,083	0,929	0,054	0,952	0,034	0,963	0,021
EP	0,706	0,171	0,797	0,123	0,859	0,094	0,91	0,06	0,946	0,038	0,96	0,025
Filli	0,712	0,166	0,805	0,121	0,875	0,087	0,922	0,058	0,949	0,036	0,962	0,022
GG	0,658	0,166	0,76	0,125	0,85	0,092	0,915	0,062	0,949	0,037	0,962	0,022
GLB	0,712	0,171	0,798	0,124	0,863	0,09	0,909	0,062	0,939	0,041	0,957	0,027
GMG	0,787	0,132	0,862	0,092	0,914	0,062	0,946	0,04	0,965	0,025	0,975	0,014
H1	0,799	0,166	0,862	0,118	0,852	0,09	0,999	53E-6	0,999	21E-8	0,999	53E-8
H2	0,728	0,209	0,8	0,153	0,775	0,116	0,764	0,085	0,762	0,062	0,766	0,045
H3	0,67	0,241	0,751	0,188	0,716	0,138	0,698	0,1	0,692	0,073	0,694	0,053
H4	0,631	0,255	0,706	0,218	0,662	0,166	0,641	0,118	0,634	0,084	0,636	0,061
JB	0,643	0,173	0,762	0,13	0,856	0,096	0,918	0,063	0,949	0,038	0,963	0,022
KS	0,585	0,104	0,723	0,072	0,789	0,056	0,836	0,041	0,905	0,025	0,939	0,016
LF	0,671	0,178	0,758	0,133	0,828	0,1	0,883	0,072	0,921	0,049	0,947	0,033
MI	0,632	0,143	0,676	0,098	0,705	0,068	0,724	0,046	0,736	0,029	0,745	0,019
N1	0,245	0,01	0,585	0,015	0,97	0,004	0,999	49E-8	0,999	49E-8	0,999	49E-9
SF	0,715	0,166	0,807	0,121	0,876	0,087	0,923	0,057	0,949	0,036	0,962	0,022
SW	0,718	0,172	0,808	0,124	0,874	0,089	0,919	0,06	0,946	0,039	0,962	0,024
SWRG	0,694	0,185	0,775	0,133	0,834	0,1	0,882	0,074	0,916	0,055	0,946	0,04
ZQ	0,513	0,23	0,574	0,201	0,628	0,16	0,668	0,123	0,696	0,092	0,717	0,066

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
ZQQstar	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15
ZQstar	0,513	0,231	0,576	0,202	0,63	0,161	0,669	0,124	0,697	0,091	0,718	0,064
ZW1	0,71	0,178	0,799	0,129	0,864	0,095	0,91	0,066	0,94	0,044	0,959	0,028
ZW2	0,715	0,17	0,806	0,122	0,869	0,087	0,912	0,06	0,939	0,041	0,957	0,027

3.2 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis asimetrinių skirstinių grupei su skirtingais imties dydžiais

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
AD	0,729	0,159	0,835	0,112	0,908	0,068	0,949	0,037	0,969	0,022	0,984	0,015
Baksh	0,173	0,013	0,385	0,026	0,809	0,027	0,997	0,002	0,999	14E-8	0,999	22E-9
BCMR	0,749	0,15	0,856	0,1	0,924	0,056	0,958	0,03	0,977	0,018	0,991	0,01
BHS	0,529	0,183	0,664	0,167	0,769	0,138	0,855	0,102	0,915	0,066	0,95	0,038
BHSBS	0,538	0,207	0,652	0,182	0,747	0,15	0,834	0,113	0,902	0,072	0,944	0,046
BM1	0,644	0,182	0,802	0,122	0,901	0,068	0,952	0,033	0,975	0,018	0,989	0,011
BM2	0,737	0,156	0,859	0,102	0,931	0,056	0,965	0,028	0,981	0,016	0,993	0,008
BS	0,506	0,236	0,588	0,226	0,665	0,205	0,738	0,177	0,805	0,143	0,859	0,109
CC1	0,735	0,155	0,846	0,103	0,919	0,059	0,957	0,031	0,977	0,019	0,991	0,01
CC2	0,579	0,23	0,682	0,196	0,777	0,148	0,856	0,096	0,916	0,051	0,956	0,025
CHI2	0,645	0,203	0,799	0,149	0,881	0,102	0,934	0,058	0,965	0,028	0,98	0,018
CVM	0,594	0,14	0,755	0,098	0,836	0,064	0,887	0,035	0,935	0,018	0,957	0,013
ChenS	0,756	0,147	0,862	0,097	0,928	0,054	0,961	0,029	0,978	0,018	0,991	0,01
Coin	0,48	0,225	0,556	0,218	0,63	0,202	0,7	0,184	0,769	0,161	0,831	0,13
DA	0,237	0,147	0,223	0,128	0,209	0,108	0,198	0,089	0,191	0,072	0,186	0,057
DAP	0,705	0,176	0,826	0,118	0,91	0,066	0,955	0,034	0,977	0,019	0,99	0,01
DH	0,724	0,161	0,845	0,107	0,921	0,058	0,957	0,03	0,977	0,019	0,991	0,01
DLDMLRn	0,576	0,227	0,686	0,197	0,78	0,156	0,856	0,114	0,912	0,075	0,949	0,046
DLDMXAPD	0,726	0,162	0,843	0,109	0,918	0,062	0,955	0,033	0,975	0,02	0,989	0,012
DLDMZEPD	0,545	0,237	0,639	0,219	0,724	0,189	0,798	0,153	0,858	0,115	0,905	0,081
EP	0,739	0,16	0,846	0,107	0,913	0,066	0,951	0,033	0,975	0,02	0,988	0,01
Filli	0,732	0,157	0,842	0,107	0,915	0,061	0,953	0,032	0,974	0,02	0,989	0,011
GG	0,672	0,202	0,805	0,132	0,898	0,073	0,949	0,035	0,973	0,019	0,988	0,011
GLB	0,725	0,162	0,831	0,115	0,905	0,071	0,947	0,039	0,969	0,023	0,984	0,015
GMG	0,683	0,171	0,751	0,148	0,809	0,122	0,859	0,098	0,901	0,072	0,932	0,05
H1	0,816	0,148	0,896	0,095	0,896	0,068	0,999	0,015	0,999	0,003	0,999	2E-7
H2	0,744	0,191	0,833	0,141	0,826	0,106	0,83	0,077	0,836	0,053	0,845	0,037
H3	0,697	0,216	0,784	0,17	0,764	0,132	0,764	0,098	0,77	0,069	0,781	0,049
H4	0,669	0,231	0,749	0,189	0,715	0,149	0,71	0,111	0,714	0,08	0,726	0,057
JB	0,662	0,179	0,808	0,12	0,904	0,067	0,953	0,033	0,975	0,018	0,989	0,011
KS	0,582	0,111	0,736	0,081	0,81	0,059	0,863	0,036	0,921	0,018	0,945	0,012
LF	0,671	0,177	0,786	0,123	0,872	0,09	0,929	0,053	0,959	0,03	0,976	0,018
MI	0,644	0,198	0,731	0,162	0,798	0,125	0,843	0,093	0,872	0,068	0,89	0,051

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
N1	0,464	0,061	0,761	0,028	0,99	0,002	0,999	5E-8	0,999	41E-8	0,999	55E-9
SF	0,736	0,156	0,846	0,105	0,918	0,059	0,955	0,031	0,975	0,02	0,989	0,011
SW	0,753	0,149	0,859	0,098	0,925	0,055	0,959	0,03	0,977	0,018	0,991	0,01
SWRG	0,758	0,145	0,861	0,097	0,927	0,056	0,96	0,03	0,977	0,018	0,99	0,011
ZQ	0,22	0,085	0,21	0,066	0,202	0,049	0,191	0,036	0,183	0,028	0,174	0,021
ZQQstar	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	4E-15
ZQstar	0,57	0,17	0,639	0,139	0,693	0,11	0,732	0,085	0,761	0,065	0,781	0,049
ZW1	0,749	0,143	0,852	0,094	0,918	0,054	0,955	0,03	0,976	0,017	0,989	0,01
ZW2	0,764	0,145	0,87	0,093	0,932	0,051	0,962	0,028	0,98	0,017	0,991	0,009

3.3 lentelė. Vidutinė kriterijų galia ir standartinis nuokrypis modifikuotų normaliųjų skirstinių grupei su skirtingais imties dydžiais

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
AD	0,662	0,176	0,756	0,134	0,825	0,098	0,872	0,072	0,905	0,055	0,931	0,042
Baksh	0,222	0,02	0,519	0,036	0,931	0,017	0,999	86E-6	0,999	4E-13	0,999	8E-14
BCMR	0,652	0,179	0,756	0,135	0,831	0,096	0,88	0,068	0,913	0,05	0,935	0,038
BHS	0,463	0,189	0,585	0,185	0,676	0,16	0,744	0,132	0,796	0,108	0,834	0,089
BHSBS	0,568	0,214	0,701	0,172	0,787	0,128	0,847	0,093	0,89	0,068	0,918	0,052
BM1	0,494	0,173	0,652	0,149	0,779	0,112	0,86	0,079	0,904	0,056	0,93	0,041
BM2	0,641	0,176	0,77	0,135	0,854	0,093	0,904	0,063	0,934	0,043	0,953	0,03
BS	0,587	0,218	0,688	0,186	0,77	0,147	0,833	0,112	0,881	0,084	0,916	0,06
CC1	0,567	0,187	0,669	0,157	0,761	0,119	0,83	0,087	0,878	0,065	0,912	0,05
CC2	0,576	0,202	0,675	0,166	0,763	0,127	0,833	0,093	0,887	0,066	0,923	0,047
CHI2	0,566	0,226	0,728	0,168	0,808	0,125	0,866	0,089	0,914	0,058	0,939	0,041
CVM	0,557	0,138	0,708	0,095	0,779	0,071	0,833	0,05	0,897	0,033	0,93	0,024
ChenS	0,656	0,18	0,759	0,136	0,833	0,096	0,882	0,07	0,915	0,051	0,937	0,038
Coin	0,579	0,214	0,691	0,18	0,781	0,137	0,846	0,099	0,889	0,07	0,918	0,052
DA	0,314	0,156	0,342	0,141	0,367	0,12	0,388	0,097	0,405	0,077	0,418	0,06
DAP	0,617	0,201	0,733	0,158	0,818	0,11	0,872	0,076	0,906	0,055	0,93	0,042
DH	0,617	0,195	0,727	0,156	0,815	0,111	0,872	0,076	0,907	0,055	0,93	0,041
DLDMLRn	0,581	0,211	0,687	0,175	0,78	0,131	0,852	0,091	0,902	0,062	0,933	0,041
DLDMXAPD	0,651	0,189	0,754	0,146	0,831	0,104	0,879	0,073	0,912	0,055	0,935	0,041
DLDMZEPD	0,595	0,22	0,698	0,184	0,78	0,142	0,842	0,106	0,887	0,077	0,919	0,056
EP	0,64	0,179	0,748	0,136	0,819	0,096	0,865	0,07	0,906	0,056	0,931	0,046
Filli	0,637	0,178	0,743	0,136	0,823	0,098	0,877	0,069	0,911	0,049	0,933	0,036
GG	0,529	0,192	0,657	0,151	0,775	0,112	0,86	0,078	0,906	0,054	0,932	0,039
GLB	0,659	0,178	0,755	0,135	0,823	0,1	0,87	0,072	0,903	0,056	0,93	0,044
GMG	0,688	0,165	0,771	0,137	0,836	0,104	0,883	0,075	0,917	0,055	0,942	0,041
H1	0,743	0,187	0,816	0,138	0,799	0,104	0,999	22E-8	0,999	17E-9	0,999	95E-9
H2	0,711	0,197	0,775	0,153	0,741	0,117	0,723	0,087	0,716	0,064	0,716	0,047

Kriterijai	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
H3	0,688	0,209	0,749	0,167	0,704	0,127	0,68	0,093	0,67	0,068	0,668	0,051
H4	0,675	0,22	0,73	0,179	0,677	0,136	0,649	0,098	0,637	0,072	0,634	0,053
JB	0,515	0,176	0,662	0,149	0,783	0,111	0,861	0,079	0,904	0,056	0,93	0,041
KS	0,564	0,108	0,71	0,073	0,772	0,058	0,825	0,042	0,893	0,026	0,924	0,02
LF	0,626	0,184	0,724	0,143	0,796	0,109	0,85	0,082	0,889	0,062	0,917	0,048
MI	0,494	0,197	0,536	0,161	0,563	0,127	0,578	0,098	0,585	0,075	0,59	0,059
N1	0,243	0,009	0,581	0,014	0,972	0,003	0,999	43E-8	0,999	27E-9	0,999	69E-11
SF	0,642	0,177	0,747	0,135	0,826	0,096	0,879	0,068	0,912	0,049	0,934	0,036
SW	0,654	0,18	0,758	0,135	0,832	0,096	0,882	0,068	0,915	0,049	0,937	0,036
SWRG	0,643	0,188	0,746	0,145	0,818	0,101	0,864	0,076	0,901	0,059	0,931	0,046
ZQ	0,382	0,189	0,422	0,182	0,459	0,165	0,489	0,144	0,562	0,125	0,522	0,103
ZQQstar	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	8E-15	0,05	3E-15	0,05	2E-15
ZQstar	0,394	0,173	0,423	0,157	0,45	0,136	0,472	0,115	0,487	0,097	0,498	0,083
ZW1	0,635	0,181	0,737	0,14	0,816	0,1	0,87	0,072	0,906	0,053	0,931	0,039
ZW2	0,64	0,183	0,749	0,14	0,826	0,097	0,876	0,069	0,907	0,051	0,931	0,039

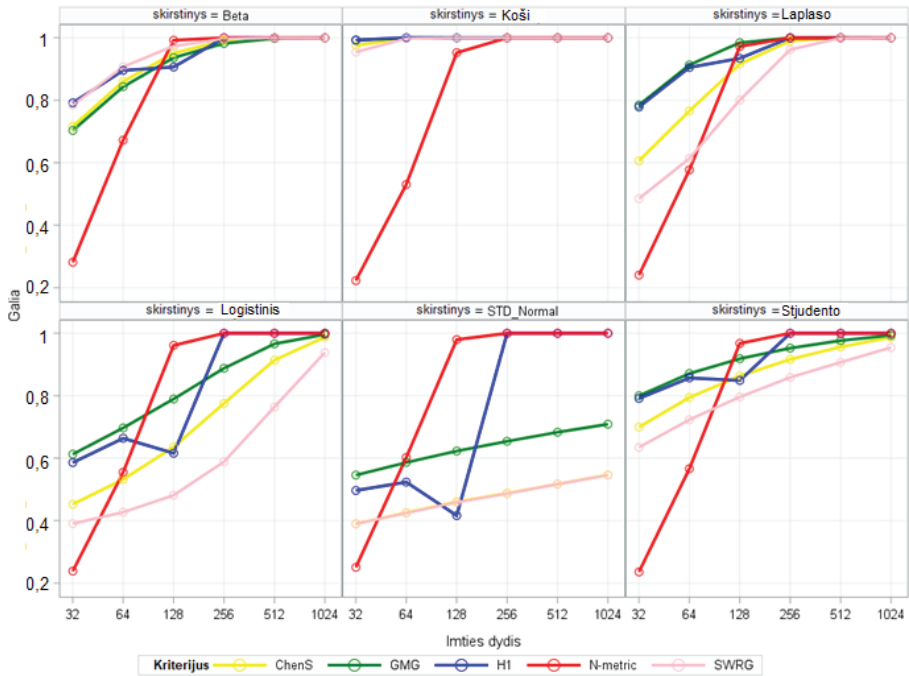
Atliktas papildomas tyrimas, siekiant įvertinti mažiausią imties dydį, ties kuriuo *NI* kriterijus tampa galingesnis simetriškų, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms. Atsižvelgiant į 3.1–3.3 lentelių rezultatus, *HoskingI* ir *NI* kriterijai buvo taikomi 80, 90, 100, 105, 110, 115 dydžio imtims. Gauti rezultatai pateikiami 3.4 lentelėje. Pagal gautus rezultatus nustatyta, kad *NI* kriterijus yra galingiausias, kai $n \geq 112$, simetrinių skirstinių grupei, taip pat, kai $n \geq 112$, asimetrinių skirstinių grupei ir kai $n \geq 88$ – modifikuotų normaliųjų skirstinių grupei. Iš 3.4 lentelės matyti, kad *NI* kriterijus yra galingiausias gama skirstinio atveju, kai $n \geq 32$. Taip pat nustatyta, kad Koši ir lognormaliojo skirstinių atvejais *NI* kriterijus yra galingiausias, kai $n \geq 255$. Tam įtakos turi ilgos šių skirstinių uodegos [**P1**].

3.4 lentelė. Minimalus imties dydis, kai *NI* kriterijus tampa galingiausias

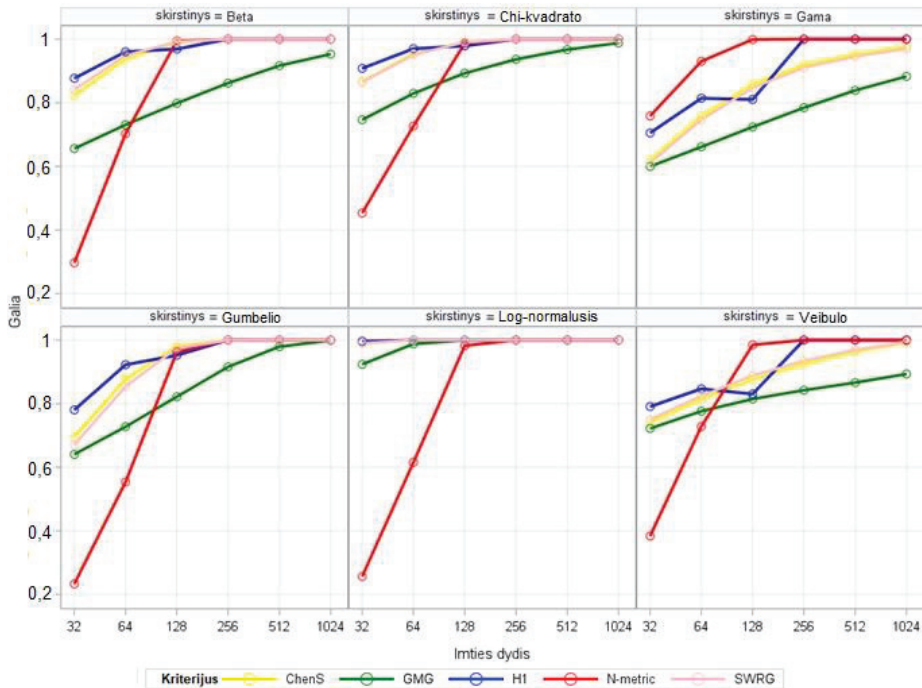
Nr.	Skirstinys	Skirstinio grupė	Minimalus imties dydis (n)
1.	Standartinis normalusis	Simetrinė	46
2.	Beta	Simetrinė	88
3.	Koši	Simetrinė	257
4.	Laplaso	Simetrinė	117
5.	Logistinis	Simetrinė	71
6.	Stjudento	Simetrinė	96
7.	Beta	Asimetrinė	108
8.	Chi-kvadrato	Asimetrinė	123
9.	Gama	Asimetrinė	< 32

Nr.	Skirstinys	Skirstinio grupė	Minimalus imties dydis (<i>n</i>)
10.	Gumbelio	Asimetrinė	125
11.	Lognormalusis	Asimetrinė	255
12.	Veibulo	Asimetrinė	65
13.	Normal1	Modifikuota	70
14.	Normal2	Modifikuota	93
15.	Normal3	Modifikuota	72
16.	Normal4	Modifikuota	117

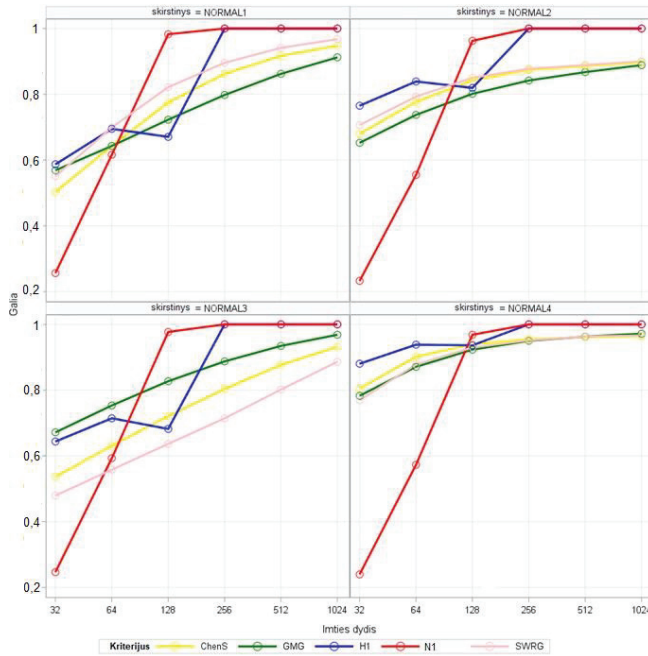
Siekiant papildyti 3.1–3.3 lentelėse pateiktus rezultatus, nubraižytos 5 galingiausių kriterijų galios kitimo diagramos kiekvienam nagrinėtam skirstiniui. Daugiamačių duomenų vizualizavimo grotelių metodu (angl. *trellis display*) [112] nubraižyti grafikai pateikiami 3.1–3.3 paveiksluose. Iš 3.1 pav. matyti, kad standartinio normaliojo skirstinio atveju *NI* kriterijus yra galingiausias, kai $n \geq 64$. 3.2 pav. parodo, kad gama skirstinio atveju *NI* kriterijus yra galingiausias visiems nagrinėtiems imčių dydžiams. Iš 3.3 pav. matyti, kad visiems nagrinėtiems modifikuotiems normaliesiems duomenų skirstiniams *NI* kriterijus yra galingiausias, kai $n \geq 128$. Apibendrinant 3.1–3.3 pav. galima teigti, kad Cheno ir Šapiro, Gelo, Miao ir Gastwirtho, Hoskingo1 bei modifikuoto Šapiro ir Vilko kriterijų galia didėja, didinant imties dydį. Pasiūlyto kriterijaus galia padidėja, kai imties dydis yra 128 ir išlieka artima 1 esant didesniems imties dydžiams.



3.1 pav. Penkių galingiausių kriterijų galios kitimas simetriniams skirstiniams



3.2 pav. Penkių galingiausių kriterijų galios kitimas asimetriniams skirstiniams



3.3 pav. Penkių galingiausių kriterijų galios kitimas modifikuotiems normaliesiems skirstiniams

3.2. Daugiamačių suderinamumo hipotezės tikrinimo kriterijų galios lyginamoji analizė

Šiame skyriuje pateikiamas daugiamačių suderinamumo hipotezės tikrinimo kriterijų lyginamasis modeliavimo tyrimas. Pasirinkti kriterijai (chi-kvadrato, Kramerio ir von Mises, Andersono ir Darlingo, Doorniko ir Hanseno, Roystono, Mardijos, Energijos, Lobato ir Velasco, Henze ir Zirklerio, Nikulino, Rao ir Rosono, Dzhaparidze ir Nikulino bei N2) lyginami pagal galią. Šiam tyrimui taip pat taikytas Monte Karlo modeliavimo metodas $n = 32, 64, 128, 256, 512, 1024$ dydžio imtims su reikšmingumo lygmeniu ($\alpha = 0,05$). Kriterijų palyginimui pasirinkti beta, Koši, Laplaso, logistinis, Stjudento, standartinis normalusis, chi-kvadrato, gama, Gumbelio, lognormalusis, Veibulo, mišrieji, normaliojo skirstinio mišinių duomenų pasiskirstymai. Šių skirstinių duomenys generuoti $d = 2, 5, 10$ dydžio dimensijoms. Modeliavimas atliktas naudojant 1 000 000 nepriklausomųjų imčių.

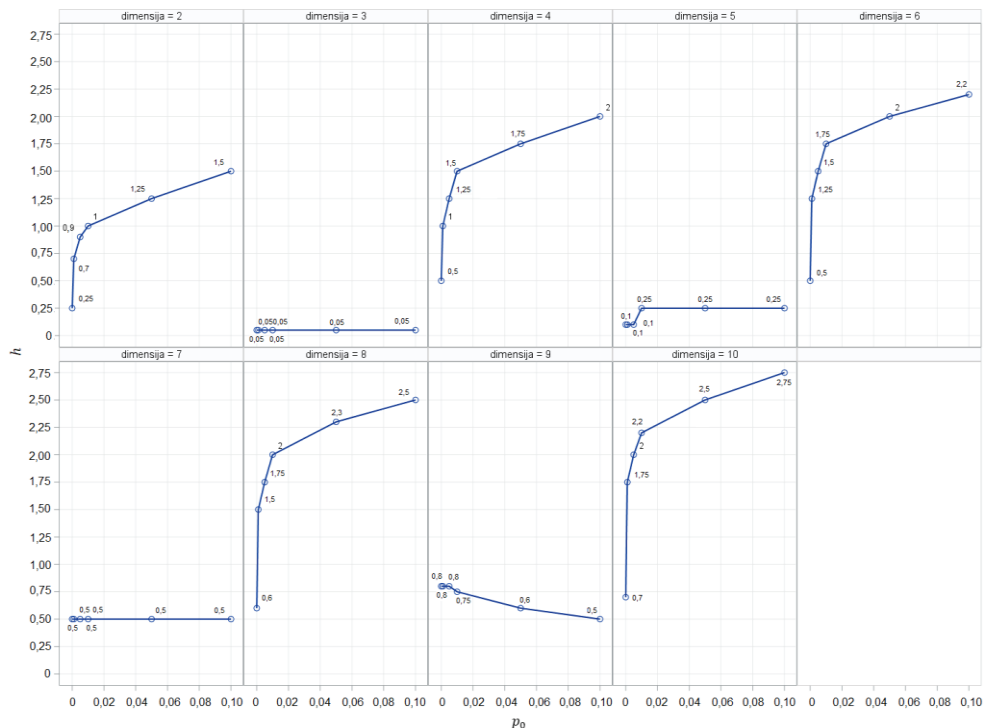
Pasiūlytas pasiskirstymo tankių skirtumo vertinimu paremtas daugiamačių kriterijus modeliuotas su įvairiomis skirtingomis glodumo (h) parametro reikšmėmis. Kriterijaus glodumo parametro reikšmės atrinktos eksperimentiniu būdu: nuo 0,05 iki 10 su žingsniu 0,1. Kriterijaus h parametro reikšmės nustatytos 2–10 dydžio dimensijoms su skirtingais triukšmo klasterio svoriais ($\hat{p}_0 = 0; 0,001; 0,005; 0,01; 0,05; 0,1$). Nustačius h reikšmes, kurioms esant vidutinės

statistikos (\mathcal{T}^*) reikšmės buvo didžiausios, tų h reikšmių aplinkoje atlikta patikslinanti h reikšmės paieška naudojant trumpesnę žingsnio ilgį. Gauti rezultatai grafiškai vaizduojami 3.4 pav., o tikslios parametrų reikšmės pateikiamos 2 priede (kartu su statistikos \mathcal{T}^* reikšme). Iš 3.4 pav. matyti, kad, kai d įgyja lyginę reikšmę, tai glodumo parametras didėja didinant triukšmo klasterio svorį, o priešingu atveju didinant p_0 , h nežymiai pakinta.

$N2$ kriterijaus modeliavimas atliktas su vidutinėmis h reikšmėmis [P2]:

- kai $d = 2$, tai $h = 1,05$;
- kai $d = 5$, tai $h = 0,1$;
- kai $d = 10$, tai $h = 2,4$.

Išsamesnis šio kriterijaus aprašymas pateiktas 2.2.4 poskyryje.



