



Vilniaus universiteto matematikos ir informatikos institutas



Tikimybių teorijos ir statistikos skyrius

# Diskrečių struktūrų statistinės analizės metodai ir jų taikymai

Ataskaita už I-ąjį doktorantūros kursą

Doktorantė: Monika Lapėnaitė-Gedvilė

Vadovas: doc. dr. Marijus Radavičius

Doktorantūros pradžios ir pabaigos metai: 2014 - 2018

2015-10-13



# Tyrimo objektas



- Diskrečios struktūros (kokybiniai požymiai ir jų tarpusavio sąryšiai, atsitiktiniai grafai, lingvistinės struktūros ir t.t.);
- Statistiniai modeliai ir metodai šių struktūrų statistinei analizei.



# Artimiausi tikslai



- Susipažinti su objektu ir jo tyrimo problemomis;
- Statistikos metodus pritaikyti uždaviniams, kurie paprastai sprendžiami duomenų sijos metodais;
- Parengti atitinkamą metodiką ir ją pritaikyti realių duomenų statistiniame tyrime.



# Planuojami rezultatai



- 1 Iširtas statistinių metodų ir algoritmų efektyvumas
- 2 Parinktų metodų realizacija
- 3 Metodų palyginimas ir įvertinimas



# 2014/2015 m. darbo planas



- 1 Išlaikyti egzaminus:
  - Bajeso modelių taikymas statistikoje
  - Baigtinių populiacijų statistika
  - Statistiniai duomenų analizės pagrindai
- 2 Apžvelgti mokslinius tyrimus susijusius su disertacijos tema.



# Ataskaita už 2014/2015 mokslo metus



- 1 Išlaikytas 1 egzaminas
- 2 Apžvelgta mokslinė literatūra
- 3 Atliktas pranešimas konferencijoje
- 4 Parengta (priimta spausdinimui) mokslinė publikacija recenzuojamame žurnale

# Mokslinė literatūra 1



Šaltiniai, kuriuose pateikiami statistiniai modeliai ir metodai:

- 1 Agresti, A. Categorical Data Analysis, second edition, Canada, 2002.
- 2 Dobson, A., J. Introduction to generalized linear models, second edition, Boca Raton, London, New York, Washington A CRC Press Company, 2001.
- 3 Abney, S. Statistical methods and linguistics. In: The Balancing Act: Combining Symbolic and Statistical Approaches to Language, MIT Press, Cambridge, 1996, pp. 1–26.
- 4 Baayen, R., H. Word Frequency Distributions, Kluwer Academic Publishers, 2001.
- 5 Robert, C., The Bayesian Choice, 2nd edition, Springer, 2007.
- 6 Sarndal C. E., Swensson B., Wretman J., Model Assisted Survey Sampling, Springer, 1992.

## Mokslinė literatūra 2



Straipsniai susiję su kalba ir jos savybėmis:

- 1 Utka, A. Labai Dažnų Lietuvių Kalbos Žodžių ir Žodžio Formų Ypatybės. Kaunas: Vytauto Didžiojo universitetas, 2005
- 2 Kazlauskienė, A.; Rimkutė, E.; Utka, A. Kiekybiniai tyrimai kalbotyroje [žiūrėta 2015m. spalio 13 d.]. Prieiga per internetą: < [http : //donelaitis.vdu.lt/lkk/pdf/I\\_dalis.pdf](http://donelaitis.vdu.lt/lkk/pdf/I_dalis.pdf) >.
- 3 Brasiūnaitė J., Straipsnių antraščių tekstų heterogeniškumas, Žurnalistikos tyrimai, 3, p. 79-98, 2010
- 4 Dabašinskienė I., Šnekamosios lietuvių kalbos morfologinės ypatybės, Acta Linguistica Lithuanica, LX, p. 1-15, 2009
- 5 Leonavičienė A., Spaudos tekstų heterogeniškumas funkcinių stilių sandūros aspektu. Kalbotyra, 55 (3), p. 38-46, 2005





## **Pranešimas Lietuvos matematikų draugijos 56-ojoje konferencijoje tema "Žymeklių naudojimas vertinant tekstų statistines savybes"**

Nagrinėjamas **objektas**: rašytinė lietuvių kalba.

Naudojami **duomenys**: 30 panašaus žanro lietuviškų kūrinių, suskaitmenintų vykdant ES struktūrinių fondų remiamą projektą „Pagrindinio ugdymo pirmojo koncentro (5–8 kl.) mokinių esminių kompetencijų ugdymas“, 2012.