3.4 pav. Glodumo parametro kitimo priklausomybė nuo triukšmo klasterio svorio ir dimensijos

Atliekant daugiamačių kriterijų modeliavimo tyrimą taikyti lygiagretieji ir paskirstytieji skaičiavimai. 13 (įskaitant ir naujai pasiūlytą, parentą apvertimo formulę) daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia buvo skaičiuojama įvairiems imties dydžiams, naudojant skirtingus skirstinius ir jų mišinius. Dėl didelio kiekio modeliavimo rezultatų, pasiskirstymai buvo sugrupuoti į simetrinių, asimetrinių, mišriųjų ir normaliojo skirstinio mišinių grupes. Šių grupių vidutinės galios ir standartinio nuokrypio reikšmės kiekvienam kriterijui su visais nagrinėtais imties dydžiais, yra pateikiamos 3.5–3.8 lentelėse. Palyginus šiose

lentelėse pateiktas galios reikšmės nustatyta, kad naujai pasiūlytas $N2$ kriterijus yra galingiausias simetrinių ir mišriųjų skirstinių grupėms. Asimetrinių skirstinių grupėje galingiausi yra $N2$ (kai $d = 2$) ir Roy (kai $d = 5$ ir 10) kriterijai. Normaliojo skirstinio mišinių grupėje taip pat galingiausi yra $N2$ (kai $d = 2$ ir 5) ir Roy (kai $d = 10$ su $n = 256, 512, 1024$) kriterijai. Visais nagrinėtais atvejais $N2$ kriterijaus sklaida buvo mažiausia. Palyginus Mardia kriterijus (Mar1 ir Mar2), paremtus asimetrijos ir eksceso koeficientais, nustatyta, kad Mar1 yra galingiausias tik asimetrinių skirstinių grupei, o simetrinių skirstinių grupei šio kriterijaus galia yra mažiausia (lyginant su kitais kriterijais) [P2, K4, K5].

3.5 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis simetrinių skirstinių grupei

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
$d = 2$												
AD	0,651	0,2	0,778	0,151	0,867	0,099	0,92	0,054	0,939	0,028	0,945	0,021
CHI2	0,57	0,205	0,692	0,176	0,798	0,135	0,873	0,086	0,912	0,057	0,932	0,041
CVM	0,652	0,2	0,779	0,152	0,868	0,098	0,92	0,053	0,94	0,029	0,945	0,021
DH	0,677	0,18	0,809	0,128	0,892	0,076	0,932	0,041	0,945	0,025	0,949	0,022
DN	0,565	0,208	0,671	0,181	0,768	0,145	0,847	0,104	0,903	0,069	0,937	0,04
Energijos	0,65	0,189	0,77	0,145	0,86	0,095	0,914	0,058	0,941	0,034	0,948	0,024
HZ	0,644	0,186	0,765	0,143	0,853	0,097	0,906	0,063	0,936	0,039	0,947	0,027
LV	0,696	0,182	0,815	0,127	0,893	0,076	0,932	0,041	0,945	0,025	0,949	0,022
N2	0,999	0,003	0,999	95E-9	0,999	9E-10	0,999	61E-10	0,999	94E-10	0,999	3E-13
Mar1	0,532	0,189	0,617	0,17	0,681	0,144	0,721	0,126	0,743	0,114	0,758	0,108
Mar2	0,605	0,183	0,751	0,147	0,857	0,099	0,917	0,055	0,942	0,03	0,95	0,022
NRR	0,608	0,202	0,736	0,165	0,842	0,116	0,91	0,068	0,941	0,034	0,949	0,023
Roy	0,703	0,177	0,819	0,123	0,891	0,078	0,929	0,043	0,944	0,028	0,95	0,022
$d = 5$												
AD	0,644	0,184	0,791	0,14	0,883	0,084	0,929	0,043	0,942	0,025	0,946	0,021
CHI2	0,531	0,191	0,656	0,177	0,773	0,145	0,864	0,093	0,916	0,055	0,941	0,033
CVM	0,624	0,181	0,775	0,144	0,876	0,089	0,926	0,046	0,942	0,026	0,946	0,021
DH	0,735	0,149	0,864	0,086	0,924	0,044	0,941	0,026	0,946	0,023	0,949	0,022
DN	0,585	0,193	0,7	0,167	0,806	0,127	0,886	0,084	0,932	0,045	0,949	0,027
Energijos	0,632	0,171	0,758	0,137	0,863	0,089	0,921	0,051	0,945	0,028	0,949	0,022
HZ	0,622	0,172	0,755	0,146	0,856	0,098	0,91	0,062	0,94	0,035	0,949	0,024
LV	0,763	0,15	0,871	0,085	0,925	0,044	0,941	0,026	0,946	0,023	0,949	0,022
N2	0,985	0,041	0,989	0,027	0,988	0,027	0,987	0,03	0,981	0,032	0,985	0,03
Mar1	0,523	0,155	0,621	0,14	0,696	0,11	0,735	0,087	0,752	0,074	0,764	0,07
Mar2	0,637	0,162	0,792	0,131	0,89	0,074	0,934	0,036	0,948	0,024	0,95	0,023
NRR	0,602	0,186	0,739	0,159	0,851	0,108	0,916	0,06	0,944	0,031	0,95	0,022

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
Roy	0,784	0,135	0,875	0,084	0,924	0,046	0,941	0,029	0,947	0,023	0,95	0,021
d = 10												
AD	0,557	0,172	0,754	0,145	0,878	0,083	0,928	0,041	0,942	0,023	0,945	0,021
CHI2	0,473	0,198	0,604	0,173	0,726	0,156	0,824	0,114	0,891	0,071	0,928	0,042
CVM	0,534	0,171	0,728	0,148	0,865	0,091	0,922	0,047	0,941	0,025	0,945	0,021
DH	0,754	0,126	0,884	0,062	0,934	0,03	0,941	0,023	0,945	0,023	0,948	0,022
DN	0,599	0,194	0,704	0,161	0,817	0,123	0,896	0,078	0,936	0,041	0,949	0,024
Energijos	0,598	0,157	0,709	0,132	0,821	0,094	0,906	0,055	0,943	0,028	0,948	0,021
HZ	0,604	0,149	0,71	0,149	0,831	0,114	0,901	0,066	0,937	0,039	0,949	0,024
LV	0,791	0,125	0,893	0,062	0,935	0,03	0,941	0,023	0,945	0,023	0,948	0,022
N2	0,997	0,026	0,998	0,023	0,998	0,017	0,998	0,013	0,991	0,037	0,991	0,04
Mar1	0,458	0,128	0,592	0,116	0,676	0,096	0,733	0,06	0,747	0,053	0,756	0,049
Mar2	0,65	0,147	0,802	0,127	0,899	0,077	0,94	0,033	0,951	0,025	0,95	0,022
NRR	0,599	0,188	0,719	0,158	0,844	0,11	0,913	0,063	0,942	0,032	0,949	0,022
Roy	0,834	0,101	0,905	0,058	0,934	0,032	0,943	0,024	0,946	0,022	0,95	0,021

3.6 lentelė. Vidutinė daugiamachių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis asimetrinių skirstinių grupei

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
d = 2												
AD	0,634	0,196	0,744	0,146	0,827	0,11	0,897	0,08	0,954	0,044	0,985	0,016
CHI2	0,631	0,227	0,767	0,173	0,861	0,125	0,931	0,083	0,977	0,039	0,996	0,012
CVM	0,639	0,196	0,744	0,144	0,822	0,112	0,892	0,083	0,949	0,048	0,982	0,019
DH	0,852	0,12	0,956	0,057	0,995	0,013	0,999	59E-5	0,999	11E-8	0,999	56E-9
DN	0,55	0,209	0,657	0,171	0,724	0,132	0,774	0,114	0,816	0,104	0,864	0,088
Energijos	0,832	0,129	0,93	0,078	0,985	0,028	0,999	0,003	0,999	0,001	0,999	7E-5
HZ	0,811	0,13	0,906	0,086	0,968	0,042	0,995	0,01	0,999	61E-5	0,999	42E-7
LV	0,87	0,115	0,961	0,054	0,995	0,012	0,999	54E-5	0,999	95E-9	0,999	4E-11
N2	0,999	0,003	0,999	5E-11	0,999	64E-11	0,999	91E-11	0,999	3E-12	0,999	35E-13
Mar1	0,813	0,151	0,941	0,073	0,992	0,018	0,999	0,001	0,999	92E-8	0,999	11E-10
Mar2	0,63	0,244	0,776	0,191	0,876	0,139	0,947	0,086	0,988	0,033	0,999	0,007
NRR	0,639	0,204	0,759	0,154	0,841	0,12	0,915	0,087	0,968	0,047	0,993	0,016
Roy	0,877	0,106	0,962	0,052	0,995	0,012	0,999	58E-5	0,999	27E-8	0,999	12E-10
d = 5												
AD	0,614	0,188	0,763	0,136	0,869	0,1	0,946	0,059	0,984	0,019	0,995	0,003
CHI2	0,6	0,212	0,779	0,155	0,892	0,108	0,965	0,056	0,994	0,015	0,999	0,002
CVM	0,608	0,186	0,761	0,135	0,869	0,101	0,947	0,059	0,985	0,02	0,995	0,003

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
DH	0,915	0,086	0,99	0,02	0,999	84E-5	0,999	46E-9	0,999	1E-10	0,999	37E-13
DN	0,551	0,203	0,675	0,156	0,748	0,121	0,812	0,107	0,872	0,089	0,926	0,063
Energijos	0,854	0,117	0,958	0,055	0,996	0,011	0,999	24E-5	0,999	12E-8	0,999	59E-9
HZ	0,798	0,128	0,907	0,088	0,974	0,038	0,998	0,007	0,999	14E-5	0,999	1E-10
LV	0,932	0,077	0,992	0,017	0,999	71E-5	0,999	33E-9	0,999	91E-9	0,999	51E-10
N2	0,982	0,055	0,989	0,043	0,997	0,018	0,997	0,017	0,999	0,007	0,999	22E-8
Mar1	0,803	0,143	0,954	0,063	0,997	0,008	0,999	67E-6	0,999	17E-8	0,999	87E-8
Mar2	0,623	0,197	0,791	0,158	0,908	0,107	0,978	0,043	0,997	0,011	0,999	0,003
NRR	0,61	0,196	0,763	0,145	0,869	0,109	0,95	0,066	0,989	0,023	0,999	0,003
Roy	0,945	0,066	0,993	0,016	0,999	62E-5	0,999	1E-7	0,999	34E-8	0,999	11E-10
d = 10												
AD	0,483	0,194	0,707	0,132	0,863	0,103	0,955	0,052	0,99	0,013	0,996	89E-5
CHI2	0,443	0,206	0,712	0,169	0,87	0,115	0,96	0,059	0,994	0,016	0,999	0,001
CVM	0,459	0,191	0,7	0,132	0,859	0,104	0,953	0,054	0,989	0,013	0,996	0,001
DH	0,944	0,065	0,998	0,007	0,999	44E-7	0,999	29E-8	0,999	9E-9	0,999	42E-11
DN	0,532	0,212	0,679	0,15	0,776	0,119	0,858	0,098	0,93	0,064	0,975	0,031
Energijos	0,829	0,122	0,956	0,054	0,997	0,008	0,999	75E-6	0,999	3E-8	0,999	73E-9
HZ	0,744	0,13	0,861	0,102	0,954	0,055	0,994	0,015	0,999	56E-5	0,999	51E-9
LV	0,96	0,054	0,998	0,005	0,999	28E-7	0,999	36E-8	0,999	54E-9	0,999	4E-12
N2	0,922	0,132	0,947	0,116	0,98	0,073	0,995	0,025	0,996	0,019	0,996	0,02
Mar1	0,693	0,15	0,931	0,078	0,997	0,009	0,999	78E-6	0,999	49E-8	0,999	7E-11
Mar2	0,573	0,192	0,746	0,142	0,898	0,106	0,978	0,046	0,998	0,008	0,999	3E-5
NRR	0,532	0,208	0,722	0,141	0,86	0,109	0,952	0,062	0,992	0,017	0,999	0,002
Roy	0,98	0,033	0,999	0,003	0,999	21E-7	0,999	29E-8	0,999	15E-9	0,999	67E-11

3.7 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis mišriųjų skirstinių grupei

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
d = 2												
AD	0,469	0,214	0,572	0,208	0,683	0,188	0,78	0,144	0,848	0,094	0,883	0,063
CHI2	0,408	0,202	0,476	0,215	0,571	0,216	0,667	0,198	0,763	0,16	0,842	0,107
CVM	0,463	0,211	0,567	0,208	0,679	0,188	0,778	0,144	0,847	0,094	0,883	0,065
DH	0,436	0,201	0,51	0,192	0,591	0,173	0,66	0,159	0,746	0,144	0,826	0,113
DN	0,439	0,2	0,511	0,195	0,59	0,192	0,674	0,18	0,763	0,147	0,835	0,105
Energijos	0,582	0,196	0,703	0,165	0,809	0,112	0,872	0,07	0,895	0,047	0,902	0,044
HZ	0,572	0,19	0,697	0,164	0,807	0,115	0,871	0,07	0,894	0,047	0,901	0,043
LV	0,453	0,203	0,547	0,195	0,667	0,163	0,749	0,13	0,808	0,109	0,857	0,088
N2	0,999	0,007	0,999	37E-5	0,999	2E-7	0,999	43E-9	0,999	7E-11	0,999	4E-12

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
Mar1	0,476	0,202	0,577	0,186	0,659	0,154	0,717	0,127	0,76	0,101	0,78	0,086
Mar2	0,412	0,193	0,527	0,216	0,651	0,209	0,762	0,162	0,843	0,108	0,884	0,066
NRR	0,444	0,202	0,533	0,203	0,641	0,191	0,741	0,16	0,827	0,111	0,878	0,074
Roy	0,451	0,184	0,513	0,183	0,572	0,178	0,643	0,165	0,72	0,138	0,764	0,115
d = 5												
AD	0,626	0,207	0,749	0,132	0,805	0,097	0,852	0,074	0,88	0,057	0,894	0,045
CHI2	0,466	0,217	0,582	0,204	0,67	0,169	0,729	0,132	0,763	0,113	0,789	0,102
CVM	0,585	0,211	0,726	0,143	0,791	0,105	0,841	0,083	0,875	0,063	0,893	0,048
DH	0,545	0,209	0,631	0,178	0,695	0,142	0,747	0,113	0,777	0,091	0,795	0,087
DN	0,538	0,216	0,684	0,177	0,771	0,121	0,825	0,096	0,865	0,075	0,891	0,057
Energijos	0,703	0,148	0,788	0,097	0,845	0,068	0,88	0,052	0,894	0,044	0,9	0,045
HZ	0,706	0,154	0,815	0,091	0,864	0,059	0,885	0,047	0,895	0,043	0,899	0,044
LV	0,553	0,21	0,628	0,179	0,692	0,142	0,751	0,113	0,81	0,091	0,883	0,062
N2	0,982	0,063	0,989	0,045	0,995	0,019	0,998	0,019	0,999	41E-7	0,999	8E-9
Mar1	0,584	0,174	0,675	0,133	0,722	0,105	0,75	0,092	0,766	0,085	0,778	0,085
Mar2	0,584	0,209	0,723	0,147	0,791	0,111	0,838	0,088	0,864	0,079	0,889	0,065
NRR	0,551	0,215	0,694	0,164	0,769	0,118	0,822	0,097	0,863	0,076	0,889	0,058
Roy	0,47	0,187	0,524	0,187	0,589	0,18	0,669	0,157	0,73	0,125	0,764	0,114
d = 10												
AD	0,688	0,161	0,753	0,095	0,775	0,092	0,802	0,091	0,833	0,086	0,866	0,072
CHI2	0,477	0,219	0,579	0,21	0,651	0,192	0,709	0,155	0,746	0,121	0,763	0,109
CVM	0,642	0,18	0,744	0,104	0,771	0,093	0,795	0,091	0,821	0,087	0,853	0,078
DH	0,58	0,207	0,69	0,143	0,736	0,11	0,761	0,097	0,778	0,091	0,791	0,087
DN	0,669	0,197	0,753	0,12	0,776	0,102	0,793	0,093	0,818	0,089	0,842	0,085
Energijos	0,719	0,104	0,744	0,086	0,777	0,084	0,823	0,077	0,875	0,058	0,897	0,043
HZ	0,745	0,106	0,78	0,081	0,821	0,081	0,87	0,061	0,892	0,044	0,899	0,043
LV	0,592	0,209	0,69	0,146	0,735	0,111	0,76	0,097	0,776	0,09	0,79	0,086
N2	0,916	0,148	0,942	0,095	0,94	0,093	0,968	0,078	0,995	0,033	0,997	0,016
Mar1	0,614	0,101	0,68	0,084	0,722	0,087	0,745	0,086	0,764	0,088	0,779	0,086
Mar2	0,731	0,169	0,796	0,098	0,795	0,094	0,811	0,094	0,84	0,092	0,861	0,082
NRR	0,679	0,188	0,753	0,115	0,774	0,1	0,79	0,092	0,814	0,089	0,837	0,086
Roy	0,475	0,194	0,529	0,195	0,602	0,184	0,689	0,156	0,745	0,13	0,769	0,122

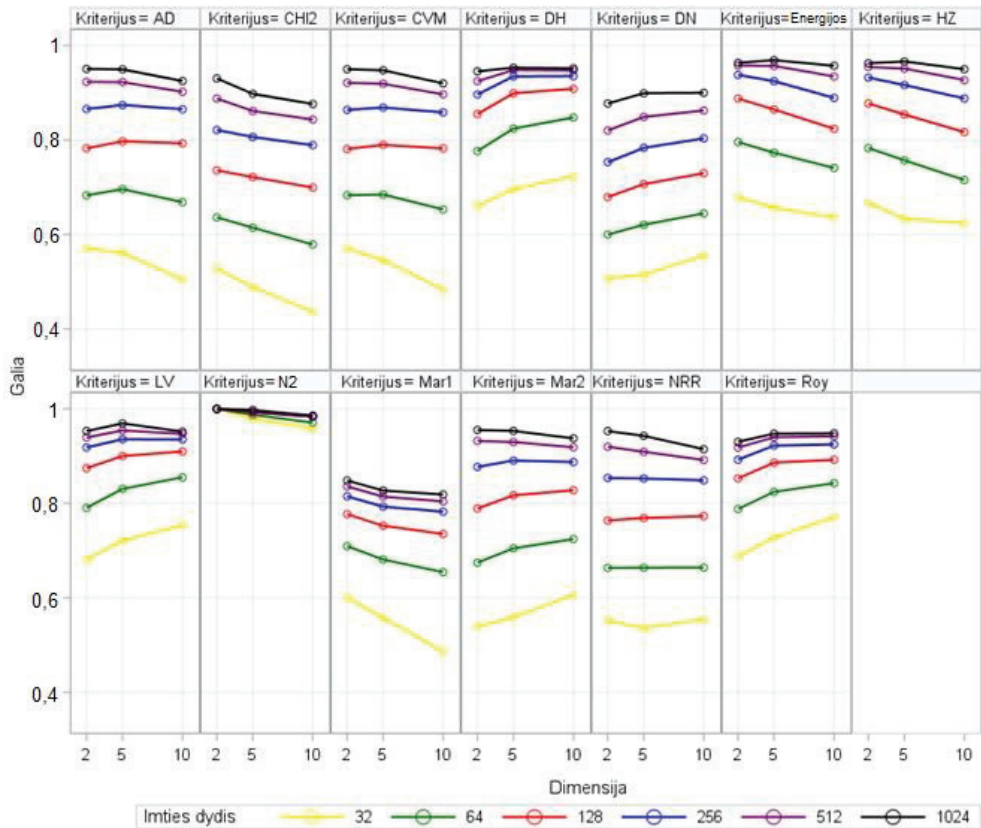
3.8 lentelė. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia ir standartinis nuokrypis normaliojo skirstinio mišinių grupei

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
d = 2												
AD	0,465	0,221	0,576	0,225	0,71	0,205	0,844	0,149	0,943	0,073	0,987	0,019
CHI2	0,422	0,211	0,508	0,228	0,618	0,229	0,738	0,196	0,845	0,136	0,917	0,077

	Imties dydis											
	32		64		128		256		512		1024	
	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD	Galia	STD
CVM	0,468	0,221	0,581	0,225	0,715	0,205	0,848	0,15	0,945	0,073	0,988	0,018
DH	0,56	0,221	0,74	0,178	0,893	0,105	0,974	0,04	0,998	0,006	0,999	22E-6
DN	0,428	0,205	0,503	0,211	0,582	0,215	0,685	0,193	0,791	0,157	0,882	0,111
Energijos	0,537	0,22	0,682	0,199	0,836	0,139	0,938	0,068	0,986	0,022	0,999	0,002
HZ	0,529	0,214	0,672	0,198	0,823	0,145	0,926	0,077	0,977	0,031	0,998	0,006
LV	0,588	0,223	0,752	0,176	0,895	0,104	0,974	0,04	0,998	0,006	0,999	22E-6
N2	0,999	0,004	0,999	2E-8	0,999	7E-9	0,999	19E-9	0,999	2E-11	0,999	9E-12
Mar1	0,433	0,211	0,544	0,204	0,633	0,178	0,701	0,152	0,733	0,138	0,757	0,132
Mar2	0,437	0,201	0,563	0,226	0,707	0,217	0,84	0,159	0,931	0,084	0,977	0,032
NRR	0,442	0,208	0,544	0,22	0,664	0,217	0,805	0,173	0,924	0,099	0,985	0,031
Roy	0,607	0,212	0,778	0,158	0,908	0,093	0,978	0,035	0,998	0,006	0,999	81E-6
d = 5												
AD	0,45	0,211	0,574	0,213	0,699	0,184	0,806	0,141	0,889	0,095	0,946	0,056
CHI2	0,399	0,196	0,491	0,218	0,598	0,212	0,702	0,183	0,787	0,145	0,857	0,105
CVM	0,441	0,206	0,563	0,213	0,689	0,186	0,798	0,144	0,883	0,098	0,942	0,059
DH	0,594	0,206	0,782	0,149	0,916	0,077	0,979	0,029	0,998	0,005	0,999	29E-5
DN	0,443	0,208	0,51	0,211	0,594	0,2	0,691	0,177	0,782	0,145	0,859	0,106
Energijos	0,503	0,21	0,64	0,193	0,781	0,145	0,894	0,085	0,963	0,038	0,992	0,011
HZ	0,485	0,199	0,621	0,201	0,761	0,163	0,877	0,106	0,95	0,052	0,985	0,02
LV	0,632	0,206	0,795	0,146	0,92	0,076	0,979	0,029	0,998	0,005	0,999	29E-5
N2	0,98	0,061	0,994	0,028	0,997	0,016	0,999	0,006	0,999	0,003	0,999	0,001
Mar1	0,384	0,199	0,516	0,196	0,619	0,169	0,694	0,138	0,736	0,115	0,758	0,103
Mar2	0,46	0,187	0,59	0,212	0,728	0,188	0,832	0,142	0,905	0,093	0,954	0,049
NRR	0,442	0,204	0,539	0,215	0,655	0,196	0,766	0,157	0,857	0,113	0,925	0,071
Roy	0,672	0,186	0,828	0,128	0,934	0,066	0,984	0,024	0,998	0,004	0,999	45E-5
d = 10												
AD	0,402	0,188	0,556	0,203	0,709	0,165	0,801	0,116	0,853	0,085	0,893	0,071
CHI2	0,392	0,19	0,47	0,212	0,587	0,207	0,688	0,17	0,76	0,138	0,818	0,115
CVM	0,396	0,185	0,537	0,201	0,692	0,172	0,79	0,122	0,846	0,089	0,886	0,074
DH	0,62	0,182	0,8	0,121	0,915	0,069	0,973	0,03	0,995	0,008	0,999	72E-5
DN	0,495	0,216	0,536	0,209	0,629	0,193	0,723	0,16	0,8	0,119	0,85	0,089
Energijos	0,476	0,204	0,599	0,182	0,723	0,143	0,834	0,1	0,912	0,062	0,96	0,033
HZ	0,487	0,165	0,581	0,189	0,708	0,166	0,815	0,118	0,887	0,081	0,939	0,052
LV	0,667	0,181	0,815	0,118	0,919	0,067	0,974	0,03	0,995	0,008	0,999	7E-4
N2	0,989	0,035	0,984	0,035	0,977	0,036	0,971	0,03	0,971	0,031	0,973	0,031
Mar1	0,287	0,162	0,472	0,175	0,582	0,15	0,669	0,119	0,711	0,095	0,739	0,085
Mar2	0,547	0,16	0,634	0,195	0,758	0,151	0,834	0,106	0,885	0,084	0,924	0,066
NRR	0,487	0,206	0,55	0,21	0,674	0,186	0,771	0,138	0,833	0,099	0,875	0,079
Roy	0,735	0,151	0,855	0,1	0,939	0,052	0,98	0,023	0,997	0,006	0,999	63E-5

Siekiant papildyti ir pabrėžti 3.5–3.8 lentelėse pateiktus rezultatus, nubraižytos jas apibendrinančios linijinės diagramos taikant daugiamačių duomenų vizualizavimo grotelių metodą (angl. *trellis display*). Gautas grafikas pateikiamas

3.5 paveiksle. Iš šio paveikslo matyti, kad mūsų pasiūlytas N2 kriterijus pasižymi kur kas didesne galia nei kiti nagrinėti kriterijai. Mar1 kriterijaus galia yra mažiausia iš kitų kriterijų. 3.5 pav. parodo, kad didinant imties dydį daugiamačių kriterijų galia didėja. Didinant dimensiją kriterijų galia 8 kriterijų (AD, CHI2, CVM, Energijos HZ, N2, Mar1 ir NRR) galia mažėja, o kitų (DH, DN, LV, Mar2 ir Roy) šiek tiek padidėja. Mažiams imties dydžiams galingiausi yra N2, Roy ir DH kriterijai. Dideliems imties dydžiams galingiausi yra N2, Energijos HZ ir LV kriterijai.



3.5 pav. Vidutinė daugiamačių suderinamumo hipotezės tikrinimo kriterijų galia visų tipų skirstinių grupėms

Taip pat daugiamačių kriterijų galia palyginta kiekvieno nagrinėto skirstinio atveju. Šiam palyginimui nubraižyti daugiamačiai grafikai, kurie yra pateikti 3 priede. Iš 3 priede esančių grafikų nustatyta, kad nagrinėti daugiamačiai kriterijai yra galingiausi Koši skirstinio duomenims, o silpniausi – standartinio normaliojo skirstinio atveju. Palyginus 3 priede pateiktus grafikus galima pastebėti, kad pristatytas N2 kriterijus yra vienas iš galingiausių. Už N2 kriterijų šiek tiek silpnesnis tik Roy kriterijus. Mar1 kriterijus silpniausias standartinio normaliojo, beta, Laplaso ir logistinio skirstinių atvejais. DH kriterijus silpniausias gama, Gumbelio ir lognormaliojo skirstinių atvejais. Palyginus 3 priedo grafikus matyti, kad didėjant

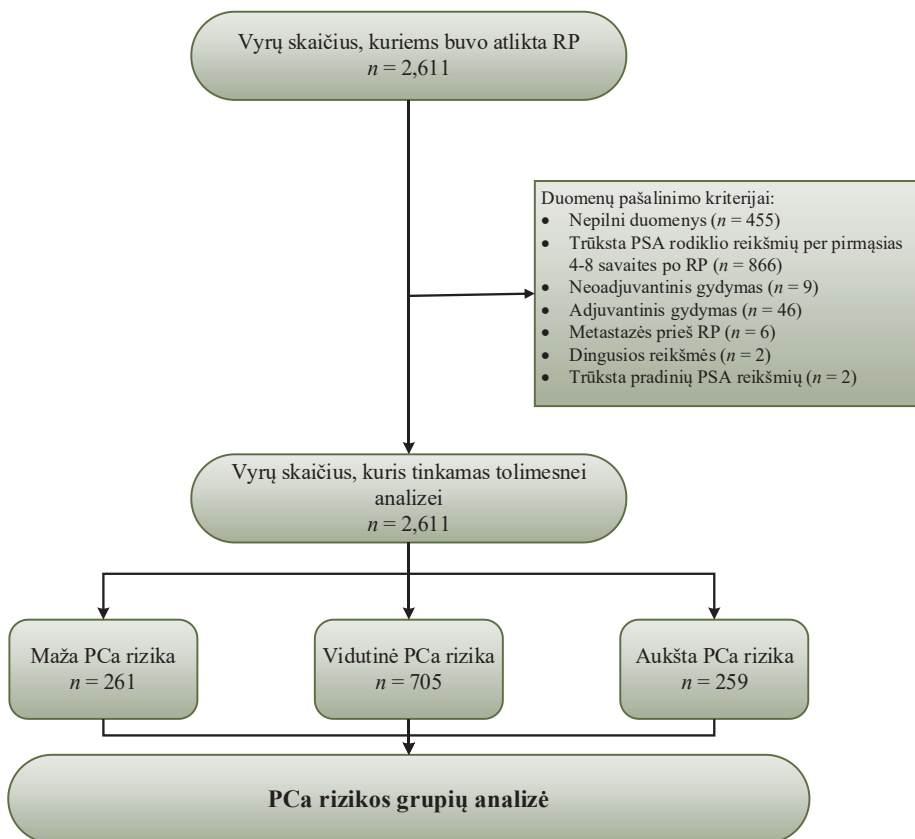
imties dydžiui daugiamačių kriterijų galia auga. Daugiamačių kriterijų galios priklausomybė nuo dimensijos kiekvieno skirstinio atveju skiriasi.

3.3. Praktinis pritaikymas realiems duomenims

Šiame skyriuje pateikiamas daugiamačių suderinamumo hipotezės tikrinimo kriterijų AD, CHI2, CVM, DH, DN, Energijos, HZ, LV, N2, Mar1, Mar2 ir NRR praktinis pritaikymas realiems duomenims. Šiam tyrimui buvo pasirinkta keletas duomenų rinkinių. Kriterijų galios apskaičiavimas turint realius duomenis buvo vykdomas pagal pirmuosius du būdus (t. y. kai kriterijaus statistikos pasiskirstymas yra žinomas arba kai nežinomas). Tik neatliekamas gautų rezultatų vidurkinimas (žr. 2.4 sk.).

3.3.1. Prostatos specifinio antigeno rodiklių normalumo prielaidos tikrinimas

Praktiniam taikymui pasirinkti 2001–2019 metais Lietuvos sveikatos mokslų universiteto Urologijos klinikos surinkti duomenys. Ši duomenų rinkinį sudarė 2611 pacientų tyrimų rodikliai: prostatos specifinio antigeno (PSA), biocheminis atkrytis (BCR), mirtingumas nuo vėžio (CSM), bendrasis mirtingumas (OM), radikali prostatektomija (RP), metastazės (MTS) ir kt. BCR įgijo reikšmę „1“, kai atlikus du tyrimus $PSA > 0,2$ ng/mL. MTS įgijo reikšmę „1“, kai kaulų skenavimo, kompiuterinės tomografijos arba magnetinio rezonanso tyrimais (pritaikius solidinių navikų atsako vertinimo kriterijus) aptikti skeleto ar vidaus organų pažeidimai. Išgyvenamumas be MTS buvo apibrėžtas kaip laikas nuo operacijos iki MTS nustatymo. OM buvo apibrėžtas kaip laikas nuo operacijos iki mirties dėl bet kokios priežasties. Gyvi pacientai buvo cenzūruoti. CSM buvo apibrėžtas kaip laikas nuo operacijos iki su vėžiu susijusios mirties. Gyvi pacientai taip pat cenzūruoti. Pacientų duomenys užregistruoti Prostatos vėžio (PCa) duomenų bazėje [113]. Duomenų rinkinio struktūra pateikta 3.6 paveiksle.



3.6 pav. Prostatos specifinio antigeno rodiklių duomenų rinkinio struktūra

Normalumo prielaida buvo tikrinta filtruojant pacientų amžių ir PSA pagal mirimo metus (t. y. mirę per pirmus 1, 2, 3, 4, 5, 6, 7, 10 ir 15 metų). Nufiltruoti duomenys buvo standartizuoti. Kriterijams buvo skaičiuojama galia ir p -reikšmė. Tyrimui taikyta $\alpha = 0,05$. Pagal gautus rezultatus nustatyta, kad visi taikyti daugiamačiai kriterijai atmeta H_0 hipotezę apie duomenų pasiskirstymą pagal normalųjį skirstinį. Kriterijų CHI2, DH, Energijos, HZ, LV, N2, Mar, NRR ir Roy galia 0,999, o p -reikšmė $<0,0001$. Išsiskiria DN kriterijus, kurio galia 0,576, o p -reikšmė 0,026.

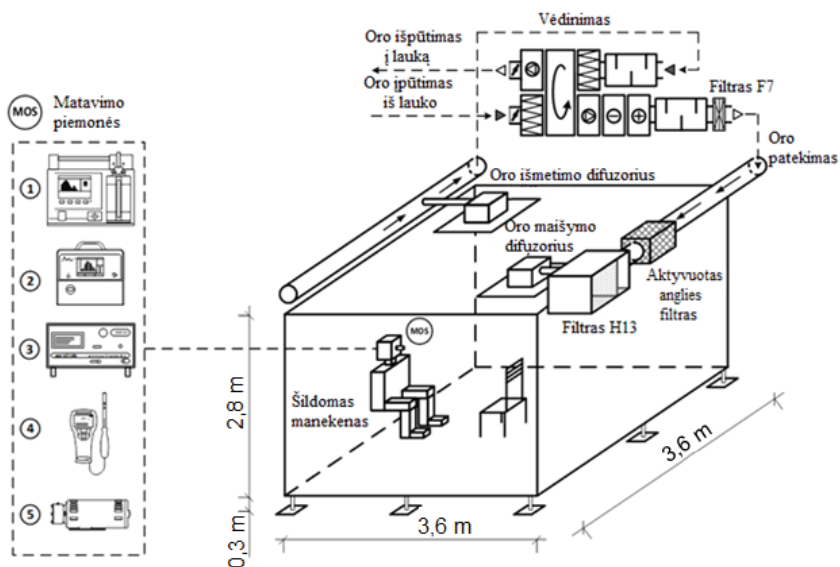
Autoriai [113] netikrino normalumo prielaidos ir taikė su ja nesusijusius išgyvenamumo analizės metodus. Susipažinus su duomenimis galima patikrinti, ar PSA įtakos turi ir kiti stebėti rodikliai (paciento amžius, gydymo būdas, pašalintų limfmazgių skaičius, patologinė stadija, rizikingumo lygis ir kt.). Pagal normalumo prielaidos patikrinimo rezultatus autorių tyrimas galėtų būti praplėstas neparametrinės dispersinės analizės metodais (pvz., Kruskalo ir Voliso neparametrinis metodas), nustatant faktorių įtaką PSA ir su juo glaudžiai susijusiai mirtingumo rizikai. Kai duomenys nėra pasiskirstę pagal normalųjį skirstinį ir didžioji duomenų dalis yra pradiniuose laiko perioduose, tada galima pritaikyti

Kaplano ir Mejerio išgyvenamumo analizės metodą, siekiant nustatyti laiko trukmę nuo apibrėžtos pradžios (pvz., gydymo pradžia, diagnozės laikas).

3.3.2. IQOS ir tradicinių cigarečių išmetamų kietųjų dalelių koncentracijos normalumo prielaidos tikrinimas

Daugiamačių suderinamumo hipotezės tikrinimo kriterijų taikymui pasirinkta 2017 metų Kauno technologijos universiteto, Cheminės technologijos fakulteto, Aplinkosaugos technologijos katedros mokslininkų atliktų 33 eksperimentų (su skirtingomis sąlygomis) duomenys. Eksperimentai buvo atlikti 13 m² ploto ir 35,8 m³ talpos patalpoje. Jos schema pateikiama 3.7 paveiksle. Ši patalpa buvo įrengta panaudojant įprastas konstrukcines medžiagas. Eksperimentų patalpoje buvo įrengti oro įpūtimo ir išpūtimo difuzoriai bei reguliuojama temperatūra, kuri eksperimentų metu buvo $+22 \pm 2$ °C. Eksperimentai buvo atliekami su manekenu. Manekeno paviršiaus plotas 1,7 m². Jis buvo uždengtas tekstiline medžiaga, o jo paviršiaus temperatūra palaikoma nuo +31 °C iki +34 °C. Aerosolių mėginiai buvo ištraukti per koperio įleidimo vamzdelį ir izokinetiniu būdu suskirstyti į mėginių ėmimo prietaisus. Šie prietaisai buvo išdėstyti iškart už patalpos. Manekenas nuo sienos buvo atitrauktas 0,2 m atstumu (žr. 3.8 pav.). Eksperimentų patalpoje buvo keistas vėdinimas (0,2 arba 0,5, arba 1), drėgmė (30 arba 50, arba 70), rūkančiųjų skaičius (1 arba 3, arba 5), atstumas tarp jų (0,5 m arba 1 m, arba 2 m) ir rūkomų cigarečių tipas (IQOS¹¹ arba tradicinė cigaretė). Pirmuose 30 eksperimentų buvo naudotas IQOS, o kituose – tradicinė cigaretė. Savanoriai vienu metu rūkė cigaretes. Eksperimentų metu buvo imituojamas pasyvus rūkymas uždaroje patalpoje ir matuojama išmetamų kietųjų dalelių skaitinė (Pn10) ir masės (Pm2,5 ir Pm10) koncentracijos, dalelių koncentracijos sumažėjimo/padidėjimo rodikliai, lakiųjų organinių junginių (VOC) koncentracija kas sekundę, neuroninių karbonilų (acetaldehido ir formaldehido) kiekis, nikotino kiekis, išpučiamo etilbenzeno (BTEX) kiekis, CO ir CO₂ kiekiai [114].

¹¹ IQOS (kaitinančioji cigaretė) – rūkymui skirtas produktas, susidedantis iš kaitinamo prietaiso ir tabako lazdelės.



3.7 pav. Eksperimentų patalpos schema

Normalumo prielaida tikrinta filtruojant Pn10, Pm2,5, Pm10 pagal eksperimento numerį rūkymo fazėje. Nufiltruoti duomenys standartizuoti. Tyrimo metu apskaičiuojama daugiamačių kriterijų galia ir p -reikšmė su $\alpha = 0,05$. Pagal gautus rezultatus nustatyta, kad visi taikyti daugiamačiai kriterijai parodo, jog normalumo prielaida yra atmetama. Daugumos taikytų daugiamačių kriterijų (CHI2, DH, Energijos, HZ, LV, N2, Mar, NRR ir Roy) galia 0,999, o p -reikšmė $<0,0001$. Kitų kriterijų galia taip pat artima 0,99 ir p -reikšmė yra apie 0,0001.

Autoriai [114] netikrino suderinamumo hipotezės apie duomenų pasiskirstymą pagal normalųjį skirstinį. Mokslininkai analizavo tyrimą cheminiu požiūriu, taip pat pritaikė žvalgomosios ir grafinės analizės metodus, kuriems nėra būtina tikrinti normalumo prielaidos. Remiantis tuo, kad šių duomenų normalumo prielaida yra atmetama, galima pritaikyti Spirmano koreliacijos koeficientą ir apskaičiuoti ryšio stiprumą tarp Pn10, Pm2,5, Pm10 ir tarp ventilacijos, atstumo, rūkančiųjų skaičiaus ir drėgmės rūkymo ir nerūkymo fazėse. Taikant Kruskalo ir Voliso neparametrinės dispersinės analizės metodą galima nustatyti, ar cigaretės tipas turi įtakos Pn10, Pm2,5, Pm10 koncentracijų vidurkiams. Tyrimą galima dar praplėsti branduoliniu regresijos metodu su Gauso branduoliu siekiant nustatyti kintamųjų ir jų porų sąveikų svarbumą Pn10, Pm2,5, Pm10 koncentracijoms rūkymo ir nerūkymo fazėse.

3.4. Darbo rezultatus apibendrinančios išvados

Atlikus vienmačių ir daugiamačių suderinamumo hipotezės tikrinimo kriterijų modeliavimo tyrimus galima padaryti išvadas.

1. Atliktas išsamus pasiūlyto ir populiarių vienmačių suderinamumo hipotezės tikrinimo kriterijų galios palyginimas simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms. Pagal gautus modeliavimo rezultatus nustatyta, kad simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms galingiausi kriterijai buvo Hosking1 (mažiems imčių dydžiams) ir naujai pristatytas (dideliems imčių dydžiams) kriterijus. Hosking1 kriterijaus galia (mažiems imčių dydžiams) 1,5–7,99 proc. didesnė už antro (pagal galią) kriterijaus simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms. Naujo kriterijaus galia (dideliems imčių dydžiams) 6,2–16,26 proc. didesnė už antro (pagal galią) kriterijaus simetrinių, asimetrinių ir modifikuotų normaliųjų skirstinių grupėms.
2. Atliktas daugiamačių suderinamumo hipotezės tikrinimo kriterijų galios palyginimas simetrinių, asimetrinių, mišriųjų ir normaliojo skirstinio mišinių grupėms. Pagal gautus modeliavimo rezultatus nustatyta, kad naujas daugiamatis kriterijus buvo galingiausias simetrinių ir mišriųjų skirstinių grupėms. Asimetrinių ir normaliojo skirstinio mišinių grupėms galingiausi buvo Roystono ($d = 10$ su $n = 256, 512, 1024$) ir naujas (kai $d = 2, 5$) kriterijai. Naujo kriterijaus galia (kai $d = 2$) 4,9–29,6 %, 0–12,2 %, 9,8–41,7 %, 0–39,2 % didesnė už antro (pagal galią) kriterijaus simetrinių, asimetrinių, mišriųjų ir normaliojo skirstinio mišinių grupėms.
3. Daugiamačiai suderinamumo hipotezės tikrinimo kriterijai (įskaitant ir naujai pasiūlytą) buvo pritaikyti realiems duomenų rinkiniams. Pagal gautus rezultatus nustatyta, kad chi-kvadrato, Doorniko ir Hanseno, Energijos, Henzės ir Zirklerio, Lobato ir Velasco, Nikulino, Rao ir Robsono, Roystono ir naujai pristatytas kriterijai galingai (galia = 0,999, o p -reikšmė yra $<0,0001$) suveikia realių duomenų atvejais.

IV. BENDROSIOS IŠVADOS IR REKOMENDACIJOS

1. Ištirtos pasiskirstytųjų skaičiavimų, imties dydžio, pasiskirstymų ir kriterijų parinkimo problemos atliekant suderinamumo su normaliuoju skirstiniu hipotezių tikrinimo kriterijų galios analizę. Pagal atliktą apžvalgą tyrimui atrinktos įvairaus tūrio imtys, skirtingi duomenų pasiskirstymai – atspindintys visas dažniausiai praktikoje sutinkamas problemas, galingiausi suderinamumo hipotezės tikrinimo kriterijai.
2. Remiantis N -metrikos teorija, pasiūlytas suderinamumo hipotezės tikrinimo kriterijus, kuris yra galingas kitų populiarių vienmačių kriterijų konkurentas. Darbe parodyta, kad pasiūlytas kriterijus galingiausias naudoti simetriniams duomenų rinkiniams, kurių imties dydis $n \geq 112$, asimetriniams duomenų rinkiniams, kurių imties dydis $n \geq 118$, ir modifikuotų normaliųjų skirstinių duomenų rinkiniams, kurių imties dydis yra $n \geq 88$.
3. Atliktas naujos branduolio funkcijos parametrų kalibravimas leido parinkti tokią jos formą, kuri vertinimo poslinkį, taikant pasiūlytą N -metrika paremtą suderinamumo hipotezės tikrinimo kriterijų, sumažino iki artimo optimaliam.
4. Sukonstruotas pasiskirstymo tankių skirtumo vertinimu ir apvertimo formulės taikymu paremtas daugiamačis suderinamumo hipotezės tikrinimo kriterijus. Daugiamačių suderinamumo hipotezės tikrinimo kriterijų palyginimas parodė, kad sukonstruotą kriterijų pakankamo dydžio imtims geriausia naudoti 2-mačiams duomenims. Didesnės dimensijos duomenims šį kriterijų rekomenduojama naudoti su dideliais imties dydžiais ($n \geq 512$).
5. Parodyta, kad daugiamačiai suderinamumo hipotezės tikrinimo kriterijai – chi-kvadrato, Doorniko ir Hanseno, Energijos, Henzės ir Zirklerio, Lobato ir Velasco, Nikulino, Rao ir Robsono, Roystono bei naujai pristatytas yra galingiausi šiame tyrime naudotų realių duomenų pasiskirstymų atvejais.

VI. SUMMARY

INTRODUCTION

Relevance of the work

In modern data analytics, making more complex decisions is not possible without testing hypotheses. Data analysts try to use statistical and a priori information. They usually start their research by testing hypotheses about the distribution of data. The information about data distribution can be useful in several of ways, for example:

- It can provide insights about the observed process.
- Parameters of the model could be inferred from the characteristics of data distributions.
- It can help to choose more specific and computationally efficient methods.

For these reasons, this study examines the issue of testing goodness of fit hypotheses. The goodness of fit hypotheses for normality is defined as

H_0 : Data distribution is normal.

H_A : Data distribution is not normal.

Testing of the goodness of fit hypotheses about the normal distribution of data is applied (as an assumption) to many methods of parametric data analysis (e.g., correlation, variance, regression analysis, etc.). The assumption of normality is one of the most important for many practical applications. One way to check the assumption of normality is to graphically (i.e., draw a histogram and estimate the distribution). However, it is not a reliable due to the high potential for error. A more accurate way to test the assumption of normality is to apply the goodness of fit hypothesis tests. Since 1900, many univariate and multivariate tests have been developed to test the assumption of normality. In 1900, Karl Pearson introduced the first test for testing the goodness of fit hypothesis, which he called the Chi-square [1]. This test is considered as a basis of modern statistics. New tests were developed later. Presently, some of the more popular tests include the Kolmogorov-Smirnov [2], Anderson-Darling [3], Cramer von Mises [4], and their modifications. Tests are based on various measures or metrics that result in different responses to data exceptions. Therefore, the choice of the appropriate goodness of fit hypothesis test remains important.

The object, subject, and measure of the research

The *object* of this research is considered to be the goodness of fit hypothesis for testing the assumption of normality. The *subject* is univariate and multivariate test for testing the goodness of fit hypothesis. A *measure* of power is used to compare the tests of the goodness of fit hypothesis.

The aim of the work

Create and examine univariate and multivariate tests for testing goodness of fit hypotheses that would be effective under the normality assumption.

The main objectives of this research are as follows:

1. To examine the main problems, solutions, and obstacles for testing the goodness of fit hypothesis.
2. To create new univariate and multivariate tests for testing the assumption of normality.
3. To perform a comparative analysis of the power of created and popular univariate and multivariate goodness of fit hypothesis test in the case of data distributions used in research.
4. To apply the created goodness of fit hypothesis tests to real data sets.

Methodology of the research

The dissertation uses various methods of probability theory, mathematical statistics, parallel and distributed computations and multivariate data visualisation methods. A univariate goodness of fit test based on N-metric is proposed. The presented multivariate test is based on the inversion formula. The empirical part presents the results obtained by the Monte Carlo modeling method. SAS and R statistical software were used in this work.

For the defense:

- Based on the theory of N-metric, a test for normality is proposed, which is more powerful than the other powerful univariate tests for large sample sizes.
- A multivariate test based on the evaluation of the difference in distribution densities and the application of the inversion formula is constructed, which has significantly higher power than the other powerful multivariate tests for a group of symmetric and mixed distributions.

Scientific novelty and practical significance of the research

The proposed test for testing the goodness of fit hypothesis extends, summarise and refines the results of Bakšhaev [5, 6], Klebanov [7], Kavaliauskas [8], and Ruzgas [9]. The proposed univariate and multivariate tests are powerful competitors to other published tests [P1, P2]. The proposed univariate test is the most appropriate for symmetric data sets of $n \geq 112$ size, for asymmetric data sets of $n \geq 118$ size, and bell-shaped distributed data sets of $n \geq 88$ size. The constructed multivariate test for moderate-sized samples is best used for 2-dimensional data. Using this test with large sample sizes ($n \geq 512$) for larger dimensions is recommended. The statistical tests presented in this work can be successfully applied in real data analysis goodness of fit hypothesis testing tasks. The results and recommendations of this research were used by Hernandez [10], Szentesi et al. [11], Sergei et al. [12], Avdović et al. [120], Ahmad et al. [121], Doz et al. [122] in their studies.

Approval of the results

There are three scientific articles published on the topic of the dissertation in the list of the Institute for Science Information (ISI) as the main list of publications with citing indexes, and one article is published in a peer-reviewed scientific journal (without ISI). One publication was published in a national and three in international

publishers. The topics of this dissertation were presented in five international and one national conferences. The doctoral dissertation was presented:

- January 14, 2022, at the meeting of Kaunas University of Technology, Faculty of Mathematics and Natural Sciences, Department of Applied Mathematics.
- February 15, 2022, Vilnius Gediminas Technical University, Faculty of Fundamental Sciences, Department of Mathematical Modeling.
- June 2, 2022 at the meeting of Composition of the Doctoral Committee of Informatics.

Scope and structure of the dissertation

This doctoral dissertation consists of an introduction, three main sections, conclusions, references, and a list of publications. The volume of the work is 150 pages. There are 23 figures, 11 tables, and a list of 122 cited references.

I. REVIEW OF SCIENTIFIC LITERATURE

The object of the research presented in this section is the goodness of fit hypothesis. Comparative studies of the goodness of fit hypothesis tests for the assumption of normality are discussed.

1.1. Goodness-of-fit hypothesis and a comparative review of tests

The goodness-of-fit hypothesis is a non-parametric hypothesis about the observed random X distribution function:

$$H_0: F_X \equiv F_0(X).$$

$$H_A: F_X \not\equiv F_0(X).$$

In 1900, this hypothesis was tested when the first Chi-square test was developed. However, the term was proposed by Lazarus and Folkman in 1984 [13] and developed by Forsythes and Compass only in 1987 [14].

The goodness-of-fit hypothesis is most commonly used to test the assumption of compatibility of data with normal distribution. Testing of this assumption is important in the selection of data analysis methods. For example, in an analysis of variance (using Fisher generalized linear model or Kruskal-Wallis nonparametric method) [15, 16] or regression analysis (using least squares linear model or nonparametric spline or kernel regression method) [17, 18]. The identification of the distribution of data is also important in neural networks both in the selection of activation functions and in the construction of the neural network [19, 20].

The tests for the goodness-of-fit hypotheses are divided into univariate and multivariate. First, univariate tests were created and analyzed. In 1900, the examination of the assumption about the distribution of data was started with the Chi-square test [1]. This test is still widely used. As technology advances, researchers created, specified, and modified univariate goodness-of-fit hypothesis tests [10, 21-34]. Since 1981, the first studies with multivariate goodness-of-fit hypothesis tests were begun. In 1981, Moore and Stubblebine introduced the first multivariate Chi-square test [37]. Since then, research into the creation and comparison of multivariate tests begun [37-46].

Scientific publications examining the issue of the goodness-of-fit tests can be divided into three parts:

1. Theoretical derivations and evidence for the goodness-of-fit hypothesis tests [5, 6, 25, 37, 39, 40, 47-49].
2. The goodness-of-fit hypothesis tests using simulation modelling research [10, 21-34, 37-46, 50-52].
3. Use of goodness-of-fit hypothesis tests in solving real world problems [15-20, 53, P4].

1.2. Review and analysis of univariate tests research

At the beginning of the 20th century, the creation and analysis of univariate tests for testing the goodness-of-fit hypothesis began [1]. Since then, new univariate tests for goodness-of-fit hypotheses are still created and compared with others. From a practical point of view, the test of the assumption of normality remains the most relevant. Univariate tests are often used for data analysis. In the 21st century, research on this topic is very active [5, 6, 10, 21-34, 47-51]. A small part of the research on this topic was based on mathematical evidence, while the majority focused on imitative modelling. The researchers compared a large number of classical goodness-of-fit hypothesis tests (up to 50) in their studies. The tests were applied to various sample sizes (from 10 to 2,000) in the studies using a significance level of α (usually $\alpha = 0.05$). A modelling study was performed for many data on symmetric, asymmetric, and modified normal distributions and mixtures. The tests were compared using a power measure.

Researches of the univariate goodness-of-fit hypothesis tests are still relevant. In the 21st century, research was focused on tests based on N-metric, empirical distribution and density functions, kurtosis and skewness, Wasserstein metric, and others. The tests were compared based on mathematical evidence [5-7, 25, 47, 48] and the Monte Carlo modelling method [10, 21, 22, 24-26, 28-30, 32-34, 50, 51, 57-59, 60-64].

1.3. Review and analysis of multivariate tests research

At the end of the 20th century, the creation and analysis of multivariate goodness-of-fit hypothesis tests began [37]. Multivariate tests are less commonly used in practice due to their complexity. In the 21st century, research on this topic was carried out [37-46, 52, 65-71]. The researchers created and compared multivariate tests using the Monte Carlo modelling method. Multivariate tests were compared for different symmetric, asymmetric distributions and mixtures for different sample sizes (10–1000) when the dimension was 2–10. The significance level $\alpha = 0.05$ was usually chosen.

The most powerful multivariate tests are based on kurtosis and skewness, undeniable functional distance, empirical distribution function, density, Euclidean distance, and sample characteristics. In addition, the proposed various data transformations help increase the power of the tests.

II. METHODOLOGY

This section defines the goodness-of-fit tests. These tests are grouped according construction. The proposed univariate and multivariate tests are described. The data distributions used in the study and the procedure for calculating the power are presented

2.1. Univariate tests for testing the goodness-of-fit hypothesis

Let X_1, X_2, \dots, X_n denotes a random sample of n , $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denotes a sample of order statistics, \bar{X} and $\hat{\sigma}^2$ denote the mean and variance of the sample, which are defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (2.1)$$

The standardized mean and dispersion size are defined as

$$Z_i = \frac{X_i - \bar{X}}{\hat{\sigma}}. \quad (2.2)$$

The kurtosis and skewness are defined as

$$s = \frac{m_3}{\sqrt{m_2}^3}, \quad k = \frac{m_4}{m_2^2}, \quad (2.3)$$

where $m_i = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^i$ is i^{th} moment.

Univariate tests are divided into four groups according to their construction:

1. Tests based on empirical distribution function:
 - Kolmogorov-Smirnov (KS) [2];
 - Anderson-Darling (AD) [3];
 - Cramer von Mises (CVM) [4];
 - Lilliefors (LF) [80];
 - Glen-Leemis-Barr (GLB) [81];
 - Zhang-Wu (ZW1 and ZW2) [34].
2. Tests based on order statistic:
 - Shapiro-Wilk (SW) [36];
 - D'Agostino (DA) [84];
 - Shapiro-Francia (SF) [85];
 - Filliben (Filli) [83];
 - Cheno-Shapiro (ChenS) [86];
 - Modified Shapiro-Wilk (SWRG) [87];
 - Zhang (ZQ, ZQstar, ZQQstar) [88];
 - Barrio-Cuestos-Albertos-Matrano-Rodríguez (BCMR) [90];
 - Coin (Coin) [91];
3. Tests based on skewness and kurtosis:

- D'Agostin-Pearson (DAP) [92];
 - Jarque-Bera (JB) [93];
 - Hosking-Wall (H1-H4) [94, 116];
 - Cabaña-Cabaña (CC1, CC2) [95, 96];
 - Bonett-Seier (BS) [97];
 - Bontemps-Meddah (BM1, BM2) [98];
 - Doornik-Hansen (DH) [38];
 - Robust Jarque-Bera (GG) [99];
 - Brys-Hubert-Struyf (BHS) [24];
 - Brys-Hubert-Struyf-Bonett-Seier (BHSBS) [24];
 - Desgagné-Lafaye de Micheaux-Leblanc (DLDMLRn) [25];
 - Desgagné-Lafaye de Micheaux-Leblanc (DLDMXAPD) [25];
 - Desgagné-Lafaye de Micheaux-Leblanc (DLDMZEPD) [25];
4. Other tests:
- Chi-Square (CHI2) [1];
 - Martinez-Iglewicz (MI) [100];
 - Epps-Pulley [101];
 - Gel-Miao-Gastwirth (GMG) [99];
 - Bakshaev(Baksh) [5, 6].
 - A new univariate test (N1) [**P1**]

A new goodness-of-fit hypothesis test (N1) based on N-metric with new kernel function was constructed. This test is defined as follow under the null hypothesis statistic

$$T_n = -n \int_0^1 \int_0^1 K(x) d(F_n^*(x) - x) \quad (2.4)$$

has the same asymptotic distribution as the quadratic form:

$$T_n = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{a_{kj}}{\pi^2 kj} \xi_k \xi_j, \quad (2.5)$$

where ξ_k are independent random variables from the standard normal distribution and

$$a_{kj} = -2 \int_0^1 \int_0^1 K(x) d \sin(\pi kx), \quad (2.6)$$

A new kernel function (Figure 2.1) was applied to statistics (2.4)

$$K(x) = \varphi(\bar{g}(x)) \bar{g}'(x), \quad (2.7)$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (2.8)$$

An additional bias is introduced when the kernel function is calculated at the sample values, i.e., for $x=X(t)$. Therefore, to eliminate this bias, the shape of the kernel function is chosen so that their influence in the environment of the sample

values is as small as possible. Let X be the standard normal random variable, Φ and φ be its distribution and density functions, respectively, $g: \mathbf{R} \rightarrow \mathbf{R}$ is an odd strictly monotonically increasing function. Then the distribution function F_Y of the random variable $Y = g(X)$ is $\Phi(\bar{g}(x))$, where \bar{g} is the inverse of the function g . The distribution density f_Y of a random variable Y is $\varphi(\bar{g}(x))\bar{g}'(x)$. Let us consider the parametric class of functions \bar{g} , which depends on three parameters:

$$\begin{aligned} \bar{g} &= x(c + |x|^b)^a, \quad a, b, c > 0, \\ \bar{g}' &= (c + |x|^b)^a + a|x|(c + |x|^b)^{a-1}b|x|^{b-1}. \end{aligned} \tag{2.9}$$

where a is variance, b is the trough, c is the peak shape parameter.