Pasirinkta tyrimui savybė – linksnis.



# Moksliniai rezultatai 1



## **Pranešimas Lietuvos matematikų draugijos 56-ojoje konferencijoje tema "Žymeklių naudojimas vertinant tekstų statistines savybes"**

Tyrimo **tikslas**: remiantis dažniau pasitaikančiomis žodžio formomis ar kitais dariniais (toliau juos vadinsime žymekliais), jų pasiskirstymu tekste, nuspėti kitų (retesnių) žodžio formų (ar kitų lingvistinių objektų) savybių pasiskirstymus.

Tyrimo pagrindinis **uždavinys**: patikrinti ar įmanoma pakankamai tiksliai (statistinių paklaidų tikslumu) prognozuoti linksnių ir jų kombinacijų pasitaikymo tekste dažnius pasirinktu žymeklių pagrindu.

# Moksliniai rezultatai 1



## **Pranešimas Lietuvos matematikų draugijos 56-ojoje konferencijoje tema "Žymeklių naudojimas vertinant tekstų statistines savybes"**

Darbe pasiūlyta metodika ir ja remiantis atliktas tyrimas leidžia padaryti tokias **išvadas**:

1. Tyrimas parodė, kad tinkamai parinkto žymeklių rinkinio ir statistinio modelio pagrindu galima gauti gana tikslus dominančios savybės (šiuo atveju – linksnių vartojimo dažnumo) pasiskirstymo įverčius.
2. Nors vieno ar kito linksnio vartojimo dažnis tarp autorių gali ženkliai skirtis, tačiau jo sąryšiai su žymeklių grupėmis, matyt, nuo autorių priklauso silpnai. Taigi, tie sąryšiai gali būti susieti su universaliosiomis lietuvių kalbos savybėmis.



## Moksliniai rezultatai 2



**Mokslinė publikacija (priimta spausdinimui) recenzuojamame el. leidinyje Lietuvos statistikos darbai 2015, 54 t., Nr. 1, ISSN2029-7262 tema "Tekstų nehomogeniškumo tyrimas naudojant žymeklius"**

Straipsnio [tikslas](#) įvertinti tekstų statistinį nehomogeniškumą pagal funkcinių žodžių ir kitų lingvistinių elementų vartoseną.



**Mokslinė publikacija (priimta spausdinimui) recenzuojamame el. leidinyje Lietuvos statistikos darbai 2015, 54 t., Nr. 1, ISSN2029-7262 tema "Tekstų nehomogeniškumo tyrimas naudojant žymeklius"**

Pastebėta, kad žymeklių dažnumai blokuose turi **ženklia perteklinę sklaidą** lyginant su lingvistikoje įprastu homogeniškumo modeliu. Pasirinktoms žymeklių grupėms parinkti kelių tipų hierarchiniai binominės logistinės regresijos modeliai, naudojantys autoriaus identifikatorių, bloko ilgį ir likusių žymeklių dažnius blokuose kaip aiškinančiuosius kintamuosius, leido paaiškinti ženklia dalį pasirinktųjų žymeklių perteklinės sklaidos.



# 2015/2016 m. darbo planas



Išlaikyti egzaminus:

- 1 Baigtinių populiacijų statistika
- 2 Statistiniai duomenų analizės pagrindai arba Taikomoji statistika
- 3 Algebra

Atlikti šiuos disertacijos rengimo etapus:

- 1 Disertacijos tikslo ir uždavinių formulavimo patikslinimas.
- 2 Tinkamos tyrimo metodikos sudarymas.
- 3 Teorinio ir empirinio tyrimo planavimas.
- 4 Teorinio tyrimo pradžia - diskrečiųjų struktūrų statistinių modelių savybių tyrimas.



# Šakinio padalinio išvada



2015 m. spalio 14 d. Tikimybių teorijos ir statistikos skyriuje vykusio svarstymo metu nutarta, kad doktorantė Monika Lapėnaitė-Gedvilė 2014/2015 m. m. darbo planą įvykdė nepilnai, nes buvo suplanuota išlaikyti 3 egzaminus, o buvo išlaikytas tik 1. Tačiau dalyvavo konferencijoje ir paruošė mokslinę publikaciją recenzuojamame leidinyje.



Ačiū už dėmesį!