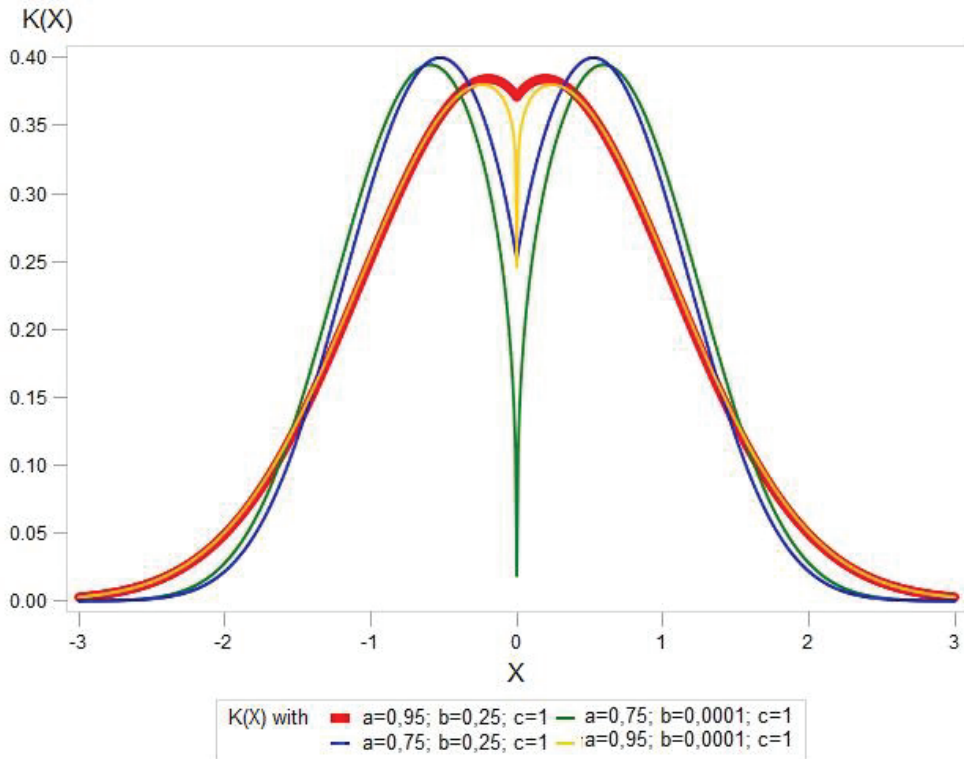


Figure 2.1. Plot of out kernel function $K(x)$ with experimentally chosen optimal parameters $a = 0.95$, $b = 0.25$, $c = 1$.

2.2. Multivariate tests for testing the goodness-of-fit hypothesis

Let denote the d -variate normal distribution as $N_d(\mu, \Sigma)$, where μ is an expectation vector $\mu = (\mu_1, \dots, \mu_d)^T$ and Σ is the nonsingular covariance matrix. N_d indicates a set of all possible d -variate normal distributions. Let X_1, X_2, \dots, X_n , where $X_k = (X_{k1}, X_{k2}, \dots, X_{kd})^T$. The covariance matrix is defined as

$$\Sigma = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})^T.$$

Multivariate tests are divided into three groups according to their construction:

1. Tests based on squared radii:
 - Chi-Square (CHI2) [37];
 - Cramer von Mises (CVM) [40];
 - Anderson-Darling (AD) [40];
 - Nikulin- Rao- Robson (NRR) [37];
 - Dzhaparidze and Nikulin (DN) [37].
2. Tests based on skewness and kurtosis:
 - Mardia (Mar1, Mar2) [75];
 - Royston (Roy) [45].
3. Other tests:
 - Henze-Zirkler (HZ) [39];
 - Lobato-Velasco (LV) [103];
 - Energy (Energy) [104];
 - A new multivariate test (N2) [**P2**].

New goodness-of-fit hypothesis test (N2) based on distribution distance has been constructed using an inversion formula. The estimation of a sample distribution density is based on the application of the characteristic function and inversion formula. This method is known for its good properties (i.e., low sensitivity) and was introduced in [9]. Marron and Wand [105] extensively compared density estimation methods (including the adapted kernel method). They concluded that density estimation based on the application of characteristic function and inversion is more accurate for non-Gaussian data sets.

The random d -variate vector $X \in \mathbb{R}^d$, which follows a distribution of a mixture model has a density function

$$f(X) = f(X, \theta) = \sum_{k=1}^q p_k f_k(X, \theta_k), \quad (2.10)$$

where q is the number of clusters (i.e., components, classes) of the mixture, and p_k ($k = 1, \dots, q$) is the a priori probability that satisfies

$$p_k > 0, \sum_{k=1}^q p_k = 1. \quad (2.11)$$

The $f_k(X, \theta_k)$ is a distribution of the k^{th} class and θ is a set of parameters $\theta = \{p_1, \dots, p_q, \theta_1, \dots, \theta_q\}$. Let denote the d -variate sample of independent and identically distributed random values X .

When examining approximations of parametric methods, it should be emphasized that as the data dimension increases, the number of model parameters increases rapidly, making it more difficult to find accurate parameter estimates. It is much easier to find the density of univariate data projections

$$x_\tau = \tau^T x, \quad (2.12)$$

than multivariate data density f because of mutually unambiguous compliance

$$f \leftrightarrow \{f_\tau, \tau \in \mathbb{R}^d\}. \quad (2.13)$$

It is natural to try to find the multivariate density f using the density estimates \hat{f}_τ of univariate observational projections [9]. In the case of Gaussian mixture model, the projection of the observations (2.12) is also distributed according to the Gaussian mixture model:

$$f_\tau(x) = f_\tau(x, \theta_\tau) = \sum_{k=1}^q p_{k,\tau} \varphi_{k,\tau}(x), \quad (2.14)$$

where $\varphi_{k,\tau}(x) = \varphi(x; m_{k,\tau}, \sigma_{k,\tau}^2)$ is univariate Gaussian density. The parameter set θ of the multivariate mixture and the distribution parameters of the data projections $\theta_\tau = (p_{k,\tau}, m_{k,\tau}, \sigma_{k,\tau}^2)$, $k = 1, \dots, q$ are related by equations:

$$\begin{aligned} p_{j,\tau} &= p_j, \\ m_{j,\tau} &= \tau^T M_j, \\ \sigma_{j,\tau}^2 &= \tau^T R_j \tau. \end{aligned} \quad (2.15)$$

The inversion formula is used

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-it^T x} \psi(t) dt, \quad (2.16)$$

where

$$\psi(t) = E e^{it^T x} \quad (2.17)$$

where $\psi(t)$ denotes the characteristic function of the random variable X . Given that $u = |t|$, $\tau = t/|t|$ and by changing the variables to a spherical coordinate system we obtain

$$f(x) = \frac{1}{(2\pi)^d} \int_{\tau: |\tau|=1} ds \int_0^\infty e^{-iu\tau^T x} \psi(u\tau) u^{d-1} du, \quad (2.18)$$

where the first integral is the surface integral of the unit sphere. The characteristic function of the projection of the observed random variable is

$$\psi_\tau(u) = E e^{iu\tau^T x}, \quad (2.19)$$

and has the property

$$\psi(u\tau) = \psi_\tau(u). \quad (2.20)$$

The design directions were selected as follows: a uniformly generated hyper-cube with an edge of 2. It is checked whether the generated point for direction indication falls within the hypersphere of a single radius. If the point was outside the sphere, then regeneration was performed. If the point was inside the sphere, then it defined the direction. By selecting the set T of uniform distributed directions ($\tau = 10,000$) on the sphere and replacing the characteristic function with its estimate, a density estimate is obtained [8,9]:

$$\hat{f}(x) = \frac{A(d)}{\#T} \sum_{\tau \in T} \int_0^\infty e^{-iu\tau^T x} \hat{\psi}_\tau(u) u^{d-1} e^{-hu^2} du, \quad (2.21)$$

where $\#T$ denotes the size of set T . Using the d -variate ball volume formula

$$V_d(\mathbb{R}) = \frac{\pi^{\frac{d}{2}} R^d}{\Gamma(\frac{d}{2}+1)} = \begin{cases} \frac{\pi^{\frac{d}{2}} R^d}{(\frac{d}{2})!}, & \text{kai } d \bmod 2 \equiv 0, \\ \frac{2^{\frac{d+1}{2}} \pi^{\frac{d-1}{2}} R^d}{d!!}, & \text{kai } d \bmod 2 \equiv 1, \end{cases} \quad (2.22)$$

the constant $A(d)$ is defined as

$$A(d) = \frac{(V_d(1))'}{(2\pi)^d} = \frac{d2^{-d}\pi^{-\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}. \quad (2.23)$$

Computer simulation studies have shown that the density estimates of the inversion formula are not smooth. Therefore, in formula (2.21), an additional multiplier e^{-hu^2} is used. This multiplier smoothes the estimate $\hat{f}(x)$ with the Gaussian kernel function. Moreover, this form of the multiplier allows the integral value to be calculated analytically. Monte Carlo studies have shown that its use significantly reduces the error of estimates. Formula (2.21) can be used to estimate the characteristic function of the projected data. Let us consider two approaches. The first one is based on the density approximation of the Gaussian distribution mixture model. In this case, the parametric estimate of the characteristic function is used:

$$\hat{\psi}_\tau(u) = \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} e^{iu\hat{m}_{k,\tau} - u^2 \hat{\sigma}_{k,\tau}^2 / 2}. \quad (2.24)$$

By substituting $\hat{\psi}_\tau(u)$ in (2.21) by (2.24), we get

$$\begin{aligned} \hat{f}(x) &= \frac{A(d)}{\#T} \sum_{\tau \in T} \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} \int_0^\infty e^{iu(\hat{m}_{k,\tau} - \tau^T x) - u^2(h + \hat{\sigma}_{k,\tau}^2 / 2)} u^{d-1} du \\ &= \frac{A(d)}{\#T} \sum_{\tau \in T} \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} I_{d-1} \left(\frac{\hat{m}_{k,\tau} - \tau^T x}{\sqrt{\hat{\sigma}_{k,\tau}^2 + 2h}} \right) \left(\sqrt{\hat{\sigma}_{k,\tau}^2 + 2h} \right)^{-d}, \end{aligned} \quad (2.25)$$

where

$$I_j(y) = \operatorname{Re} \left[\int_0^\infty e^{iyt - t^2/2} t^j dt \right]. \quad (2.26)$$

Note, that only the real part of the expression is considered here (the sum of the imaginary parts must be equal to zero) in other words, the density estimate $\hat{f}(x)$ can acquire only real values. The chosen form of the smoothing multiplier e^{-hu^2} allows relating the smoothing parameter h with the variances of the projection clusters, i.e., in the calculations, the variances are simply increased by $2h$. Next, the expression (2.26) is evaluated. Let

$$C_j(y) = \int_0^\infty \cos(yt) \cdot e^{-t^2/2} \cdot t^j dt, \quad (2.27)$$

$$S_j(y) = \int_0^\infty \sin(yt) \cdot e^{-t^2/2} \cdot t^j dt, \quad (2.28)$$

then (2.26) can be written as

$$\int_0^\infty e^{-iyt - t^2/2} t^j dt = C_j(y) + iS_j(y). \quad (2.29)$$

By integrating in parts, can get

$$\begin{aligned} C_j(y) &= e^{-\frac{t^2}{2}} t^{j-1} \cos(yt) \Big|_0^\infty + \int_0^\infty e^{-\frac{t^2}{2}} \left((j-1)t^{j-2} \cos(yt) - \right. \\ &\left. yt^{j-1} \sin(yt) \right) dt = \mathbf{1}_{\{j=1\}} + (j-1)C_{j-2}(y) - yS_{j-1}(y), j \geq 1. \end{aligned} \quad (2.30)$$

$S_j(y)$ is expressed analogously. Concerning the limitations of the j index, the following recursive equations are obtained:

$$C_j(y) = (j - 1)C_{j-2}(y) - yS_{j-1}(y), j \geq 2, \quad (2.31)$$

$$C_1(y) = 1 - yS_0(y), \quad (2.32)$$

$$S_j(y) = (j - 1)S_{j-2}(y) - yC_{j-1}(y), j \geq 2, \quad (2.33)$$

$$S_1(y) = YC_0(y). \quad (2.34)$$

The initial function $S_0(y)$ is founded by starting with the relation

$$(S_0(y))'_y = \int_0^\infty t \cos(yt) \cdot e^{-t^2/2} dt = C_1(y). \quad (2.35)$$

From (2.32) and (2.35) it follows that S_0 satisfies the differential equation

$$S'_0(y) = 1 - yS_0(y), \dots, S_0(0) = 0. \quad (2.36)$$

which is solved by writing down S_0 as the Taylor series:

$$S'_0(y) = \sum_{l=0}^\infty c_{l+1}(l+1)y^{l+1} = 1 - \sum_{l=2}^\infty c_{l-1}y^l. \quad (2.37)$$

By equating the coefficients of the same powers, its values are obtained

$$\begin{aligned} c_0 &= 0, & c_1 &= 1, \\ c_l &= -c_{l-2}/l, & l &\geq 2. \end{aligned} \quad (2.38)$$

which gives us

$$S_0(y) = \sum_{l=0}^\infty \frac{(-1)^l y^{2l+1}}{(2l+1)!!} = y - \frac{y^3}{3!!} + \frac{y^5}{5!!} - \frac{y^7}{7!!} + \dots \quad (2.39)$$

C_0 is found from expression (2.27):

$$\begin{aligned} C_0(y) &= \int_0^\infty \cos(yt) \cdot e^{-t^2/2} dt = \frac{1}{2} \int_{-\infty}^\infty \cos(yt) \cdot e^{-t^2/2} dt \\ &= \frac{1}{2} \int_{-\infty}^\infty (\cos(yt) - i \sin(yt)) \cdot e^{-t^2/2} dt = \sqrt{\frac{\pi}{2}} e^{-y^2/2}. \end{aligned} \quad (2.40)$$

The value of the integral (2.21) then is

$$I_j(y) = C_j(y). \quad (2.41)$$

One of the disadvantages of the inversion formula method (defined by (2.21)) is that the Gaussian distribution mixture model (2.10) described by this estimate (for $f_k = \varphi_k$) does not represent density accurately, except around observations. When approximating the density under study with a mixture of Gaussian distributions, the estimation of density using the inversion formula often becomes complicated due to a large number of components. Thus, we merge the components with small a priori probabilities into one noise cluster.

We have developed and examined a modification of the algorithm which is based on the use of a multivariate Gaussian distribution mixture model. The parametric estimate of the characteristic function of uniform distribution density is defined as

$$\hat{\psi}(u) = \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}. \quad (2.42)$$

in the inversion formula (2.16). In the density estimate calculation formula (2.21), the estimation of the characteristic function is constructed as a union of the characteristic functions of a mixture of Gaussian distributions and uniform distribution with corresponding a priori probabilities:

$$\hat{\psi}_\tau(u) = \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} e^{iu\hat{m}_{k,\tau} - u^2 \hat{\sigma}_{k,\tau}^2 / 2} + \hat{p}_{0,\tau} \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}, \quad (2.43)$$

where the second member describes a uniformly distributed noise cluster, \hat{p}_0 – noise cluster weight, $a = a(\tau)$, $b = b(\tau)$. Based on the established estimates of the parameters of the uniform distribution and data projections, it is possible to define the range

$$a = (\tau^T x)_{\min} - \frac{(\tau^T x)_{\max} - (\tau^T x)_{\min}}{2(n-1)}, \quad (2.44)$$

$$b = (\tau^T x)_{\max} + \frac{(\tau^T x)_{\max} - (\tau^T x)_{\min}}{2(n-1)}. \quad (2.45)$$

By inserting (2.43) into (2.21) we obtain

$$\hat{f}(x) = \frac{A(d)}{\#T} \sum_{\tau \in T} \left[\sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} \int_0^\infty e^{iu(\hat{m}_{k,\tau} - \tau^T x) - u^2(h + \hat{\sigma}_{k,\tau}^2/2)} u^{d-1} du + \frac{2\hat{p}_{0,\tau}}{b-a} \int_0^\infty e^{iu\left(\frac{a+b}{2} - \tau^T x\right) - u^2 h} \cdot \sin\left(\frac{(b-a)u}{2}\right) \cdot u^{d-2} du \right]. \quad (2.46)$$

Using notations such as (2.25), we define the density estimate as

$$\hat{f}(x) = \frac{A(d)}{\#T} \sum_{\tau \in T} \left[\sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} I_{d-1} \left(\frac{\hat{m}_{k,\tau} - \tau^T x}{\sqrt{\hat{\sigma}_{k,\tau}^2 + 2h}} \right) (\hat{\sigma}_{k,\tau}^2 + 2h)^{-\frac{d}{2}} + \frac{2\hat{p}_{0,\tau}}{b-a} J_{d-2} \left(\frac{a+b-2\tau^T x}{2\sqrt{2h}}, \frac{b-a}{2\sqrt{2h}} \right) \cdot (2h)^{-\frac{d-1}{2}} \right], \quad (2.47)$$

where $I_j(y)$ is given in (2.26) which is evaluated by (2.41) and

$$J_j(y, t) = \operatorname{Re} \left[\int_0^\infty e^{iyu - u^2/2} \cdot \sin(tu) \cdot u^j du \right]. \quad (2.48)$$

By integrating, can get

$$\begin{aligned} \int_0^\infty e^{iyu - \frac{u^2}{2}} \cdot \sin(tu) \cdot u^j du &= \int_0^\infty (\cos(yu) + i \sin(yu)) \cdot \sin(tu) \cdot \\ e^{-\frac{u^2}{2}} \cdot u^j du &= \int_0^\infty \left(\frac{\sin(y+t)u + \sin(t-y)u}{2} + i \frac{\cos(y-t)u - \cos(y+t)u}{2} \right) \cdot e^{-\frac{u^2}{2}} \cdot \\ u^j du &= \frac{1}{2} S_j(y+t) + \frac{1}{2} S_j(t-y) + i \frac{1}{2} C_j(y-t) - i \frac{1}{2} C_j(y+t), \end{aligned} \quad (2.49)$$

where $S_j(y)$ and $C_j(y)$ are defined in (2.27) and (2.28). Then the integral (2.48) evaluates to

$$J_j(y, t) = \frac{1}{2} S_j(y+t) + \frac{1}{2} S_j(t-y). \quad (2.50)$$

The above procedure is called *a modified inversion formula density estimate*. Our proposed normality test is based on the distance function

$$\mathcal{T} = n^{-1} \sum_{i=1}^n |f(z_{(i)}) - \hat{f}(z_{(i)})|, \quad (2.51)$$

where z is a standardized value, $\hat{f}(z)$ is an estimate of density function. \mathcal{T} does not depend on a moderate sample volume (≥ 32) but depends on the data dimension. It is convenient to use the test statistics:

$$\mathcal{T}^* = -\log \left(n^{-1} \sum_{i=1}^n |f(z_{(i)}) - \hat{f}(z_{(i)})| \right), \quad (2.52)$$

which had the lowest sensitivity based on the exploratory study. Under the null hypothesis statistic \mathcal{T}^* approximately follows the Johnson S_U distribution which is

specified by the shape ($\delta > 0, \gamma$), scale ($\lambda > 0$), and location (ξ) parameters and has the density function

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}} g' \left(\frac{x - \xi}{\lambda} \right) \exp \left(-0,5 \left[\gamma + \delta g \left(\frac{x - \xi}{\lambda} \right) \right]^2 \right),$$

where $g(y) = \ln \left[y + \sqrt{y^2 + 1} \right]$, when $x \in (-\infty, +\infty)$. Johnson S_U distribution function:

$$F(x) = \int_{-\infty}^x f(u) du.$$

The S_U system is presented in detail in [106, 107]. Estimates of \mathcal{T}^* statistic Johnson S_U distribution parameters for different dimensions are given in Table 2.1.

Table 2.1. Statistic \mathcal{T}^* Johnson S_U distribution parameter estimates.

Parameter	Symbol	Estimate
d=2		
Location	$\hat{\xi}$	4.342807
Scale	$\hat{\lambda}$	0.585038
Shape	$\hat{\delta}$	1.498293
Shape	$\hat{\gamma}$	0.764906
d=5		
Location	$\hat{\xi}$	7.025845
Scale	$\hat{\lambda}$	0.088023
Shape	$\hat{\delta}$	0.895003
Shape	$\hat{\gamma}$	0.400035
d=10		
Location	$\hat{\xi}$	5.195174
Scale	$\hat{\lambda}$	1.578613
Shape	$\hat{\delta}$	2.24856
Shape	$\hat{\gamma}$	-1.83037

For statistic \mathcal{T}^* , the invariance and contingency properties were checked. The invariance property is confirmed because standardized data was used. The contingency property is confirmed experimentally.

2.3. Statistical distributions

2.3.1. Univariate distributions

Statistical distributions are grouped into three groups: symmetric, asymmetric, and modified normal distributions [28, P1]. A description of these distribution groups is presented below.

1. Symmetric distributions considered in this research are

- one case of the standard normal $N(0;1)$ distribution;
- three cases of the $Beta(a; b)$ distribution;

- three cases of the *Cauchy*($t; s$) distribution;
 - one case of the *Laplace*(0;1) distribution;
 - one case of the *Logistic*(2;2) distribution;
 - four cases of the *t - Student*(v) distribution;
 - five cases of the *Tukey*(λ) distribution.
2. Asymmetric distributions considered in this research are
- four cases of the *Beta*($a; b$) distribution;
 - four cases of the *Chi - squared*(v) distribution;
 - six cases of the *Gamma*($a; b$) distribution;
 - one case of the *Gumbel*(1;2) distribution;
 - one case of the *Lognormal*(0;1) distribution;
 - four cases of the *Weibull*($a; b$) distribution.
3. Modified normal distributions considered in this research are
- six cases of the standard normal distribution truncated at a and b *Trunc*($a; b$). This is hereinafter referred to as NORMAL1.
 - nine cases of a location-contaminated standard normal distribution. Hereon termed *LoConN*($p; a$). This is hereinafter referred to as NORMAL2.
 - nine cases of a scale-contaminated standard normal distribution. Hereon termed *ScConN*($p; b$). This is hereinafter referred to as NORMAL3.
 - twelve cases of a mixture of normal distributions. Hereon termed *MixN*(p, a, b) = $N(1 - p, 0, 1) + N(p, a, b)$ This is hereinafter referred to as NORMAL4.

2.3.2. Multivariate distributions

Statistical distributions are grouped into four groups: symmetric, asymmetric, mixed, and normal mixture distributions [42, 46, P2]. A description of these distribution groups is given in the following subsections.

1. Symmetric multivariate distributions are taken from research:
 - one case of the standard normal $N(0;1)$ distribution;
 - three cases of the *Beta*($a; b$) distribution;
 - one case of the *Cauchy*(0;1) distribution;
 - one case of the *Laplace*(0;1) distribution;
 - one case of the *Logistic*(0;1) distribution;
 - four cases of the *t - Student*(v) distribution.
2. Asymmetric multivariate distributions are taken from research
 - five cases of the *Chi - squared*(v) distribution;
 - two cases of the *Gamma*($a; b$) distribution;
 - one case of the *Gumbel*(1;2) distribution;
 - two cases of the *Lognormal* ($t; s$) distribution;
 - three cases of the *Weibull*($a; b$) distribution.
3. Mixed distribution. The generated mixed data distribution

$$X_k = (X_{k1}, X_{k2}, \dots, X_{km}, \dots, X_{kd})^T, k = 1, 2, \dots, n$$

is such that the first m variates (i.e.. $X_{k1}, X_{k2}, \dots, X_{km}$) follow the standard normal distribution and the distribution of the remaining variates is one of the non-normal distributions

$$Laplace(0, 1), \chi^2(5), t(5), Beta(1,1), Beta(1,2), Beta(2,2).$$

The experimental research covers the cases for $m = d - 1, m = d/2$ and $m = 1$.

4. Normal mixture distributions. Normal mixture distributions are considered in this research: nine cases of the multivariate normal mixture distribution $MVNMIX(a, b, c, d)$. The multivariate normal mixture distribution with density

$$aN(0, \Sigma_1) + (1 - a)N(b\mathbf{1}, \Sigma_2),$$

where $\mathbf{1}$ is the column vector with all elements being 1.

$$\Sigma_1 = (1 - c)I + c\mathbf{1}\mathbf{1}^T \text{ and } \Sigma_2 = (1 - d)I + d\mathbf{1}\mathbf{1}^T.$$

2.4. The Power of Test

The power of the test is defined as the probability of rejecting a false H_0 hypothesis. Power is the opposite of type II error. Decreasing the probability of type I error α increases the probability of type II error and decreases the power of the test. The smaller the error is the more powerful the test. In practice, the tests are designed to minimize type II error for a fixed type I error. The most commonly chosen value for α is 0.05. The probability of the opposite event is calculated as $1 - \beta$. i.e., the power of the test β is the probability of rejecting hypothesis H_0 when it is false. The power of the test makes it possible to compare the significance level and sample sizes of two tests. A more powerful test has a higher value of $1 - \beta$. Increasing the sample size usually increases the power of the test [108, 109].

The power of test calculation procedure was performed in two different ways:

1. When the distribution of the test statistics (actual or approximate) is known, then the power calculation scheme in Figure 2.2 applies. Software implementations of goodness-of-fit hypothesis tests (e.g., AD, H1 – H4, etc.) that include power of test calculation functionality are presented in [56]. Figure 2.2 illustrates the relationship between the probabilities of hypothesis testing errors and the power of tests. Based on Figure 2.2. the following steps can be performed to assess the power of the test:

1. The probability α of type I error of the hypothesis test and the corresponding critical value according to the conditional distribution of the test statistics (theoretical or empirical) when the hypothesis H_0 is correct ($f(\gamma|H_0)$).
2. For a specific sample, the value of the test statistic is calculated.
3. The distribution of the test statistic is constructed when the hypothesis H_1 is correct $f(\gamma|H_1)$. Typically, it is assumed that $f(\gamma|H_1)$ is from the same family of distributions as ($f(\gamma|H_0)$) only moved relative to the estimated test statistic value.
4. The probability of type II error β , which is limited by the critical value and $f(\gamma|H_1)$, is estimated.
5. Calculate the power of test $1 - \beta$.

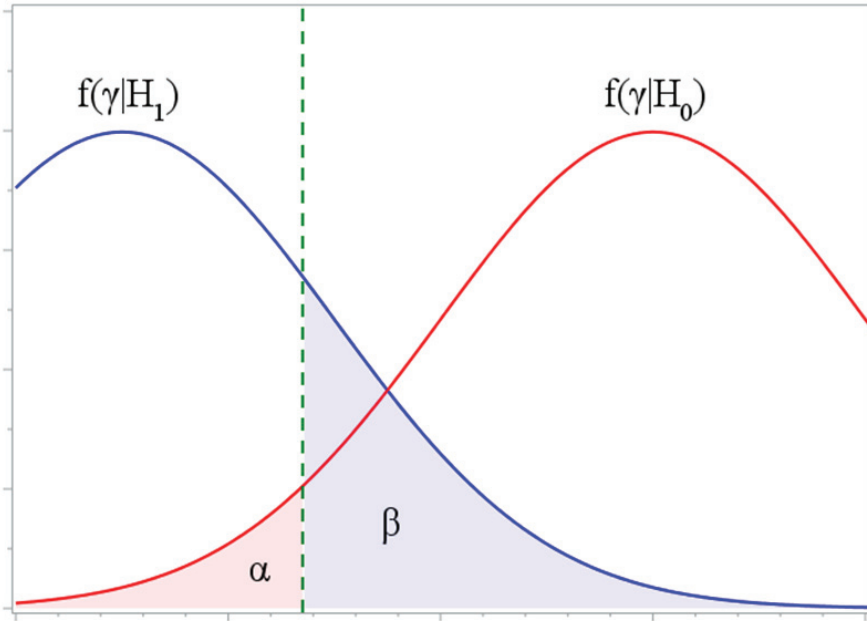


Figure 2.2 Test power scheme

2. When the distribution of the test statistic is unknown or cannot be approximated by any other known distribution, then a modeling study is performed to form a stepwise function. When the exact null distribution of a goodness-of-fit test statistic is a step function created by the summation of the exact probabilities for each possible value of the test statistic, it is possible to obtain the same critical value for a number of different adjacent significance levels α . Linear interpolation of the power of the test statistic using the power for significance levels less than (denoted α_1) and greater than (denoted α_2) the desired significance level (denoted as α) is preferred by many authors to overcome this problem (see for example [96]). Linear interpolation gives a weighting to the power based on how close α_1 and α_2 are to α (see in Figure 2.3). In this case, the power of the test is calculated according to the formula [21]:

$$\mathbf{Power} = \frac{(\alpha - \alpha_1)P(T \geq \gamma_2(\alpha)|H_1) + (\alpha_2 - \alpha)P(T \geq \gamma_1(\alpha)|H_1)}{\alpha_2 - \alpha_1}, \quad (2.92)$$

where $\gamma_1(\alpha)$ and $\gamma_2(\alpha)$ are the critical values immediately below and above the significance level α . $\alpha_1 = P(T \geq \gamma_1(\alpha)|H_0)$ and $\alpha_2 = P(T \geq \gamma_2(\alpha)|H_0)$ are the significance levels for $\gamma_1(\alpha)$ and $\gamma_2(\alpha)$ respectively.

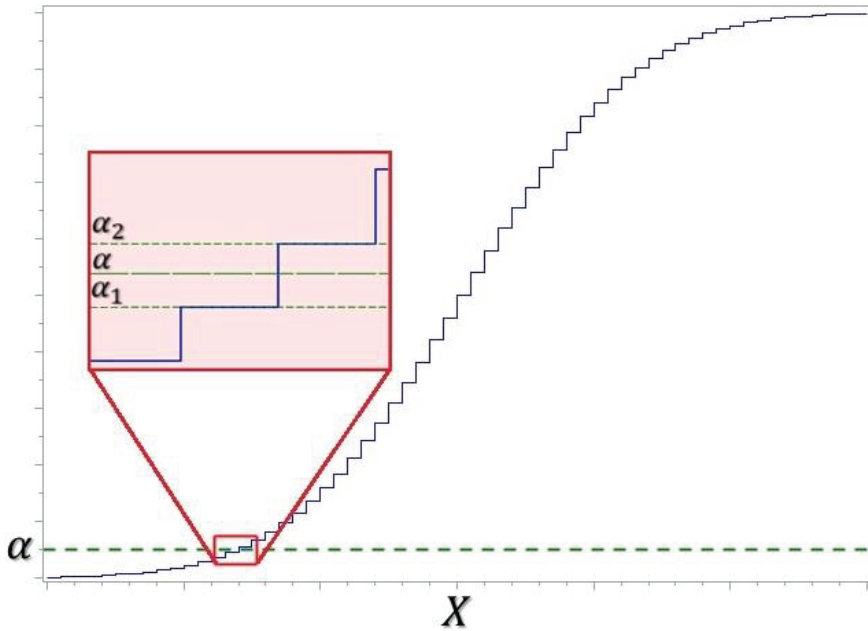


Figure 2.3 Significance levels for the statistic step function

III. RESEARCH RESULTS AND DISCUSSION

This section presents the results of a modelling comparison of the univariate and multivariate tests of the goodness-of-fit hypothesis. Then, the application of the tests to real data is presented. Finally, conclusions summarizing the results of the work are presented.

3.1. A comparative power analysis of univariate tests

This section provides a comprehensive modeling study that is designed to evaluate the power of selected normality tests. This modeling study takes into account the effects of sample size, the level of significance ($\alpha = 0.05$) chosen, and the alternative type of distribution (Beta, Cauchy, Laplace, Logistic, Student, Chi-Square, Gamma, Gumbel, Lognormal, Weibull, and modified standard normal). The study was performed by applying 42 normality tests (including our proposed normality test) for 1 000 000 generated units with standardized samples of size 32, 64, 128, 256, 512, and 1,024 [P1].

The best set of parameters (a , b , c) was selected experimentally: the value of a was examined from 0.001 to 0.99 by a step of 0.01, the value of b was examined from 0.01 to 10 by a step of 0.01, and the value of c was examined from 0.5 to 50 by a step of 0.25. The N1 gave the most powerful results with the values of the parameters: $a = 0.95$. $b = 0.25$. $c = 1$.

Table 3.1. Average empirical power results for groups of symmetric distributions

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
AD	0.714	0.169	0.799	0.123	0.863	0.089	0.909	0.062	0.939	0.041	0.957	0.031
Baksh	0.211	0.032	0.472	0.061	0.866	0.053	0.995	0.005	0.999	83E-8	0.999	83E-8
BCMR	0.718	0.171	0.809	0.123	0.875	0.089	0.92	0.059	0.947	0.038	0.962	0.023
BHS	0.431	0.189	0.551	0.194	0.663	0.175	0.752	0.139	0.818	0.108	0.868	0.084
BHSBS	0.68	0.191	0.778	0.14	0.85	0.105	0.903	0.07	0.938	0.044	0.959	0.03
BM1	0.628	0.173	0.754	0.13	0.853	0.096	0.917	0.064	0.949	0.038	0.962	0.022
BM2	0.726	0.166	0.835	0.118	0.905	0.08	0.945	0.049	0.965	0.029	0.974	0.017
BS	0.717	0.181	0.81	0.129	0.877	0.09	0.92	0.06	0.947	0.039	0.961	0.024
CC1	0.658	0.177	0.743	0.135	0.816	0.103	0.877	0.075	0.921	0.051	0.948	0.033
CC2	0.712	0.176	0.805	0.121	0.873	0.085	0.92	0.056	0.949	0.035	0.962	0.02
CHI2	0.663	0.21	0.778	0.177	0.842	0.153	0.884	0.144	0.941	0.089	0.945	0.096
CVM	0.591	0.135	0.733	0.093	0.805	0.069	0.855	0.049	0.919	0.028	0.949	0.017
ChenS	0.717	0.174	0.806	0.126	0.871	0.091	0.915	0.062	0.943	0.042	0.96	0.026
Coin	0.735	0.171	0.83	0.117	0.891	0.081	0.93	0.053	0.952	0.034	0.963	0.021
DA	0.266	0.121	0.295	0.095	0.314	0.066	0.319	0.043	0.315	0.028	0.311	0.021
DAP	0.723	0.185	0.82	0.13	0.883	0.09	0.924	0.059	0.948	0.037	0.962	0.023
DH	0.709	0.174	0.805	0.126	0.877	0.09	0.925	0.059	0.95	0.036	0.963	0.022
DLDMLRn	0.705	0.18	0.8	0.128	0.869	0.09	0.918	0.06	0.947	0.038	0.962	0.023
DLDMXAPD	0.725	0.174	0.817	0.126	0.883	0.088	0.925	0.057	0.95	0.036	0.962	0.022
DLDMZEPD	0.73	0.178	0.826	0.122	0.889	0.083	0.929	0.054	0.952	0.034	0.963	0.021
EP	0.706	0.171	0.797	0.123	0.859	0.094	0.91	0.06	0.946	0.038	0.96	0.025
Filli	0.712	0.166	0.805	0.121	0.875	0.087	0.922	0.058	0.949	0.036	0.962	0.022
GG	0.658	0.166	0.76	0.125	0.85	0.092	0.915	0.062	0.949	0.037	0.962	0.022
GLB	0.712	0.171	0.798	0.124	0.863	0.09	0.909	0.062	0.939	0.041	0.957	0.027
GMG	0.787	0.132	0.862	0.092	0.914	0.062	0.946	0.04	0.965	0.025	0.975	0.014
H1	0.799	0.166	0.862	0.118	0.852	0.09	0.999	53E-6	0.999	21E-8	0.999	53E-8
H2	0.728	0.209	0.8	0.153	0.775	0.116	0.764	0.085	0.762	0.062	0.766	0.045
H3	0.67	0.241	0.751	0.188	0.716	0.138	0.698	0.1	0.692	0.073	0.694	0.053
H4	0.631	0.255	0.706	0.218	0.662	0.166	0.641	0.118	0.634	0.084	0.636	0.061
JB	0.643	0.173	0.762	0.13	0.856	0.096	0.918	0.063	0.949	0.038	0.963	0.022
KS	0.585	0.104	0.723	0.072	0.789	0.056	0.836	0.041	0.905	0.025	0.939	0.016
Lillie	0.671	0.178	0.758	0.133	0.828	0.1	0.883	0.072	0.921	0.049	0.947	0.033
MI	0.632	0.143	0.676	0.098	0.705	0.068	0.724	0.046	0.736	0.029	0.745	0.019
NI	0.245	0.01	0.585	0.015	0.97	0.004	0.999	49E-8	0.999	49E-8	0.999	49E-9
SF	0.715	0.166	0.807	0.121	0.876	0.087	0.923	0.057	0.949	0.036	0.962	0.022
SW	0.718	0.172	0.808	0.124	0.874	0.089	0.919	0.06	0.946	0.039	0.962	0.024
SWRG	0.694	0.185	0.775	0.133	0.834	0.1	0.882	0.074	0.916	0.055	0.946	0.04
ZQ	0.513	0.23	0.574	0.201	0.628	0.16	0.668	0.123	0.696	0.092	0.717	0.066
ZQqstar	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15
ZQstar	0.513	0.231	0.576	0.202	0.63	0.161	0.669	0.124	0.697	0.091	0.718	0.064

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
ZW1	0.71	0.178	0.799	0.129	0.864	0.095	0.91	0.066	0.94	0.044	0.959	0.028
ZW2	0.715	0.17	0.806	0.122	0.869	0.087	0.912	0.06	0.939	0.041	0.957	0.027

Table 3.2. Average empirical power results for groups of asymmetric distributions

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
AD	0.729	0.159	0.835	0.112	0.908	0.068	0.949	0.037	0.969	0.022	0.984	0.015
Baksh	0.173	0.013	0.385	0.026	0.809	0.027	0.997	0.002	0.999	14E-8	0.999	22E-9
BCMR	0.749	0.15	0.856	0.1	0.924	0.056	0.958	0.03	0.977	0.018	0.991	0.01
BHS	0.529	0.183	0.664	0.167	0.769	0.138	0.855	0.102	0.915	0.066	0.95	0.038
BHSBS	0.538	0.207	0.652	0.182	0.747	0.15	0.834	0.113	0.902	0.072	0.944	0.046
BM1	0.644	0.182	0.802	0.122	0.901	0.068	0.952	0.033	0.975	0.018	0.989	0.011
BM2	0.737	0.156	0.859	0.102	0.931	0.056	0.965	0.028	0.981	0.016	0.993	0.008
BS	0.506	0.236	0.588	0.226	0.665	0.205	0.738	0.177	0.805	0.143	0.859	0.109
CC1	0.735	0.155	0.846	0.103	0.919	0.059	0.957	0.031	0.977	0.019	0.991	0.01
CC2	0.579	0.23	0.682	0.196	0.777	0.148	0.856	0.096	0.916	0.051	0.956	0.025
CHI2	0.645	0.203	0.799	0.149	0.881	0.102	0.934	0.058	0.965	0.028	0.98	0.018
CVM	0.594	0.14	0.755	0.098	0.836	0.064	0.887	0.035	0.935	0.018	0.957	0.013
ChenS	0.756	0.147	0.862	0.097	0.928	0.054	0.961	0.029	0.978	0.018	0.991	0.01
Coin	0.48	0.225	0.556	0.218	0.63	0.202	0.7	0.184	0.769	0.161	0.831	0.13
DA	0.237	0.147	0.223	0.128	0.209	0.108	0.198	0.089	0.191	0.072	0.186	0.057
DAP	0.705	0.176	0.826	0.118	0.91	0.066	0.955	0.034	0.977	0.019	0.99	0.01
DH	0.724	0.161	0.845	0.107	0.921	0.058	0.957	0.03	0.977	0.019	0.991	0.01
DLDMLRn	0.576	0.227	0.686	0.197	0.78	0.156	0.856	0.114	0.912	0.075	0.949	0.046
DLDMXAPD	0.726	0.162	0.843	0.109	0.918	0.062	0.955	0.033	0.975	0.02	0.989	0.012
DLDMZEPD	0.545	0.237	0.639	0.219	0.724	0.189	0.798	0.153	0.858	0.115	0.905	0.081
EP	0.739	0.16	0.846	0.107	0.913	0.066	0.951	0.033	0.975	0.02	0.988	0.01
Filli	0.732	0.157	0.842	0.107	0.915	0.061	0.953	0.032	0.974	0.02	0.989	0.011
GG	0.672	0.202	0.805	0.132	0.898	0.073	0.949	0.035	0.973	0.019	0.988	0.011
GLB	0.725	0.162	0.831	0.115	0.905	0.071	0.947	0.039	0.969	0.023	0.984	0.015
GMG	0.683	0.171	0.751	0.148	0.809	0.122	0.859	0.098	0.901	0.072	0.932	0.05
H1	0.816	0.148	0.896	0.095	0.896	0.068	0.999	0.015	0.999	0.003	0.999	2E-7
H2	0.744	0.191	0.833	0.141	0.826	0.106	0.83	0.077	0.836	0.053	0.845	0.037
H3	0.697	0.216	0.784	0.17	0.764	0.132	0.764	0.098	0.77	0.069	0.781	0.049
H4	0.669	0.231	0.749	0.189	0.715	0.149	0.71	0.111	0.714	0.08	0.726	0.057
JB	0.662	0.179	0.808	0.12	0.904	0.067	0.953	0.033	0.975	0.018	0.989	0.011
KS	0.582	0.111	0.736	0.081	0.81	0.059	0.863	0.036	0.921	0.018	0.945	0.012
Lillie	0.671	0.177	0.786	0.123	0.872	0.09	0.929	0.053	0.959	0.03	0.976	0.018
MI	0.644	0.198	0.731	0.162	0.798	0.125	0.843	0.093	0.872	0.068	0.89	0.051
N1	0.464	0.061	0.761	0.028	0.99	0.002	0.999	5E-8	0.999	41E-8	0.999	55E-9
SF	0.736	0.156	0.846	0.105	0.918	0.059	0.955	0.031	0.975	0.02	0.989	0.011

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
SW	0.753	0.149	0.859	0.098	0.925	0.055	0.959	0.03	0.977	0.018	0.991	0.01
SWRG	0.758	0.145	0.861	0.097	0.927	0.056	0.96	0.03	0.977	0.018	0.99	0.011
ZQ	0.22	0.085	0.21	0.066	0.202	0.049	0.191	0.036	0.183	0.028	0.174	0.021
ZQstar	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	4E-15
ZQstar	0.57	0.17	0.639	0.139	0.693	0.11	0.732	0.085	0.761	0.065	0.781	0.049
ZW1	0.749	0.143	0.852	0.094	0.918	0.054	0.955	0.03	0.976	0.017	0.989	0.01
ZW2	0.764	0.145	0.87	0.093	0.932	0.051	0.962	0.028	0.98	0.017	0.991	0.009

Table 3.3. Average empirical power results for groups of modified normal distributions

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
AD	0.662	0.176	0.756	0.134	0.825	0.098	0.872	0.072	0.905	0.055	0.931	0.042
Baksh	0.222	0.02	0.519	0.036	0.931	0.017	0.999	86E-6	0.999	4E-13	0.999	8E-14
BCMR	0.652	0.179	0.756	0.135	0.831	0.096	0.88	0.068	0.913	0.05	0.935	0.038
BHS	0.463	0.189	0.585	0.185	0.676	0.16	0.744	0.132	0.796	0.108	0.834	0.089
BHSBS	0.568	0.214	0.701	0.172	0.787	0.128	0.847	0.093	0.89	0.068	0.918	0.052
BM1	0.494	0.173	0.652	0.149	0.779	0.112	0.86	0.079	0.904	0.056	0.93	0.041
BM2	0.641	0.176	0.77	0.135	0.854	0.093	0.904	0.063	0.934	0.043	0.953	0.03
BS	0.587	0.218	0.688	0.186	0.77	0.147	0.833	0.112	0.881	0.084	0.916	0.06
CC1	0.567	0.187	0.669	0.157	0.761	0.119	0.83	0.087	0.878	0.065	0.912	0.05
CC2	0.576	0.202	0.675	0.166	0.763	0.127	0.833	0.093	0.887	0.066	0.923	0.047
CHI2	0.566	0.226	0.728	0.168	0.808	0.125	0.866	0.089	0.914	0.058	0.939	0.041
CVM	0.557	0.138	0.708	0.095	0.779	0.071	0.833	0.05	0.897	0.033	0.93	0.024
ChenS	0.656	0.18	0.759	0.136	0.833	0.096	0.882	0.07	0.915	0.051	0.937	0.038
Coin	0.579	0.214	0.691	0.18	0.781	0.137	0.846	0.099	0.889	0.07	0.918	0.052
DA	0.314	0.156	0.342	0.141	0.367	0.12	0.388	0.097	0.405	0.077	0.418	0.06
DAP	0.617	0.201	0.733	0.158	0.818	0.11	0.872	0.076	0.906	0.055	0.93	0.042
DH	0.617	0.195	0.727	0.156	0.815	0.111	0.872	0.076	0.907	0.055	0.93	0.041
DLMLRn	0.581	0.211	0.687	0.175	0.78	0.131	0.852	0.091	0.902	0.062	0.933	0.041
DLDMXAPD	0.651	0.189	0.754	0.146	0.831	0.104	0.879	0.073	0.912	0.055	0.935	0.041
DLDMZEPD	0.595	0.22	0.698	0.184	0.78	0.142	0.842	0.106	0.887	0.077	0.919	0.056
EP	0.64	0.179	0.748	0.136	0.819	0.096	0.865	0.07	0.906	0.056	0.931	0.046
Filli	0.637	0.178	0.743	0.136	0.823	0.098	0.877	0.069	0.911	0.049	0.933	0.036
GG	0.529	0.192	0.657	0.151	0.775	0.112	0.86	0.078	0.906	0.054	0.932	0.039
GLB	0.659	0.178	0.755	0.135	0.823	0.1	0.87	0.072	0.903	0.056	0.93	0.044
GMG	0.688	0.165	0.771	0.137	0.836	0.104	0.883	0.075	0.917	0.055	0.942	0.041
H1	0.743	0.187	0.816	0.138	0.799	0.104	0.999	22E-8	0.999	17E-9	0.999	95E-9
H2	0.711	0.197	0.775	0.153	0.741	0.117	0.723	0.087	0.716	0.064	0.716	0.047
H3	0.688	0.209	0.749	0.167	0.704	0.127	0.68	0.093	0.67	0.068	0.668	0.051
H4	0.675	0.22	0.73	0.179	0.677	0.136	0.649	0.098	0.637	0.072	0.634	0.053
JB	0.515	0.176	0.662	0.149	0.783	0.111	0.861	0.079	0.904	0.056	0.93	0.041

Tests	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
KS	0.564	0.108	0.71	0.073	0.772	0.058	0.825	0.042	0.893	0.026	0.924	0.02
Lillie	0.626	0.184	0.724	0.143	0.796	0.109	0.85	0.082	0.889	0.062	0.917	0.048
MI	0.494	0.197	0.536	0.161	0.563	0.127	0.578	0.098	0.585	0.075	0.59	0.059
N1	0.243	0.009	0.581	0.014	0.972	0.003	0.999	43E-8	0.999	27E-9	0.999	69E-11
SF	0.642	0.177	0.747	0.135	0.826	0.096	0.879	0.068	0.912	0.049	0.934	0.036
SW	0.654	0.18	0.758	0.135	0.832	0.096	0.882	0.068	0.915	0.049	0.937	0.036
SWRG	0.643	0.188	0.746	0.145	0.818	0.101	0.864	0.076	0.901	0.059	0.931	0.046
ZQ	0.382	0.189	0.422	0.182	0.459	0.165	0.489	0.144	0.562	0.125	0.522	0.103
ZQqstar	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	8E-15	0.05	3E-15	0.05	2E-15
ZQstar	0.394	0.173	0.423	0.157	0.45	0.136	0.472	0.115	0.487	0.097	0.498	0.083
ZW1	0.635	0.181	0.737	0.14	0.816	0.1	0.87	0.072	0.906	0.053	0.931	0.039
ZW2	0.64	0.183	0.749	0.14	0.826	0.097	0.876	0.069	0.907	0.051	0.931	0.039

Because an extensive study of the power of tests for testing the goodness-of-fit hypothesis for different sample sizes and various distributions, has yielded many results, the best tests from their modifications are presented here. Tables 3.1–3.3 present averaged power results for the symmetric, asymmetric, and modified normal distribution sets, for samples sizes of 32, 64, 128, 256, 512, and 1024. Comparing Tables 3.1–3.3, it can be seen that the most powerful test for symmetric, asymmetric, and modified normal distribution groups was Hosking1 (*H1*) only for small samples, and for large sample sizes the most powerful test was our presented test (N1). In all examined cases, the standard deviation of N1 was lower than H1. According to Tables 3.1–3.3, it is observed that for large sample sizes, the power of most tests is approaching 1, except for the D’Agostin (DA) test, the power of which is significantly lower.

An additional study has been conducted to determine the exact minimal sample size at which the N1 test becomes the most powerful for groups of symmetric, asymmetric, and modified normal distributions. Hosking1 and N1 tests have been applied for data sets of sizes: 80, 90, 100, 105, 110, and 115. The obtained results showed that the N1 test is the most powerful where the mean sample size is 112 for the symmetric distributions, the mean sample size is 118 for the asymmetric distributions, and the mean sample size is 88 for a group of modified normal distributions (see in Table 3.4). The N1 test is the most powerful for the Gamma distribution for samples sizes less than 32. It has been observed that in the case of Cauchy and Lognormal distributions, the *N-metric* test is the most powerful when the sample size is >255 influenced by the long tail of these distributions [P1].

Table 3.4. The minimal sample size at which the N1 test is most powerful.

Nr.	Distribution	Groups of Distributions	Minimal sample size
1.	Standard normal	Symmetric	46
2.	Beta	Symmetric	88
3.	Cauchy	Symmetric	257
4.	Laplace	Symmetric	117

Nr.	Distribution	Groups of Distributions	Minimal sample size
5.	Logistic	Symmetric	71
6.	Student	Symmetric	96
7.	Beta	Asymmetric	108
8.	Chi-square	Asymmetric	123
9.	Gamma	Asymmetric	< 32
10.	Gumbel	Asymmetric	125
11.	Lognormal	Asymmetric	255
12.	Weibull	Asymmetric	65
13.	Normal1	Modified normal	70
14.	Normal2	Modified normal	93
15.	Normal3	Modified normal	72
16.	Normal4	Modified normal	117

To complement the results given in Tables 3.1–3.3, presents the average power results of the most powerful goodness-of-fit tests. The graphs drawn by the trellis display method [112] are presented in Figures 3.1–3.3. In Figure 3.1, it can be seen that for standard normal distribution, our proposed test (N1) is the most powerful when the sample size is 64 or larger. Figure 3.2 shows that our proposed test (N1) is the most powerful in the case of Gamma data distribution for all examined sample sizes. Figure 3.3, shows that for all the modified normal distributions, the N1 test is the strongest when $n \geq 128$. In summary, the power of the Chen–Shapiro (ChenS), Gel–Miao–Gastwirth (GMG), Hosking1 (H1), and Modified Shapiro–Wilk (SWRG) tests increases gradually with increasing sample size. On the other hand, the power of our proposed test (N1) increases abruptly when the sample size is 128 and its power value remains close to 1 for larger sample sizes.

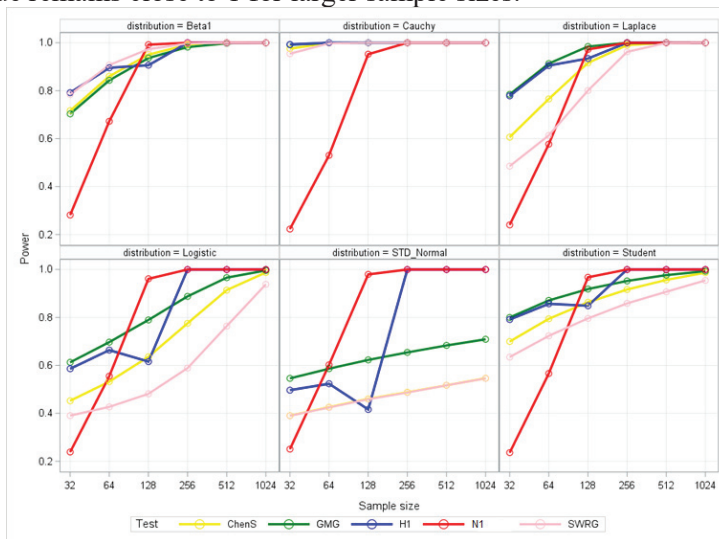


Figure 3.1. Average empirical power results. for all sample sizes. for the symmetric distributions of five powerful goodness-of-fit tests.

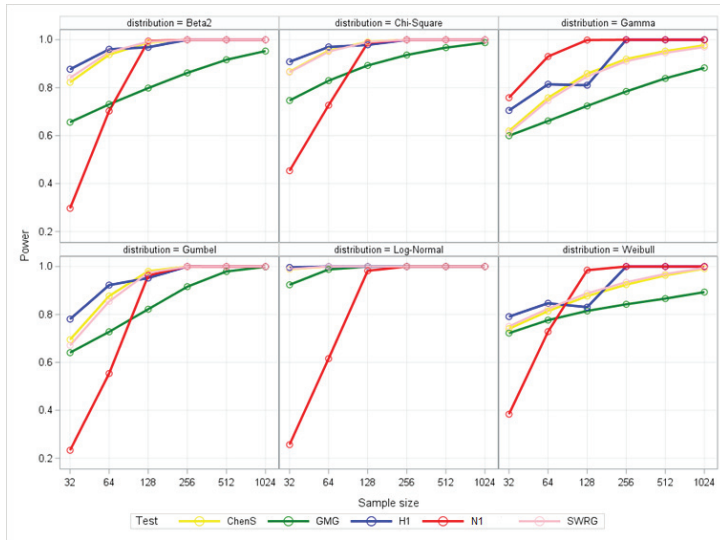


Figure 3.2. Average empirical power results. for all sample sizes. for the asymmetric distributions of five powerful goodness-of-fit tests.

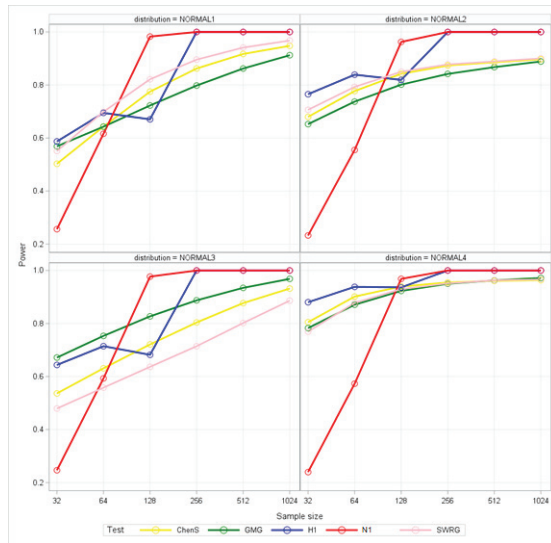


Figure 3.3. Average empirical power results. for all sample sizes. for the modified normal distributions of five powerful goodness-of-fit tests.

3.2. A comparative power analysis of multivariate tests

This section provides a modeling study that evaluates the power of selected multivariate normality tests. We used the Monte Carlo method to compare our proposed test with 13 multivariate tests described above for dimensions $d = 2, 5,$ and 10 with sample sizes $n = 32, 64, 128, 256, 512,$ and $1,024$ at significance level $\alpha=0.05$. Power was estimated by applying the tests to 1,000,000 randomly drawn

samples from the alternative distribution (Beta, Cauchy, Laplace, Logistic, Student, Standard normal, Chi-Square, Gamma, Gumbel, Lognormal, Weibull, Mixed, Normal mixture).

The test smoothness parameter (h) values were selected experimentally: from 0.1 to 5 with a step of 0.1. The value of the test h parameter was determined for each considered dimension. It was found that the best results are obtained (i.e., maximum statistical value) for $d = 2$ with $h = 1.05$, for $d = 5$ with $h = 0.1$, and for $d = 10$ with $h = 2.4$. These smoothness parameter h values were used to carry out the numerical experiments.

The power of 13 (including our proposed test) multivariate goodness-of-fit hypothesis tests was calculated for different sample sizes, distributions, and mixtures. The mean power values for the groups for distributions for each test and sample size have been computed and are presented in Tables 3.5–3.8. The new test for the groups of symmetric and mixed distributions is the most powerful one. In the group of asymmetric distributions, the new (for $d = 2$) and Roy (for $d = 5$ and 10) tests are the most powerful ones. The new (for $d = 2$ and 5) and Roy (for $d = 10$ with sample sizes $n = 256, 512,$ and $1,024$) tests are also the most powerful in the group of normal distribution mixtures. Comparing the Mardia (Mar1 and Mar2) tests, based on asymmetry and excess coefficients, it has been found that Mar1 is the most powerful only for the group of asymmetric distributions. For the group of symmetric distributions, the power of this test is the lowest (compared to other tests) [P2, K4, K5].

Table 3.5. An average empirical power for a group of symmetric distributions.

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
d=2												
AD	0.651	0.2	0.778	0.151	0.867	0.099	0.92	0.054	0.939	0.028	0.945	0.021
CHI2	0.57	0.205	0.692	0.176	0.798	0.135	0.873	0.086	0.912	0.057	0.932	0.041
CVM	0.652	0.2	0.779	0.152	0.868	0.098	0.92	0.053	0.94	0.029	0.945	0.021
DH	0.677	0.18	0.809	0.128	0.892	0.076	0.932	0.041	0.945	0.025	0.949	0.022
DN	0.565	0.208	0.671	0.181	0.768	0.145	0.847	0.104	0.903	0.069	0.937	0.04
Energy	0.65	0.189	0.77	0.145	0.86	0.095	0.914	0.058	0.941	0.034	0.948	0.024
HZ	0.644	0.186	0.765	0.143	0.853	0.097	0.906	0.063	0.936	0.039	0.947	0.027
LV	0.696	0.182	0.815	0.127	0.893	0.076	0.932	0.041	0.945	0.025	0.949	0.022
N2	0.999	0.003	0.999	95E-9	0.999	9E-10	0.999	61E-10	0.999	94E-10	0.999	3E-13
Mar1	0.532	0.189	0.617	0.17	0.681	0.144	0.721	0.126	0.743	0.114	0.758	0.108
Mar2	0.605	0.183	0.751	0.147	0.857	0.099	0.917	0.055	0.942	0.03	0.95	0.022
NRR	0.608	0.202	0.736	0.165	0.842	0.116	0.91	0.068	0.941	0.034	0.949	0.023
Roy	0.703	0.177	0.819	0.123	0.891	0.078	0.929	0.043	0.944	0.028	0.95	0.022
d=5												
AD	0.644	0.184	0.791	0.14	0.883	0.084	0.929	0.043	0.942	0.025	0.946	0.021
CHI2	0.531	0.191	0.656	0.177	0.773	0.145	0.864	0.093	0.916	0.055	0.941	0.033
CVM	0.624	0.181	0.775	0.144	0.876	0.089	0.926	0.046	0.942	0.026	0.946	0.021
DH	0.735	0.149	0.864	0.086	0.924	0.044	0.941	0.026	0.946	0.023	0.949	0.022

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
DN	0.585	0.193	0.7	0.167	0.806	0.127	0.886	0.084	0.932	0.045	0.949	0.027
Energy	0.632	0.171	0.758	0.137	0.863	0.089	0.921	0.051	0.945	0.028	0.949	0.022
HZ	0.622	0.172	0.755	0.146	0.856	0.098	0.91	0.062	0.94	0.035	0.949	0.024
LV	0.763	0.15	0.871	0.085	0.925	0.044	0.941	0.026	0.946	0.023	0.949	0.022
N2	0.985	0.041	0.989	0.027	0.988	0.027	0.987	0.03	0.981	0.032	0.985	0.03
Mar1	0.523	0.155	0.621	0.14	0.696	0.11	0.735	0.087	0.752	0.074	0.764	0.07
Mar2	0.637	0.162	0.792	0.131	0.89	0.074	0.934	0.036	0.948	0.024	0.95	0.023
NRR	0.602	0.186	0.739	0.159	0.851	0.108	0.916	0.06	0.944	0.031	0.95	0.022
Roy	0.784	0.135	0.875	0.084	0.924	0.046	0.941	0.029	0.947	0.023	0.95	0.021
d=10												
AD	0.557	0.172	0.754	0.145	0.878	0.083	0.928	0.041	0.942	0.023	0.945	0.021
CHI2	0.473	0.198	0.604	0.173	0.726	0.156	0.824	0.114	0.891	0.071	0.928	0.042
CVM	0.534	0.171	0.728	0.148	0.865	0.091	0.922	0.047	0.941	0.025	0.945	0.021
DH	0.754	0.126	0.884	0.062	0.934	0.03	0.941	0.023	0.945	0.023	0.948	0.022
DN	0.599	0.194	0.704	0.161	0.817	0.123	0.896	0.078	0.936	0.041	0.949	0.024
Energy	0.598	0.157	0.709	0.132	0.821	0.094	0.906	0.055	0.943	0.028	0.948	0.021
HZ	0.604	0.149	0.71	0.149	0.831	0.114	0.901	0.066	0.937	0.039	0.949	0.024
LV	0.791	0.125	0.893	0.062	0.935	0.03	0.941	0.023	0.945	0.023	0.948	0.022
N2	0.997	0.026	0.998	0.023	0.998	0.017	0.998	0.013	0.991	0.037	0.991	0.04
Mar1	0.458	0.128	0.592	0.116	0.676	0.096	0.733	0.06	0.747	0.053	0.756	0.049
Mar2	0.65	0.147	0.802	0.127	0.899	0.077	0.94	0.033	0.951	0.025	0.95	0.022
NRR	0.599	0.188	0.719	0.158	0.844	0.11	0.913	0.063	0.942	0.032	0.949	0.022
Roy	0.834	0.101	0.905	0.058	0.934	0.032	0.943	0.024	0.946	0.022	0.95	0.021

Table 3.6. An average empirical power for a group of asymmetric distributions.

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
d=2												
AD	0.634	0.196	0.744	0.146	0.827	0.11	0.897	0.08	0.954	0.044	0.985	0.016
CHI2	0.631	0.227	0.767	0.173	0.861	0.125	0.931	0.083	0.977	0.039	0.996	0.012
CVM	0.639	0.196	0.744	0.144	0.822	0.112	0.892	0.083	0.949	0.048	0.982	0.019
DH	0.852	0.12	0.956	0.057	0.995	0.013	0.999	59E-5	0.999	11E-8	0.999	56E-9
DN	0.55	0.209	0.657	0.171	0.724	0.132	0.774	0.114	0.816	0.104	0.864	0.088
Energy	0.832	0.129	0.93	0.078	0.985	0.028	0.999	0.003	0.999	0.001	0.999	7E-5
HZ	0.811	0.13	0.906	0.086	0.968	0.042	0.995	0.01	0.999	61E-5	0.999	42E-7
LV	0.87	0.115	0.961	0.054	0.995	0.012	0.999	54E-5	0.999	95E-9	0.999	4E-11
N2	0.999	0.003	0.999	5E-11	0.999	64E-11	0.999	91E-11	0.999	3E-12	0.999	35E-13
Mar1	0.813	0.151	0.941	0.073	0.992	0.018	0.999	0.001	0.999	92E-8	0.999	11E-10
Mar2	0.63	0.244	0.776	0.191	0.876	0.139	0.947	0.086	0.988	0.033	0.999	0.007
NRR	0.639	0.204	0.759	0.154	0.841	0.12	0.915	0.087	0.968	0.047	0.993	0.016
Roy	0.877	0.106	0.962	0.052	0.995	0.012	0.999	58E-5	0.999	27E-8	0.999	12E-10
d=5												
AD	0.614	0.188	0.763	0.136	0.869	0.1	0.946	0.059	0.984	0.019	0.995	0.003
CHI2	0.6	0.212	0.779	0.155	0.892	0.108	0.965	0.056	0.994	0.015	0.999	0.002
CVM	0.608	0.186	0.761	0.135	0.869	0.101	0.947	0.059	0.985	0.02	0.995	0.003
DH	0.915	0.086	0.99	0.02	0.999	84E-5	0.999	46E-9	0.999	1E-10	0.999	37E-13
DN	0.551	0.203	0.675	0.156	0.748	0.121	0.812	0.107	0.872	0.089	0.926	0.063
Energy	0.854	0.117	0.958	0.055	0.996	0.011	0.999	24E-5	0.999	12E-8	0.999	59E-9

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
HZ	0.798	0.128	0.907	0.088	0.974	0.038	0.998	0.007	0.999	14E-5	0.999	1E-10
LV	0.932	0.077	0.992	0.017	0.999	71E-5	0.999	33E-9	0.999	91E-9	0.999	51E-10
N2	0.982	0.055	0.989	0.043	0.997	0.018	0.997	0.017	0.999	0.007	0.999	22E-8
Mar1	0.803	0.143	0.954	0.063	0.997	0.008	0.999	67E-6	0.999	17E-8	0.999	87E-8
Mar2	0.623	0.197	0.791	0.158	0.908	0.107	0.978	0.043	0.997	0.011	0.999	0.003
NRR	0.61	0.196	0.763	0.145	0.869	0.109	0.95	0.066	0.989	0.023	0.999	0.003
Roy	0.945	0.066	0.993	0.016	0.999	62E-5	0.999	1E-7	0.999	34E-8	0.999	11E-10
d=10												
AD	0.483	0.194	0.707	0.132	0.863	0.103	0.955	0.052	0.99	0.013	0.996	89E-5
CHI2	0.443	0.206	0.712	0.169	0.87	0.115	0.96	0.059	0.994	0.016	0.999	0.001
CVM	0.459	0.191	0.7	0.132	0.859	0.104	0.953	0.054	0.989	0.013	0.996	0.001
DH	0.944	0.065	0.998	0.007	0.999	44E-7	0.999	29E-8	0.999	9E-9	0.999	42E-11
DN	0.532	0.212	0.679	0.15	0.776	0.119	0.858	0.098	0.93	0.064	0.975	0.031
Energy	0.829	0.122	0.956	0.054	0.997	0.008	0.999	75E-6	0.999	3E-8	0.999	73E-9
HZ	0.744	0.13	0.861	0.102	0.954	0.055	0.994	0.015	0.999	56E-5	0.999	51E-9
LV	0.96	0.054	0.998	0.005	0.999	28E-7	0.999	36E-8	0.999	54E-9	0.999	4E-12
N2	0.922	0.132	0.947	0.116	0.98	0.073	0.995	0.025	0.996	0.019	0.996	0.02
Mar1	0.693	0.15	0.931	0.078	0.997	0.009	0.999	78E-6	0.999	49E-8	0.999	7E-11
Mar2	0.573	0.192	0.746	0.142	0.898	0.106	0.978	0.046	0.998	0.008	0.999	3E-5
NRR	0.532	0.208	0.722	0.141	0.86	0.109	0.952	0.062	0.992	0.017	0.999	0.002
Roy	0.98	0.033	0.999	0.003	0.999	21E-7	0.999	29E-8	0.999	15E-9	0.999	67E-11

Table 3.7. An average empirical power for a group of mixed distributions.

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
d=2												
AD	0.469	0.214	0.572	0.208	0.683	0.188	0.78	0.144	0.848	0.094	0.883	0.063
CHI2	0.408	0.202	0.476	0.215	0.571	0.216	0.667	0.198	0.763	0.16	0.842	0.107
CVM	0.463	0.211	0.567	0.208	0.679	0.188	0.778	0.144	0.847	0.094	0.883	0.065
DH	0.436	0.201	0.51	0.192	0.591	0.173	0.66	0.159	0.746	0.144	0.826	0.113
DN	0.439	0.2	0.511	0.195	0.59	0.192	0.674	0.18	0.763	0.147	0.835	0.105
Energy	0.582	0.196	0.703	0.165	0.809	0.112	0.872	0.07	0.895	0.047	0.902	0.044
HZ	0.572	0.19	0.697	0.164	0.807	0.115	0.871	0.07	0.894	0.047	0.901	0.043
LV	0.453	0.203	0.547	0.195	0.667	0.163	0.749	0.13	0.808	0.109	0.857	0.088
N2	0.999	0.007	0.999	37E-5	0.999	2E-7	0.999	43E-9	0.999	7E-11	0.999	4E-12
Mar1	0.476	0.202	0.577	0.186	0.659	0.154	0.717	0.127	0.76	0.101	0.78	0.086
Mar2	0.412	0.193	0.527	0.216	0.651	0.209	0.762	0.162	0.843	0.108	0.884	0.066
NRR	0.444	0.202	0.533	0.203	0.641	0.191	0.741	0.16	0.827	0.111	0.878	0.074
Roy	0.451	0.184	0.513	0.183	0.572	0.178	0.643	0.165	0.72	0.138	0.764	0.115
d=5												
AD	0.626	0.207	0.749	0.132	0.805	0.097	0.852	0.074	0.88	0.057	0.894	0.045
CHI2	0.466	0.217	0.582	0.204	0.67	0.169	0.729	0.132	0.763	0.113	0.789	0.102
CVM	0.585	0.211	0.726	0.143	0.791	0.105	0.841	0.083	0.875	0.063	0.893	0.048
DH	0.545	0.209	0.631	0.178	0.695	0.142	0.747	0.113	0.777	0.091	0.795	0.087
DN	0.538	0.216	0.684	0.177	0.771	0.121	0.825	0.096	0.865	0.075	0.891	0.057
Energy	0.703	0.148	0.788	0.097	0.845	0.068	0.88	0.052	0.894	0.044	0.9	0.045
HZ	0.706	0.154	0.815	0.091	0.864	0.059	0.885	0.047	0.895	0.043	0.899	0.044
LV	0.553	0.21	0.628	0.179	0.692	0.142	0.751	0.113	0.81	0.091	0.883	0.062
N2	0.982	0.063	0.989	0.045	0.995	0.019	0.998	0.019	0.999	41E-7	0.999	8E-9
Mar1	0.584	0.174	0.675	0.133	0.722	0.105	0.75	0.092	0.766	0.085	0.778	0.085
Mar2	0.584	0.209	0.723	0.147	0.791	0.111	0.838	0.088	0.864	0.079	0.889	0.065
NRR	0.551	0.215	0.694	0.164	0.769	0.118	0.822	0.097	0.863	0.076	0.889	0.058

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
Roy	0.47	0.187	0.524	0.187	0.589	0.18	0.669	0.157	0.73	0.125	0.764	0.114
d=10												
AD	0.688	0.161	0.753	0.095	0.775	0.092	0.802	0.091	0.833	0.086	0.866	0.072
CHI2	0.477	0.219	0.579	0.21	0.651	0.192	0.709	0.155	0.746	0.121	0.763	0.109
CVM	0.642	0.18	0.744	0.104	0.771	0.093	0.795	0.091	0.821	0.087	0.853	0.078
DH	0.58	0.207	0.69	0.143	0.736	0.11	0.761	0.097	0.778	0.091	0.791	0.087
DN	0.669	0.197	0.753	0.12	0.776	0.102	0.793	0.093	0.818	0.089	0.842	0.085
Energy	0.719	0.104	0.744	0.086	0.777	0.084	0.823	0.077	0.875	0.058	0.897	0.043
HZ	0.745	0.106	0.78	0.081	0.821	0.081	0.87	0.061	0.892	0.044	0.899	0.043
LV	0.592	0.209	0.69	0.146	0.735	0.111	0.76	0.097	0.776	0.09	0.79	0.086
N2	0.916	0.148	0.942	0.095	0.94	0.093	0.968	0.078	0.995	0.033	0.997	0.016
Mar1	0.614	0.101	0.68	0.084	0.722	0.087	0.745	0.086	0.764	0.088	0.779	0.086
Mar2	0.731	0.169	0.796	0.098	0.795	0.094	0.811	0.094	0.84	0.092	0.861	0.082
NRR	0.679	0.188	0.753	0.115	0.774	0.1	0.79	0.092	0.814	0.089	0.837	0.086
Roy	0.475	0.194	0.529	0.195	0.602	0.184	0.689	0.156	0.745	0.13	0.769	0.122

Table 3.8. An average empirical power for a group of normal mixture distributions.

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
d=2												
AD	0.465	0.221	0.576	0.225	0.71	0.205	0.844	0.149	0.943	0.073	0.987	0.019
CHI2	0.422	0.211	0.508	0.228	0.618	0.229	0.738	0.196	0.845	0.136	0.917	0.077
CVM	0.468	0.221	0.581	0.225	0.715	0.205	0.848	0.15	0.945	0.073	0.988	0.018
DH	0.56	0.221	0.74	0.178	0.893	0.105	0.974	0.04	0.998	0.006	0.999	22E-6
DN	0.428	0.205	0.503	0.211	0.582	0.215	0.685	0.193	0.791	0.157	0.882	0.111
Energy	0.537	0.22	0.682	0.199	0.836	0.139	0.938	0.068	0.986	0.022	0.999	0.002
HZ	0.529	0.214	0.672	0.198	0.823	0.145	0.926	0.077	0.977	0.031	0.998	0.006
LV	0.588	0.223	0.752	0.176	0.895	0.104	0.974	0.04	0.998	0.006	0.999	22E-6
N2	0.999	0.004	0.999	2E-8	0.999	7E-9	0.999	19E-9	0.999	2E-11	0.999	9E-12
Mar1	0.433	0.211	0.544	0.204	0.633	0.178	0.701	0.152	0.733	0.138	0.757	0.132
Mar2	0.437	0.201	0.563	0.226	0.707	0.217	0.84	0.159	0.931	0.084	0.977	0.032
NRR	0.442	0.208	0.544	0.22	0.664	0.217	0.805	0.173	0.924	0.099	0.985	0.031
Roy	0.607	0.212	0.778	0.158	0.908	0.093	0.978	0.035	0.998	0.006	0.999	81E-6
d=5												
AD	0.45	0.211	0.574	0.213	0.699	0.184	0.806	0.141	0.889	0.095	0.946	0.056
CHI2	0.399	0.196	0.491	0.218	0.598	0.212	0.702	0.183	0.787	0.145	0.857	0.105
CVM	0.441	0.206	0.563	0.213	0.689	0.186	0.798	0.144	0.883	0.098	0.942	0.059
DH	0.594	0.206	0.782	0.149	0.916	0.077	0.979	0.029	0.998	0.005	0.999	29E-5
DN	0.443	0.208	0.51	0.211	0.594	0.2	0.691	0.177	0.782	0.145	0.859	0.106
Energy	0.503	0.21	0.64	0.193	0.781	0.145	0.894	0.085	0.963	0.038	0.992	0.011
HZ	0.485	0.199	0.621	0.201	0.761	0.163	0.877	0.106	0.95	0.052	0.985	0.02
LV	0.632	0.206	0.795	0.146	0.92	0.076	0.979	0.029	0.998	0.005	0.999	29E-5
N2	0.98	0.061	0.994	0.028	0.997	0.016	0.999	0.006	0.999	0.003	0.999	0.001
Mar1	0.384	0.199	0.516	0.196	0.619	0.169	0.694	0.138	0.736	0.115	0.758	0.103
Mar2	0.46	0.187	0.59	0.212	0.728	0.188	0.832	0.142	0.905	0.093	0.954	0.049
NRR	0.442	0.204	0.539	0.215	0.655	0.196	0.766	0.157	0.857	0.113	0.925	0.071
Roy	0.672	0.186	0.828	0.128	0.934	0.066	0.984	0.024	0.998	0.004	0.999	45E-5
d=10												
AD	0.402	0.188	0.556	0.203	0.709	0.165	0.801	0.116	0.853	0.085	0.893	0.071
CHI2	0.392	0.19	0.47	0.212	0.587	0.207	0.688	0.17	0.76	0.138	0.818	0.115
CVM	0.396	0.185	0.537	0.201	0.692	0.172	0.79	0.122	0.846	0.089	0.886	0.074

	Sample size											
	32		64		128		256		512		1024	
	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD	Power	STD
DH	0.62	0.182	0.8	0.121	0.915	0.069	0.973	0.03	0.995	0.008	0.999	72E-5
DN	0.495	0.216	0.536	0.209	0.629	0.193	0.723	0.16	0.8	0.119	0.85	0.089
Energy	0.476	0.204	0.599	0.182	0.723	0.143	0.834	0.1	0.912	0.062	0.96	0.033
HZ	0.487	0.165	0.581	0.189	0.708	0.166	0.815	0.118	0.887	0.081	0.939	0.052
LV	0.667	0.181	0.815	0.118	0.919	0.067	0.974	0.03	0.995	0.008	0.999	7E-4
N2	0.989	0.035	0.984	0.035	0.977	0.036	0.971	0.03	0.971	0.031	0.973	0.031
Mar1	0.287	0.162	0.472	0.175	0.582	0.15	0.669	0.119	0.711	0.095	0.739	0.085
Mar2	0.547	0.16	0.634	0.195	0.758	0.151	0.834	0.106	0.885	0.084	0.924	0.066
NRR	0.487	0.206	0.55	0.21	0.674	0.186	0.771	0.138	0.833	0.099	0.875	0.079
Roy	0.735	0.151	0.855	0.1	0.939	0.052	0.98	0.023	0.997	0.006	0.999	63E-5

In order to supplement and emphasize the results presented in Tables 3.5–3.8, the generalized line diagrams were drawn. The resulting graph is shown in Figure 3.4 showing that the New test (N2) is significantly more powerful than the other tests. The power of the Mar1 tests is the lowest compared with other tests. Figure 3.4 indicates that the power of tests increases as the sample size increases. By increasing the dimension the power of eight tests (AD, CHI2, CVM, Energy, HZ, N2, Mar1, and NRR) decreases while the power of the other tests (DH, DN, LV, Mar2, and Roy) increases slightly. For small sample sizes, the most powerful tests are N2, Roy, and DH. For large sample sizes, the most powerful tests are N2, Energy, HZ, and LV.

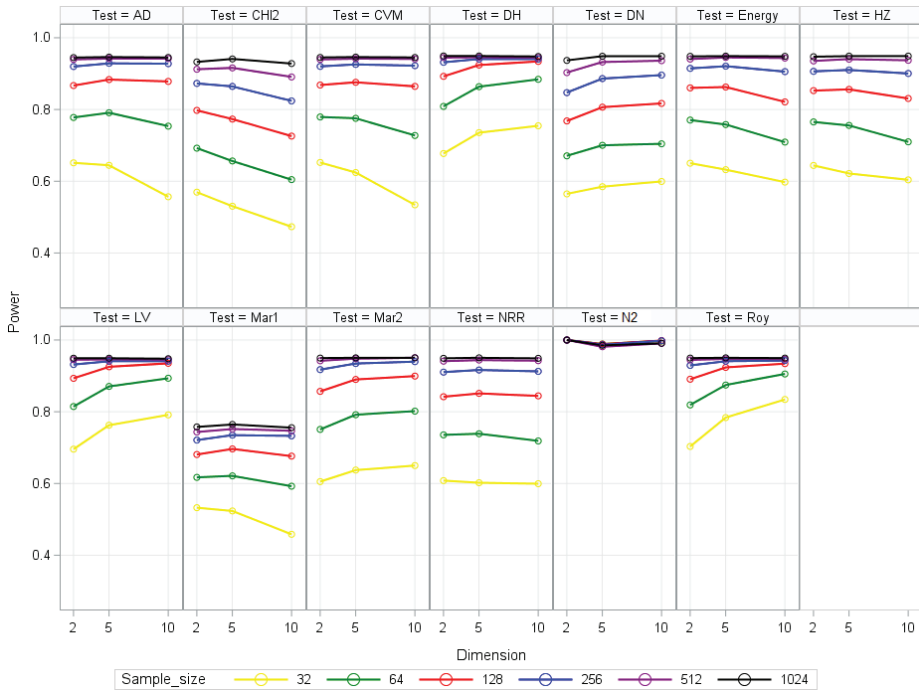


Figure 3.4. A summary of the average empirical power of all examined distribution groups by sample size and dimensionality.

3.3. Practical application to real data

The data set collected from 2001 to 2020 by the Head of the Department of Urology Clinic of the Lithuanian University of Health Sciences [113] illustrates the practical application. This dataset consists of study data from 2,423 patients and two different continuous attributes (patient age and prostate-specific antigen (PSA)). The assumption of normality was verified by filtering patients' age and PSA by year of death (i.e. deaths during the first 1, 2, 3, 4, 5, 6, 7, 10, and 15 years). The filtered data were standardized. The power and p -value were calculated for the multivariate tests. The significance level of $\alpha = 0.05$ was used for the study. Based on the obtained results, it was found that all the applied multivariate tests rejected the H_0 hypothesis of normality. The power of tests CHI2, DH, Energy, HZ, LV, New, Mar, NRR, and Roy was 0.999 and the p -value was <0.0001 , except for the DN test, the power of which was 0.576 and the p -value was 0.026.

In 2017, the data set of pollution research with IQOS and traditional cigarettes [114] was used by the Faculty of Chemical Technology and Department of Environmental Technology at Kaunas University of Technology for practical application. This data set consists of 33 experiments (with different conditions) in which the numerical (Pn10) and mass concentrations (Pm2.5, Pm10) of particles were measured. The assumption of normality was checked by filtering Pn10, Pm2.5, and Pm10 according to the number of the experiment in the smoking phase. The filtered data were standardized. The power and p -values of multivariate tests with a significance level of $\alpha = 0.05$ were calculated. Based on the obtained results, all the applied multivariate tests show that the assumption of normality is rejected. Most of the multivariate tests used (CHI2, DH, Energy, HZ, LV, New, Mar, NRR, and Roy) had a power of 0.999 and p -value of <0.0001 . The power of the other tests was also close to 0.99 and the p -value was about 0.0001.

CONCLUSIONS AND RECOMMENDATIONS

1. Problems of distributed computations, sample size, distributions, and tests selection were examined for testing the goodness-of-fit hypothesis. According to the review, samples of different sizes were selected for the research, different distributions of data covering all the most common problems in practice, the most powerful tests for testing the goodness-of-fit hypothesis were applied.
2. According to the theory of N-metric, a goodness-of-fit test for normality was proposed, which is a powerful competitor to other popular univariate tests. Furthermore, the proposed test is the most appropriate for symmetric data sets of size $n \geq 112$, asymmetric data sets of size $n \geq 118$, and bell-shaped distributed data sets of size $n \geq 88$.
3. The calibration of the parameters of the new kernel function allowed to choose its form which reduced the evaluation bias to near-optimal using the proposed N-metric based goodness-of-fit test.
4. New multivariate goodness of fit test for normality was constructed based on the evaluation of the difference in distribution densities and the application of

the inversion formula. The multivariate goodness-of-fit test showed that the constructed test for moderate-sized samples is best used for 2-dimensional data. It is recommended to use this test with large sample sizes ($n \geq 512$) for larger dimensions.

5. Chi-square, Doornick-Hansen, Energy, Henze-Zircler, Lobato-Velasco, Nickelin-Rao-Robson, Royston, and the new tests are most powerful for real data used in this study.

LITERATŪROS ŠALTINIAI

1. Pearson, K. (1900). On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. *Philosophical Magazine Series* 5, 50, 157-175.
2. Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari.* 4, 83–91.
3. Anderson, T. W., Darling, D. A. (1952) Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics.* 23, 193–212.
4. Anderson, T. (1962). On the Distribution of the Two-Sample Cramér-von Mises Criterion. *The Annals of Mathematical Statistics*, 33(3), 1148-1159.
5. Bakshaev A. (2009). Goodness of fit and homogeneity tests on the basis of N-distances. *Journal of Statistical Planning and Inference*, 510 139, 3750-3758.
6. Bakshaev A. (2010). Statistical tests based on N-distances. (Disertacija). *TECHNIKA*, 29-125.
7. Klebanov, L. (2005). *N-distances and their applications*. Karolinum, Prague.
8. Kavaliauskas M., Rudzkiš R., Ruzgas T. (2004) The Projection-based Multivariate Distribution Density Estimation, *Acta et Commentationes Universitatis Tartuensis de Mathematica.* 8, p. 135–141.
9. Ruzgas T., the nonparametric estimation of multivariate distribution density applying clustering procedures, PhD thesis, 2007, 161.
10. Hernandez H. (2021). Testing for Normality-Whatis the Best Method. *ForsChem Research.* 6, 1-38.
11. Szentesi S. G., Cuc L. D., Feher A., Cuc P. N. (2021) Does COVID-19 Affect Safety and Security Perception in the Hospitality Industry? A Romanian Case Study. *Sustainability*, 13(20), P. 1-20.
12. Gričar S., Štefan B. (2021). Technical Analysis of Tourism Price Process in the Eurozone. *Journal of Risk and Financial Management* 14(517), 1-25.
13. Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
14. Forsythe, C. J., & Compass, B. E. (1987). Interaction of cognitive appraisals of stressful events and coping: Testing the goodness of fit hypothesis. *Cognitive Therapy and Research*, 11, 473–485.
15. Sivam S.P.S.S., Loganathan G.B., Saravanan K., Dinesh Guhan S., Banerjee A. (2021) Effects of Drilling Process Parameters Using ANOVA and Graphical Methods. *Advances in Materials Research.* 5, 331-353.
16. Khanmohammadi Khorrami, MR, Shokri Aghbolagh, Z. (2020) Synthesis and non-parametric evaluation studies on high performance of catalytic oxidation-extraction desulfurization of gasoline using the novel

- TBAPW₁₁Zn@TiO₂@PAni nanocomposite. *Appl Organometal Chem.* 34, 1-14.
17. Tsagris M., Alenazi A., Stewart C. (2021) Non-parametric regression models for compositional data. *arXiv.* 3, 1-23.
 18. Ghafil W., Shamkhi R. (2021) Parametric regression analysis of bivariate the proportional hazards model with current status data. *International Journal of Nonlinear Analysis and Applications.* 12(2), 1591-1598.
 19. Hancock, J.T., Khoshgoftaar, T.M. (2020) Survey on categorical data for neural networks. *J Big Data.* 7(28), 1-41.
 20. Hu J., Shen L., Albanie S., Sun G., Vedaldi A. (2019) Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks. *arXiv.* 3, 1-14.
 21. Steele, Michael C. (2003). The Power of Categorical Goodness-Of-Fit Statistics (Disertacija). 9-94.
 22. Ahmad F., Sherwani R. A. K. (2015). Power Comparison of Various Normality Tests. *Pakistan Journal of Statistics and Operation Research.* 11(3), 331.
 23. Andersson J., Burberg M. (2015). Testing for Normality of Censored Data. 8-13.
 24. Brys G., Hubert M., Struyf A. (2008). Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics.* 23, 429–442.
 25. Desgagné, A.; Lafaye de Micheaux P. (2017) A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family. *Journal of Applied Statistics.* 45(13), 2307-2327.
 26. Friday I. A., Runyi E. F. (2018) Comparison of Goodness of Fit Tests for Normal Distribution. *Asian Journal of Probability and Statistics.* 1(2): 1-32.
 27. Górecki T., Horváth L., Kokoszka P. (2020). Tests of Normality of Functional Data. *International Statistical Review.* 88(3), 677–697.
 28. Romao, X.; Delgado, R. Costa A. (2010). An empirical power comparison of univariate goodness-of-fit tests of normality. *Journal of Statistical Computation and Simulation.* 80(5), 545–91.
 29. Adefisoye, J.; Golam Kibria, B. (2016) George, F. Performances of several univariate tests of normality: An empirical study. *J. Biom. Biostat.* 7(4), 1-8.
 30. Tianyi L. (2016) Power Comparison of Some Goodness-of-fit Tests (magistrinis darbas).
 31. Özonur D., Gokpinar E., Gokpinar, F., Bayrak H., Gul H. H. (2013). Comparisons of the Goodness of Fit Tests for the Geometric Distribution. *GUJ Sci.* 26(3), 369-375.
 32. Razali N. M., Wah Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics.* 2(1), 21-33.

33. Torabi H., Montazeri N. H., Grane A. (2016). A test for normality based on the empirical distribution function. *SORT*. 40 (1), 3-36.
34. Zhang J., Wu Y. (2005). Likelihood-ratio tests for normality. *Computational Statistics & Data Analysis*. 49, 709 – 721.
35. Choulakian, V., Lockhart, R.A. and Stephens, M.A. (1994) Cramer-von Mises Test of Discrete Distribution. *The Canadian Journal of Statistics*, 22, 125-137.
36. Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*. 52 (3–4), 591–611.
37. Moore D.S., Stubblebine J.B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Commun. Stat.-Theory and Methods*. A10(8), 713–738.
38. Doornik J. A., Hansen H. (2008). An Omnibus Test for Univariate and Multivariate Normality. *OXFORD BULLETIN OF ECONOMICS AND STATISTICS*. 70, 927-939.
39. Henze N., Zirkler B. (1990). A class of invariant consistent tests for multivariate normality. *COMMUN. STATIST.-THEORY METH*. 19(10), 3595-3617.
40. Henze N. (2002). Invariant tests for multivariate normality: A critical review. *Stat*. 43, 467–506.
41. Kesemen O., Tiryaki B. K., Tezel Ö., Özkul, E. (2021). A new goodness of fit test for multivariate normality. *Hacet. J. Math. Stat*. 50(3), 872-894.
42. Korkmaz S., Goksuluk D., Zararsiz G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, 6(2), 151-162.
43. Pinto L.P., Mingoti S.A. (2015). On hypothesis tests for covariance matrices under multivariate normality. *Pesquisa Operacional*. 35(1), 123-142.
44. Quiroz A. J., Dudley R. M. (1991). Some new tests for multivariate normality. *Probab. Th. Rel. Fields*. 87, 521-546.
45. Royston J. P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*. 31(2), 115–124.
46. Zhou M., Shao Y. (2014). A powerful test for multivariate normality. *Journal of Applied Statistics*. 41(2), 351-363.
47. Krauczi E. O. (2015). Goodness of fit to family of distribution. *Gradus*. 2(5), 52-62.
48. Krauczi E. O. (2016). Goodness of fit to the family of distribution (Disertacija).
49. Foudjo A. N. (2013). Robust normality test and robust power (Disertacija). 11-77.
50. Steele M., Chaseling J. (2005). Simulated Power of the Discrete Cramer-von Mises Goodness-of-Fit Tests. *Communications in Statistics - Simulation and Computation*. 35(4), 1067-1075.
51. Steele M., Hurst C., Chaseling J. (2007). Simulated Power of Discrete Goodness-of-Fit Tests for Likert Type Data (Proceedings Paper).

- International Congress on Modelling and Simulation (MODSIM07)*. 954-957.
52. Gnanadesikan R., Kettenring J. R. (1972). Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. *BIOMETRICS* 28, 81-124.
 53. Olivares A M., Forero C. G. (2010). Goodness-of-Fit Testing. *International Encyclopedia of Education*, 7, 190-196.
 54. Bahadur, R. (1960). Stochastic comparison of tests. *The Annals of Statistics*, 31(2), 276–295.
 55. Nikitin, Y. (1995). Asymptotic Efficiency of Nonparametric Tests. *Cambridge University Press*, New York.
 56. Lafaye de Micheaux P., Tran V. A. (2016). PoweR: A Reproducible Research Tool to Ease Monte Carlo Power Simulation Studies for Goodness-of-fit Tests in R. *Journal of Statistical Software*, 69(3), 1-42.
 57. Steele M., Chaseling J. (2006). Powers of Discrete Goodness-of-Fit Test Statistics for a Uniform Null Against a Selection of Alternative Distributions. *Communications in Statistics - Simulation and Computation*. 35(4), 1067-1075.
 58. Zhang J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *J. Roy. Statist. Soc.*, 64, 281-294
 59. Stephens M.A. (1974). EDF statistics for goodness-of-fit and some comparisons. *J. Amer. Statist. Assoc.*, 69, 730-737.
 60. Dyer A.R. (1974). Comparisons of tests for normality with a cautionary note. *Biometrika*, 61, 185-189.
 61. Zhang J. (2001). Powerful goodness-of-fit and multi-sample tests. Ph.D. Thesis, York University.
 62. Farrel, P.J., Stewart, K.R. (2006). Comprehensive Study Of Tests For Normality And Symmetry: Extending The Spiegelhalter Test. *Journal of Statistical Computation and Simulation*, 76(9), 803–816.
 63. Mendes, M. and Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Pakistan Journal of Information and Technology*, 2(2), 135-139.
 64. Keskin, S. (2006). Comparison of Several Univariate Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation Based Small Samples. *Journal of Applied Science Research*, 2(5), 296-300.
 65. Henze N., Wagner T. (1997), A new approach to the class of BHEP tests for multivariate normality, *J. Multiv. Anal.*, 62, 1–23.
 66. Voinov V., Pya N., Makarov R., Voinov Y. (2016). New invariant and consistent chi-squared type goodness-of-fit tests for multivariate normality and a related comparative simulation study. *Communications in Statistics - Theory and Methods*. 45(11), 3249-3263.
 67. Szekely G.J., Rizzo M.L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93, 58–80.

68. Enomoto R., Hanusz Z., Hara A., Seo T. (2020). Multivariate normality test using normalizing transformation for Mardia's multivariate kurtosis. *Communications in Statistics - Simulation and Computation*, 49(3), 684-698.
69. Song, Y., Zhao, X. (2021). Normality Testing of High-Dimensional Data Based on Principle Component and Jarque–Bera Statistics. *Stats*, 4, 216-227.
70. Dörr P., Ebner B., Henze N. (2021). A new test of multivariate normality by a double estimation in a characterizing PDE. *Metrika*, 84, 401-427
71. Dörr P., Ebner B., Henze N. (2020). Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. *Scand J Stat.* 1–34.
72. Shenton, L. R. and Bowman, K. O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and b_2 . *Journal of the American Statistical Association*, 72, 206-211.
73. McCulloch, C.E. (1985). Relationships among some chi-square goodness of fit statistics. *Commun. Stat.—Theory and Methods*, 14, 593–603.
74. Székely, G.J., Rizzo, M.L. (2005). A new test for multivariate normality. *J. Multivariate Anal.* 93, 58–80.
75. Mardia K.V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57, 519–530.
76. Thode H.C. Jr., Testing for Normality, Marcel Dekker, Inc., NewYork, 2002.
77. Small N. (1980). *Marginal skewness and kurtosis in testing multivariate normality*, *Appl. Stat.* 29, 85–87.
78. Timm N.H. *Applied Multivariate Analysis*, Springer, NewYork, 2002.
79. Saculinggan, M., Balase, E. A. (2013). Empirical Power Comparison of Goodness of Fit Tests for Normality In The Presence of Outliers. *Journal of Physics: Conference Series*, 435.
80. Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, 62, 399–402.
81. Glen, A. G., Leemis, L. M., Barr, D. R. (2001). Order statistics in goodness-of-fit testing. *IEEE Transactions on Reliability*, 50(2), 209–213.
82. D'Agostino R., Stephens M. *Goodness-of-fit Techniques*, Marcel Dekker, Inc., NewYork, 1986.
83. Filliben J.J. (1975). The Probability Plot Correlation Coefficient Test for Normality. *Technometrics*, 17(1), 111–117.
84. D'Agostino R.B. (1971). An Omnibus Test of Normality for Moderate and Large Size Samples. *Biometrika*, 58, 341–348.
85. Shapiro S. S., Francia R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67 215–216.

86. Chen L, Shapiro S.S. (1995). An Alternative Test for Normality Based on Normalized Spacings. *Journal of Statistical Computation and Simulation*, 53, 269–288.
87. Rahman MM, Govindarajulu Z (1997). A Modification of the Test of Shapiro and Wilk for Normality. *Journal of Applied Statistics*, 24(2), 219–235.
88. Zhang P. (1999). Omnibus Test of Normality Using the Q Statistic. *Journal of Applied Statistics*, 26(4), 519–528.
89. Hwang Y. T., Wei P. F. (2007). A Remark on the Zhang omnibus test for normality, *J. Appl. Stat.* 34(2), 177–184.
90. Barrio E., Cuesta, J.A., Albertos, Matrán C., Rodríguez J.M. (1999). Tests of goodness of fit based on the L2-Wasserstein distance, *Ann. Stat.* 27(4), 1230–1239.
91. Coin D. (2008). A goodness-of-fit test for normality based on polynomial regression, *Comput. Stat. Data Anal.* 52(4), 2185–2198.
92. D'Agostino, R. B., Pearson, E. S. (1973). Testing for departures from normality. I. Fuller empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, 60, 613–622.
93. Jarque C.M., Bera A.K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review. Revue Internationale de Statistique*, 55(2), 163–172.
94. Hosking, J. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1), 105–124.
95. Cabaña, A., Cabaña, E. M. (1994). Goodness-of-fit and comparison tests of the Kolmogorov-Smirnov type for bivariate populations. *Annals of Statistics*, 22, 1447–1459
96. Cabaña, A., Cabaña, E. M. (2003). Tests of normality based on transformed empirical processes. *Methodol. Comput. Appl. Probab.*, 5(3): 309–335.
97. Bonett, D.G.; Seier, E. (2002). A test of normality with high uniform power. *Comput. Stat. Data Anal.* 40, 435–445.
98. Bontemps C., Meddahi N. (2005). Testing Normality: A GMM Approach. *Journal of Econometrics*, 124(1), 149–186.
99. Gel, Y. R., Miao, W., Gastwirth, J. L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, 51(5), 2734–2746.
100. Martinez J., Iglewicz B. (1981). A test for departure from normality based on a biweight estimator of scale, *Biometrika*, 68(1), 331–333.
101. Epps T.W., Pulley L.B. (1983). A test for normality based on the empirical characteristic function, *Biometrika* 70(3), 723–726.
102. Ross G. J. S. MLP, Maximum Likelihood Program. *Harpenden: Rothamsted Experimental Station*, 1980.
103. Lobato I., Velasco C. (2004). A simple Test of Normality for Time Series. *Econometric Theory*, 20, 671–689.

104. Szekely G. J., Rizzo M. L. (2013). Energy statistics: A class of statistics based on distances, *Journal of Statistical Planning and Inference*, 143, 1249–1272.
105. Marron J.S., Wand M.P. (1992) Exact mMean Integrated Squared Error, *Annals of Statistics*, 20, p. 712–736.
106. Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Translation. *Biometrika*. 36 (1/2), p.149–176
107. Wicksell, S. D. (1917). *Ark. Mat. Astr. Fys.* 12(20).
108. Hill T., Lewicki P. (2007). Statistics methods and applications. *StatSoft*, Tulsa, OK. 2007.
109. Kasiulevičius V., Denapienė G. (2008). Statistikos taikymas mokslinių tyrimų analizėje. *Gerontologija*, 9(3), p. 176–180
110. Damianou C., Kemp, A. W. (1990). New goodness of statistics for discrete and continuous data. *American Journal of Mathematical and Management Sciences*, 10, p. 275–307.
111. Kaufman M. Gpu-Based Parallel Implementation of Swarm Intelligence Algorithms. 2016, p. 145-165.
112. Dzemyda G., Kurasova O., Žilinskas J. Daugiamačių duomenų vizualizavimo metodai. 2008, p. 37-38.
113. Milonas D., Ruzgas T., Venclovas Z., Jievaltas M., Joniau S. (2021). The significance of prostate specific antigen persistence in prostate cancer risk groups on long-term oncological outcomes. *Cancers*. 13(10), p. 2453-2466.
114. Martuzevicius D., Prasauskas T., Setyan A., O'Connell G., Cahours X., Julien R., Colard S. (2019). Characterization of the Spatial and Temporal Dispersion Differences Between Exhaled E-Cigarette Mist and Cigarette Smoke. *Nicotine & Tobacco Research*, 21(10), p. 1371–1377.
115. Weisberg S., Bingham C. (1975) An approximate analysis of variance test for non-normality suitable for machine calculation. *Technometrics*, 17(1), p. 133–4.
116. Elamir E. A. H., Seheult A. H. (2003). Trimmed L-moments. *Computational Statistics & Data Analysis*, 43(3), p. 299 – 314.
117. Vaičiulytė I. (2014). Markovo grandinės Monte-Karlo metodo tyrimas ir taikymas. (Disertacija), p. 24-26.
118. Stepanauskas G. (2001-2015). Monte Carlo metodas. *Visuotinė lietuvių enciklopedija*. Prieiga per: <https://www.vle.lt/straipsnis/monte-carlo-metodas/>
119. Steen Maarten van, Tanenbaum A.S., Distributed Systems, 3rd ed., distributed-systems.net. 2017.
120. Avdović A., Jevremović V. (2022) Quantile-Zone Based Approach to Normality Testing. *Mathematics*, 10(11), 1828.
121. Haj Ahmad, H.; Almetwally, E.M. (2022) Generating Optimal Discrete Analogue of the Generalized Pareto Distribution under Bayesian Inference with Applications. *Symmetry*, 14, 1457.

122. Doz, D.; Felda, D.; Cotič, M. (2022) Combining Students' Grades and Achievements on the National Assessment of knowledge: A Fuzzy Logic Approach. *Axioms*, 11, 359

AUTORĖS GYVENIMO APRAŠYMAS



Asmeninė informacija

Jurgita Arnastauskaitė

1993-09-25

Kaunas, Lietuva

jurgita.arnastauskaite@ktu.lt

Studentų g. 50-211, Kaunas, Lietuva.

Išsilavinimas:

2000–2012 Kauno Viktoro Kuprevičiaus vidurinė mokykla (dabar progimnazija).

2012–2016 Kauno technologijos universitete, Matematikos ir gamtos mokslų fakultete, įgytas bakalauro (taikomoji matematika) laipsnis.

2016–2018 Kauno technologijos universitete, Matematikos ir gamtos mokslų fakultete, įgytas magistro (taikomoji matematika) laipsnis.

2018–2022 Kauno technologijos universitete, Matematikos ir gamtos mokslų fakultete, doktorantūros studijos (Informatika N 009).

Profesinė patirtis:

2018.06–2018.06 Laborantė Kauno technologijos universiteto Taikomosios matematikos katedroje.

2018.09–2020.06 Asistentė Kauno technologijos universiteto Kompiuterių katedroje.

2020.07–dabar Lektorė Kauno technologijos universiteto Kompiuterių katedroje.

Mokslinių interesų sritys:

Duomenų analitika ir modeliavimas, statistika, duomenų tyrimas socialiniuose tinkluose, ekonomikos ir sveikatos mokslų srityse, kompiuterių architektūra.

AUTORĖS MOKSLINIŲ PUBLIKACIJŲ SĄRAŠAS

Web of Science duomenų bazėje indeksuotuose leidiniuose su cituojamumo rodikliu

Tarptautinėse leidyklose

- P1. [S1; CH; OA] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas; Bražėnas, Mindaugas. An exhaustive power comparison of normality tests // Mathematics. Basel: MDPI. ISSN 2227-7390. 2021, vol. 9, iss. 7, art. no. 788, p. 1-20. DOI: 10.3390/math9070788. [Science Citation Index Expanded (Web of Science); Scopus; DOAJ] [IF: 2,258; AIF: 1,221; IF/AIF: 1,849; Q1 (2019, InCites JCR SCIE)] [CiteScore: 2,20; SNIP: 1,290; SJR: 0,495; Q1 (2020, Scopus Sources)] [M.kr.: N 009, N 001] [Indėlis: 0,334].
- P2. [S1; CH; OA] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas; Bražėnas, Mindaugas. A new goodness of fit test for multivariate normality and comparative simulation study // Mathematics. Basel : MDPI. ISSN 2227-7390. 2021, vol. 9, iss. 23, art. no. 3003, p. 1-20. DOI: 10.3390/math9233003. [Science Citation Index Expanded (Web of Science); Scopus; DOAJ] [IF: 2,258; AIF: 1,221; IF/AIF: 1,849; Q1 (2020, InCites JCR SCIE)] [CiteScore: 2,20; SNIP: 1,290; SJR: 0,495; Q1 (2020, Scopus Sources)] [M.kr.: N 009, N 001] [Indėlis: 0,334].

Nacionalinėse leidyklose

- P3. [S1; LT; OA] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas. Accuracy of nonparametric density estimation for univariate Gaussian mixture models: a comparative study // Mathematical modelling and analysis. Vilnius: VGTU press. ISSN 1392-6292. eISSN 1648-3510. 2020, vol. 25, no. 4, p. 622-641. DOI: 10.3846 /mma.2020.10505. [Science Citation Index Expanded (Web of Science); Scopus; Academic Search Complete] [IF: 0,957; AIF: 0,993; IF/AIF: 0,963; Q2 (2019, InCites JCR SCIE)] [CiteScore: 2,10; SNIP: 1,165; SJR: 0,491; Q2 (2020, Scopus Sources)] [M.kr.: N 009, N 001] [Indėlis: 0,500].

Kituose recenzuojamuose mokslo leidiniuose

(Periodiniuose leidiniuose, tęstiniuose ir vienkartinuose straipsnių rinkiniuose)

Tarptautinėse leidyklose

- P4. [S5; IN] Ruzgas, Tomas; **Arnastauskaitė, Jurgita**; Kižauskienė, Laura; Bražėnas, Mindaugas. Vector quantization for selecting the number of neurons // Journal of communications technology, electronics and computer science. [S.l.] : JCTECS. ISSN 2457-905X. 2021, iss. 34, p. 14-17. DOI: 10.22385/jctecs.v0i0.299. [M.kr.: N 009, N 001] [Indėlis: 0,250].

MOKSLINIŲ TYRIMŲ REZULTATŲ SKELBIMAS KONFERENCIJOSE

Kitos konferencijų tezės ir straipsniai nerecenzuojamoje konferencijų pranešimų medžiagoje

- K1. [T2; TR] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas; Bražėnas, Mindaugas. Power analysis of goodness of fit tests for normality // 9th (online) international conference on applied analysis and mathematical modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey: abstracts book /M. Bayram, A. Seçer (eds.). Istanbul: Biruni University, 2021. ISBN 9786056918148. p. 155. [M.kr.: N 009, N 001].
- K2. [T2; VU] Ruzgas, Tomas; **Arnastauskaitė, Jurgita**. Accuracy of nonparametric density estimation for univariate Gaussian mixture models and application on municipal solid waste // 11th international workshop on data analysis methods for software systems, Druskininkai, Lithuania, November 28-30, 2019 / Lithuanian Computer Society, Vilnius University Institute of Data Science and Digital Technologies, Lithuanian Academy of Sciences. Vilnius: Vilnius University, 2019. ISBN 9786090703243. eISBN 9786090703250. p. 68-69. [M.kr.: N 009, N 001].
- K3. [T2; EE] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas. Accuracy of nonparametric density estimation for univariate Gaussian mixture models: a comparative study // MMA2019: Mathematical modelling and analysis: 24th international conference, May 28–31, 2019, Tallinn, Estonia: abstracts. Tallinn: Tallinn University of Technology, 2019. ISBN 9789949834396. p. 4. [M.kr.: N 001].
- K4. [T2; PH] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas; Bražėnas, Mindaugas; Kraujalis, Vytautas. Power comparison of multivariate normality tests // Sab-atan 2021: International conference on mathematical sciences and applications, November 24-26, 2021, Baguio city, Philippines: book of abstracts / editors: P.S.P. Ignacio, W.V. Alangui. Baguio city : University of the Philippines Baguio. 2021, 2A3, p. 37. [M.kr.: N 009, N 001].
- K5. [T2; LT; OA] **Arnastauskaitė, Jurgita**; Ruzgas, Tomas; Bražėnas, Mindaugas. Power analysis of multivariate goodness of fit tests // DAMSS 2021: 12th conference on data analysis methods for software systems, Druskininkai, Lithuania, December 2–4, 2021 / Lithuanian computer society, Vilnius university Institute of data science and digital technologies, Lithuanian academy of sciences. Vilnius : Vilnius university press, 2021. ISBN 9786090706732. eISBN 9786090706749. p. 5. [M.kr.: N 001].

PADĖKA

Iš visos širdies norėčiau padėkoti savo disertacijos moksliniam vadovui gerbiamam doc. dr. Tomui Ruzgui. Jis nepaliaujamai dalinasi sukauptomis žiniomis ir patirtimi. Visu studijų laikotarpiu visapusiškai padėjo vykdant mokslinius tyrimus ir rašant mokslinius straipsnius bei disertaciją. Tariu didžiausią ačiū moksliniam vadovui už tai, kad visada manimi tikėjo ir palaikė, net ir sunkiausiomis studijų akimirkomis. Siekiant darbo tobulumo vadovas netausojo jėgų ir laiko mokslinių tyrimų aptarimams ir diskusijoms. Labiausiai esu dėkinga už T. Ruzgo begalinę kantrybę, išmintį, atsakingą ir atidų disertacijos skaitymą. Jūsų išsakytos pastabos padėjo pagerinti disertacijos kokybę. Be jūsų palaikymo, padrąsinimo, pagalbos ir supratimo doktorantūros studijos ir disertacijos rašymas būtų buvęs sunkus procesas. Ačiū, kad mane pakentėte šitiek metų.

Taip pat noriu padėkoti Matematikos ir Gamtos mokslų fakulteto, Modeliavimo katedros lektoriui dr. Mindaugui Bražėnui. Ačiū, gerb. Mindaugai, už pagalbą rengiant mokslines publikacijas ir ruošiantis mokslinių rezultatų pristatymams. Jūsų pastebėtos klaidos ir pasiūlymai padėjo gerinti publikacijų kokybę. Aš esu dėkinga už naudingus pasiūlymus rengiant disertaciją. Jie tikrai padėjo lengvinti rašymo procesą.

Ir galiausiai norėčiau padėkoti Informatikos fakulteto, Kompiuterių katedros kolektyvui už pagalbą ir skatinimą tobulėti. Jūsų parama ir palaikymas man buvo labai svarbūs. Aš esu laiminga būdama tokio draugiško kolektyvo dalimi. Aš jaučiuosi dėkinga kolegei doc. dr. Laurai Kižauskienei už pagalbą rengiant ir koreguojant mokslines publikacijas bei ruošiantis mokslinėms konferencijoms. Taip pat jaučiuosi dėkinga prof. Jevgenijui Toldinui už pagalbą stengiantis įveikti doktorantūros studijų sunkumus. Ypač noriu padėkoti Gediminui Jakučiui ir Petruui Jakučiui už darbo vietos įrengimą ir pagalbą su kompiuterine technika.

Širdingai norėčiau padėkoti ir gerbiamiems recenzentams. Tikiuosi, šio darbo skaitymas nebus nuobodus ir varginantis.

Su didžiausia pagarba,
Jurgita Arnastauskaitė

PRIEDAI

1 priedas. Darbe naudoti skirstiniai

1p.1 lentelė. Simetriniai skirstiniai

Skirstinys	Parametrai	Forma
Normalusis: $N(\mu, \sigma)$	<ul style="list-style-type: none"> • Tankis: $\frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi}\sigma}$; • generavimas: $\sigma Z + \mu$; • dispersija: σ^2; • parametrai: $\mu = 0, \sigma = 1$. 	
Beta: $Beta(\alpha, \beta)$	<ul style="list-style-type: none"> • Tankis: $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 \leq x \leq 1$; • generavimas: $rbeta(\alpha, \beta)$; • parametrai: $\alpha = 0,5, \beta = 0,5$; $\alpha = 1, \beta = 1$; $\alpha = 2, \beta = 2$. 	
Koši: $Cauchy(l, s)$	<ul style="list-style-type: none"> • Tankis: $\frac{1}{s\pi\left(1+\left(\frac{x-l}{s}\right)^2\right)}$; • generavimas: $rcauchy(l, s)$; • parametrai: $l = 0, s = 0,5$; $l = 0, s = 1$; $l = 0, s = 2$. 	
Laplaso: $Lp(\mu, b)$	<ul style="list-style-type: none"> • Tankis: $\frac{1}{2b} \exp\left(-\frac{ x-\mu }{b}\right)$; • generavimas: $\mu - b \cdot \text{sgn}\{U - 0,5\} \ln(1 - 2 U - 0,5)$; • dispersija: $2b^2$; • parametrai: $\mu = 0, b = 1$. 	

Skirstinys	Parametrai	Forma
Logistinis: $Lg(\mu, s)$	<ul style="list-style-type: none"> Tankis: $\frac{1}{s} \exp\left(-\frac{x-\mu}{s}\right) \cdot \left(1 + \exp\left(-\frac{x-\mu}{s}\right)\right)^{-2}$; generavimas: $\mu + \ln\left(\frac{U}{1-U}\right)$; dispersija: $\frac{\pi^2}{3} s^2$; parametrai: $\mu = 2, s = 2$. 	
Stjudento: <i>Student</i> – $t(k)$	<ul style="list-style-type: none"> Tankis: $\frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k} \Gamma(k/2)} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}, k > 0$; Generavimas: $\frac{Z}{\sqrt{\chi^2(k)/k}}$; Dispersija: $k \leq 2: \infty, k > 2: \frac{k}{k-2}$; Parametrai: $k = 1, k = 2, k = 4, k = 10$. 	

1p.2 lentelė. Asimetriniai skirstiniai

Skirstinys	Parametrai	Forma
Beta: $Beta(\alpha, \beta)$	<ul style="list-style-type: none"> parametrai: $\alpha = 2, \beta = 1$; $\alpha = 2, \beta = 5$; $\alpha = 4, \beta = 0,5$; $\alpha = 5, \beta = 1$ 	
Chi-kvadrato: $\chi^2(k)$	<ul style="list-style-type: none"> Tankis: $2^{-k/2} \Gamma(k/2)^{-1} x^{k/2-1} \exp\left(-\frac{x}{2}\right), x > 0$; generavimas: $\sum_{i=1}^k Z_i^2$; dispersija: $2k$; parametrai $k = 1, k = 2, k = 4, k = 10$. 	

Skirstinys	Parametrai	Forma
Gama: $Gamma(a, b)$	<ul style="list-style-type: none"> Tankis: $\frac{1}{(1/b)\Gamma(a)} x^{a-1} \exp(-xb), x > 0, a, b > 0;$ generavimas: $rgamma(a, 1/b);$ dispersija: $a/b^2;$ parametrai: $a = 2, b = 2; a = 3, b = 2; a = 5, b = 1; a = 9, b = 1; a = 15, b = 1; a = 100, b = 1.$ 	
Gumbel: $Gumbel(\mu, \sigma)$	<ul style="list-style-type: none"> Tankis: $\frac{1}{\sigma} \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right] - \left(\frac{x-\mu}{\sigma}\right)\right\};$ generavimas: $\mu - \sigma \ln(E);$ dispersija: $\frac{\pi^2}{6} \sigma^2;$ parametrai: $\mu = 1, \sigma = 2.$ 	
Log-normalusis: $LN(\mu, \sigma)$	<ul style="list-style-type: none"> Tankis: $\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right);$ generavimas: $\exp\{N(\mu, \sigma^2);$ dispersija: $(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2);$ parametrai: $\mu = 0, \sigma = 1.$ 	
Veibulo: $W(\lambda, k)$	<ul style="list-style-type: none"> Tankis: $\frac{\lambda}{k} \left(\frac{x}{k}\right)^{\lambda-1} \exp\left(-\left(x/k\right)^\lambda\right);$ generavimas: $k(-\ln(U))^{1/\lambda};$ dispersija: $k^2\Gamma\left(1 + \frac{2}{\lambda}\right) - \mu^2;$ parametrai $\lambda = 0,5, k = 1; \lambda = 1, k = 2; \lambda = 2, k = 3,4; \lambda = 3, k = 4.$ 	

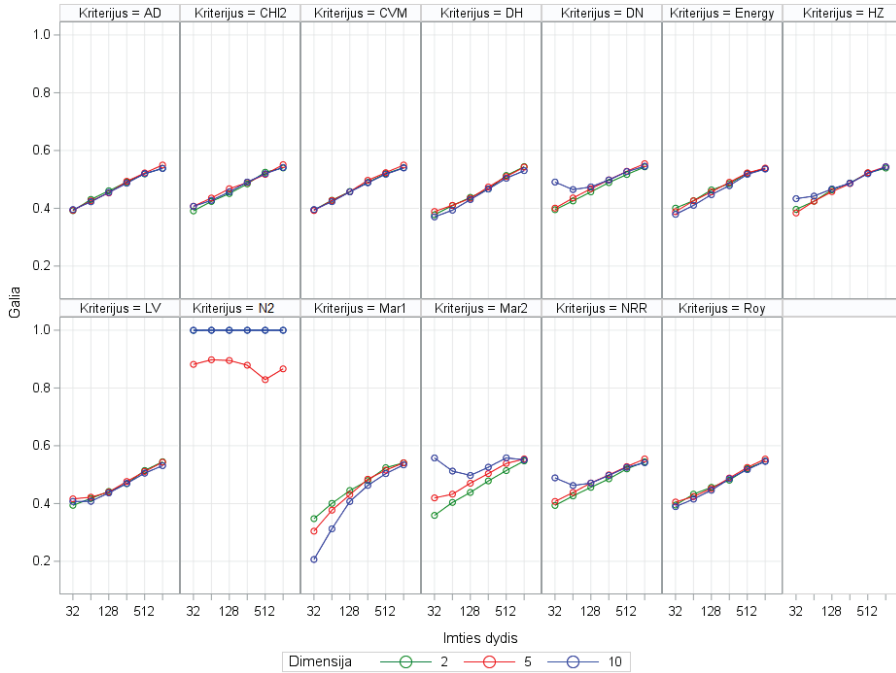
2 priedas. Naujo daugiamačio kriterijaus parametru rinkiniai

2p.1 lentelė. Glodumo parametro kitimas nuo triukšmo klasterio svorio (p_0) ir dimensijos (d)

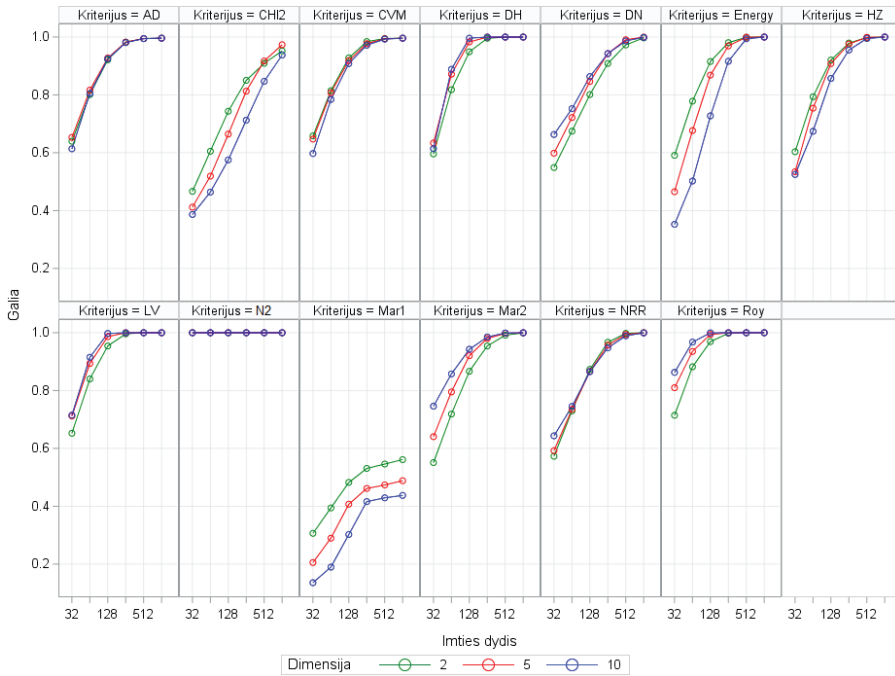
Dimensija (d)	Triukšmo klasterio svoris (p_0)	Glodumo parametras (h)	Vidutinė N2 kriterijaus statistikos (T^*) reikšmė
2	0	0,25	3,7850
2	0,001	0,70	3,2503
2	0,005	0,90	3,1316
2	0,010	1,00	3,0851
2	0,050	1,25	2,9792
2	0,100	1,50	2,9470
3	0	0,05	6,2800
3	0,001	0,05	6,3112
3	0,005	0,05	6,4146
3	0,010	0,05	6,5190
3	0,050	0,05	7,1085
3	0,100	0,05	6,7563
4	0	0,50	5,7014
4	0,001	1,00	5,3396
4	0,005	1,25	5,2694
4	0,010	1,50	5,2337
4	0,050	1,75	5,1912
4	0,100	2,00	5,1686
5	0	0,10	7,0022
5	0,001	0,10	7,0005
5	0,005	0,10	6,9945
5	0,010	0,25	6,9929
5	0,050	0,25	7,0044
5	0,100	0,25	7,0033
6	0	0,50	8,0353
6	0,001	1,25	7,6723
6	0,005	1,50	7,6390
6	0,010	1,75	7,6275
6	0,050	2,00	7,6083
6	0,100	2,20	7,6019
7	0	0,50	9,0344
7	0,001	0,50	9,0354
7	0,005	0,50	9,0397
7	0,010	0,50	9,0450

Dimensija (d)	Triukšmo klasterio svoris (ρ_0)	Glodumo parametras (h)	Vidutinė N2 kriterijaus statistikos (T^*) reikšmė
7	0,050	0,50	9,0889
7	0,100	0,50	9,1448
8	0	0,60	10,3732
8	0,001	1,50	10,1379
8	0,005	1,75	10,1243
8	0,010	2,00	10,1203
8	0,050	2,30	10,1139
8	0,100	2,50	10,1120
9	0	0,80	11,4251
9	0,001	0,80	11,4262
9	0,005	0,80	11,4306
9	0,010	0,75	11,4361
9	0,050	0,60	11,5178
9	0,100	0,50	11,7495
10	0	0,70	12,7642
10	0,001	1,75	12,6331
10	0,005	2,00	12,6285
10	0,010	2,20	12,6271
10	0,050	2,50	12,6252
10	0,100	2,75	12,6247

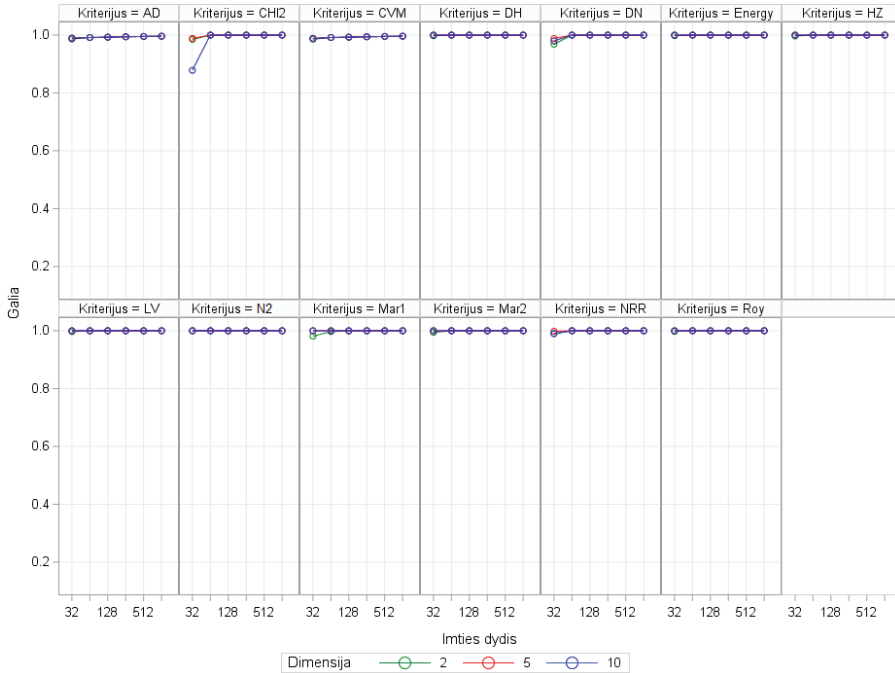
3 priedas. Daugiamačių kriterijų galios palyginimo grafikai pagal skirstinį Simetriniai skirstiniai



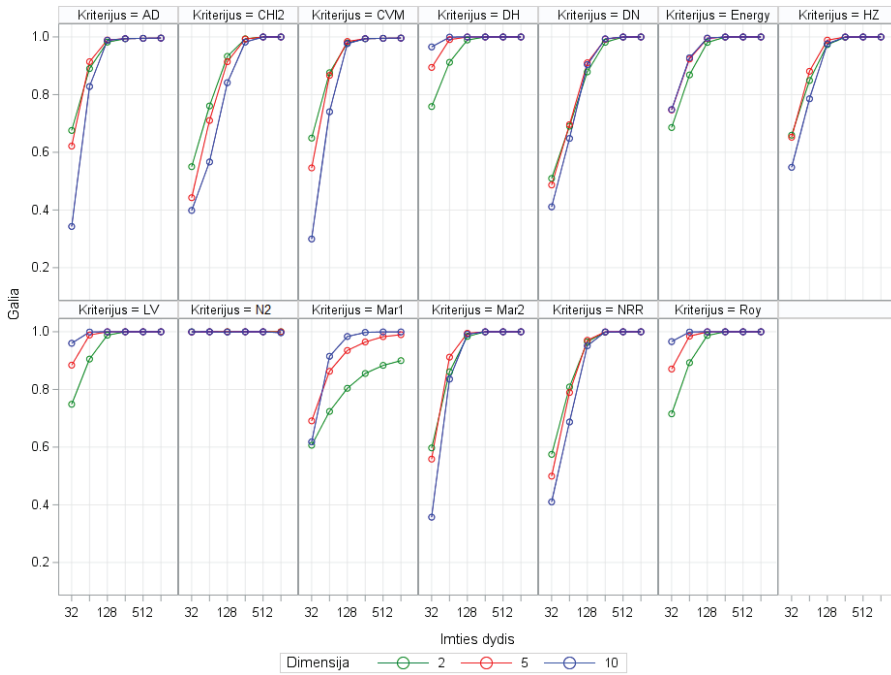
3p.1 pav. Daugiamačių kriterijų galios kitimas standartinio normaliojo skirstinio atveju



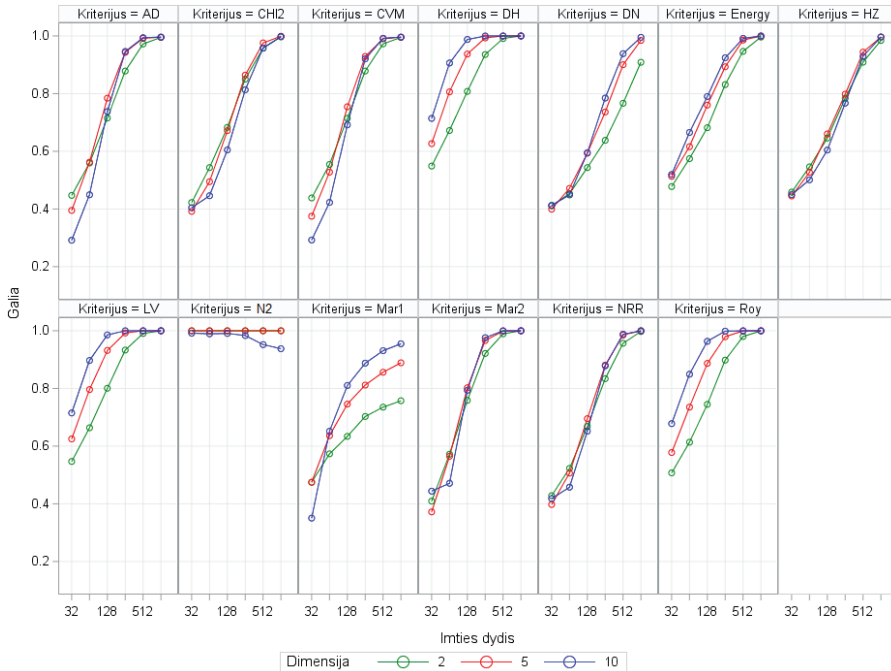
3p.2 pav. Daugiamačių kriterijų galios kitimas beta skirstinio atveju



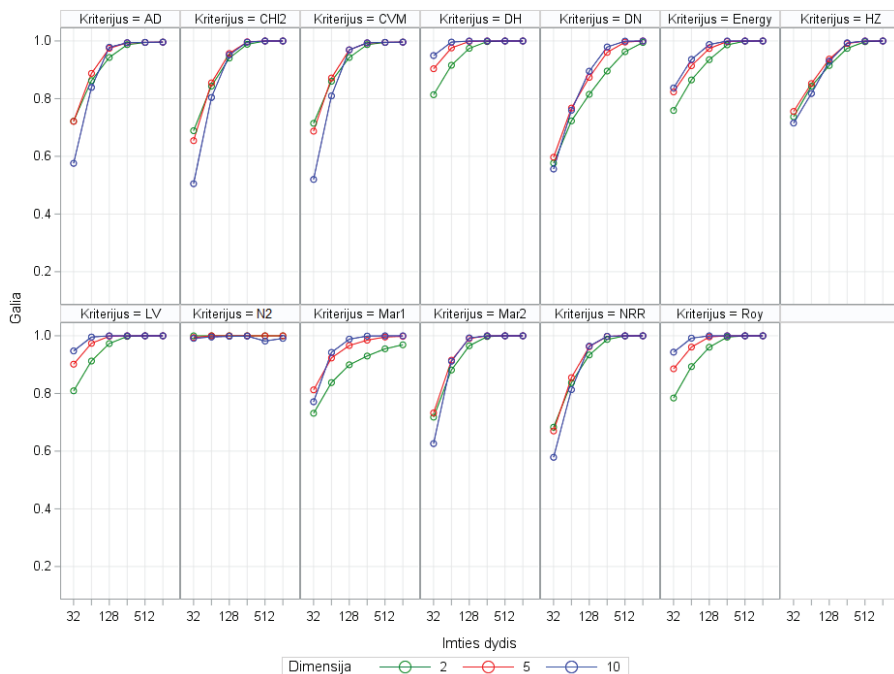
3p.3 pav. Daugiamačių kriterijų galios kitimas Koši skirstinio atveju



3p.4 pav. Daugiamačių kriterijų galios kitimas Laplaso skirstinio atveju

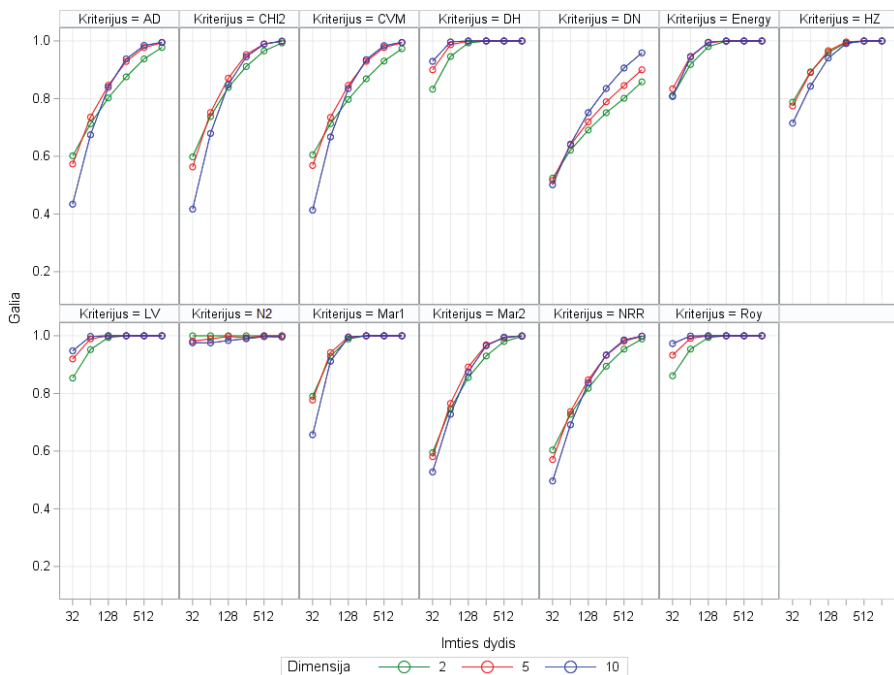


3p.5 pav. Daugiamačių kriterijų galios kitimas logistinio skirstinio atveju

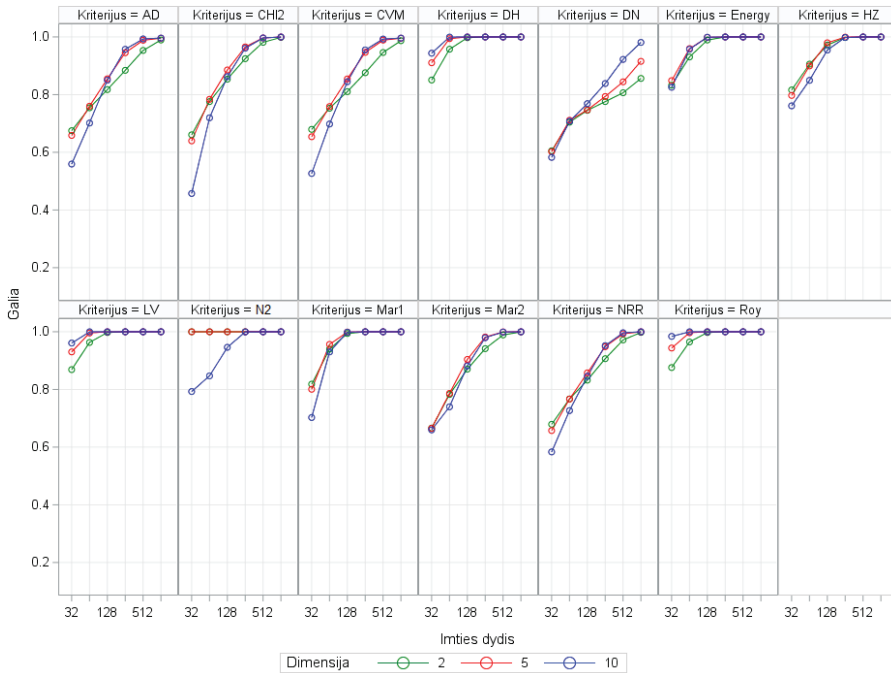


3p.6 pav. Daugiamačių kriterijų galios kitimas Stjudento skirstinio atveju

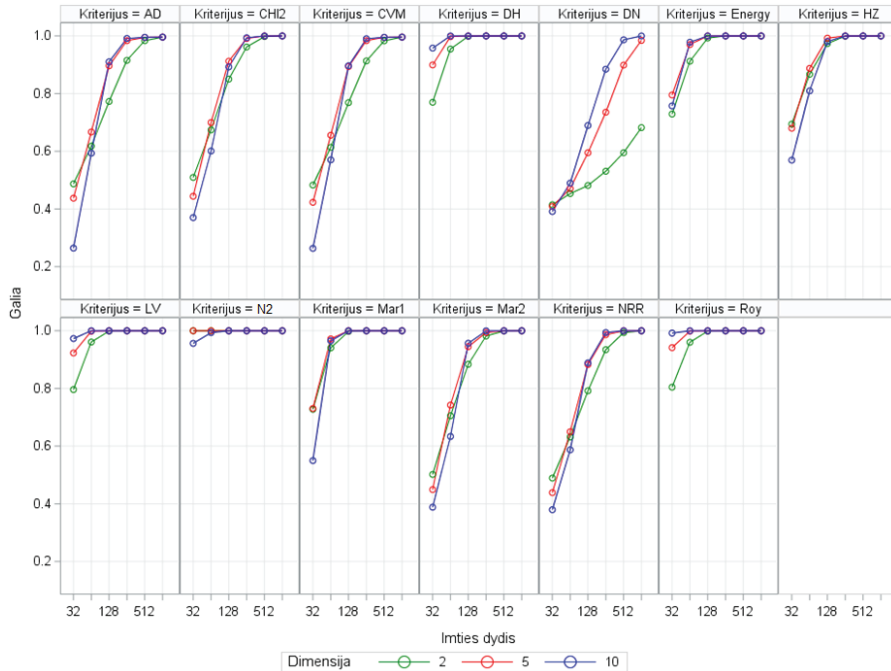
Asimetriniai skirstiniai



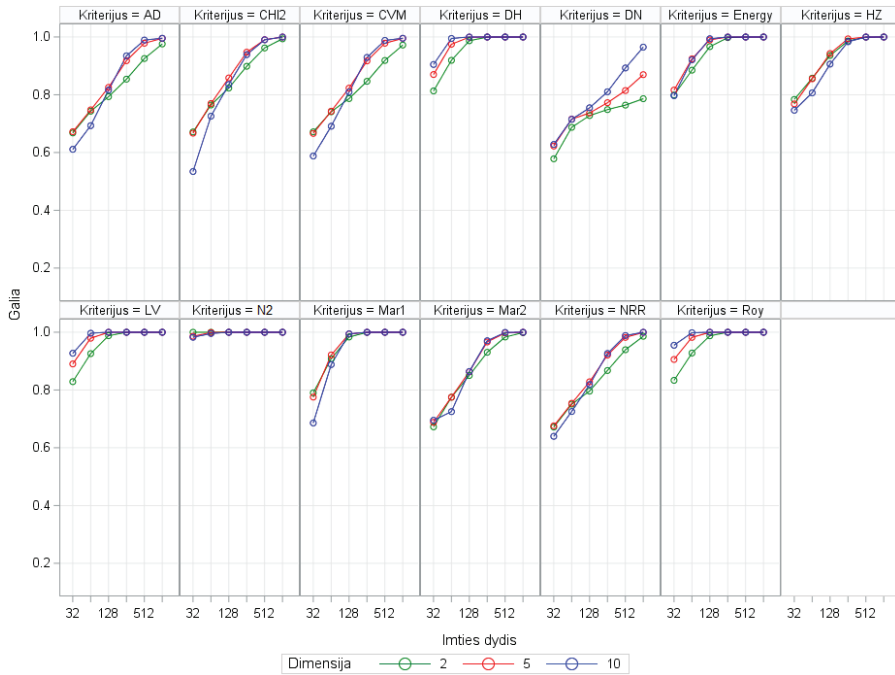
3p.7 pav. Daugiamačių kriterijų galios kitimas chi-kvadrato skirstinio atveju



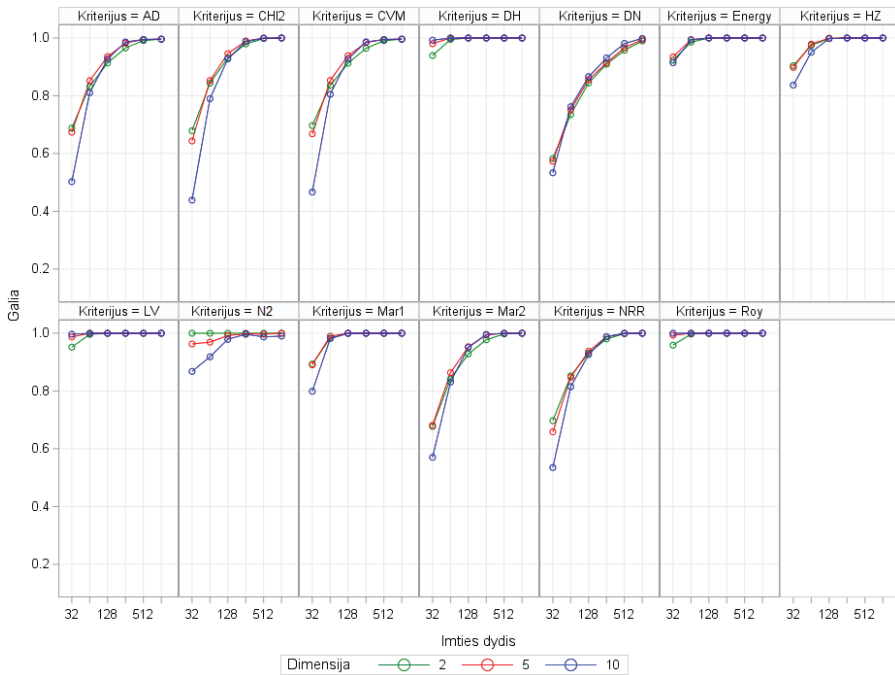
3p.8 pav. Daugiamačių kriterijų galios kitimas gama skirstinio atveju



3p.9 pav. Daugiamačių kriterijų galios kitimas Gumbelio skirstinio atveju



3p.10 pav. Daugiamačių kriterijų galios kitimas lognormaliojo skirstinio atveju



3p.11 pav. Daugiamačių kriterijų galios kitimas Veibulo skirstinio atveju

4 priedas. IQOS ir tradicinių cigarečių išmetamų kietųjų dalelių eksperimentų sąlygos

4p.1 lentelė. Eksperimentų patalpos sąlygos

Eksperimento nr.	Vėdinimas, proc.	Atstumas, m	Savanorių skaičius	Drėgmė, proc.	Cigaretės tipas
N1	0,2	0,5	1	30	IQOS
N2	0,5	0,5	1	30	IQOS
N3	1	0,5	1	30	IQOS
N4	0,2	0,5	5	30	IQOS
N5	0,5	0,5	5	30	IQOS
N6	1	0,5	5	30	IQOS
N7	0,2	0,5	1	70	IQOS
N8	0,5	0,5	1	70	IQOS
N9	1	0,5	1	70	IQOS
N10	0,2	0,5	5	70	IQOS
N11	0,5	0,5	5	70	IQOS
N12	1	0,5	5	70	IQOS
N13	0,2	2	1	30	IQOS
N14	0,5	2	1	30	IQOS
N15	1	2	1	30	IQOS
N16	0,2	2	5	30	IQOS
N17	0,5	2	5	30	IQOS
N18	1	2	5	30	IQOS
N19	0,2	2	1	70	IQOS
N20	0,5	2	1	70	IQOS
N21	1	2	1	70	IQOS
N22	0,5	0,5	5	70	IQOS
N23	1	0,5	5	30	IQOS
N24	0,2	0,5	5	70	IQOS
N25–N27	0,5	1	3	50	IQOS
N28–N30	0,5	2	1	50	IQOS
N31–N33	0,5	2	1	50	Tradicinė cigaretė

UDK 519.245+004.942](043.3)

SL 344. 2022-08-24, 18,75 leidyb. apsk. 1. Tiražas 14 egz. Užsakymas 165.
Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas

