

ŠIAULIŲ UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS KATEDRA

Robertas Mockus

Informatikos specialybės II kurso (neakivaizdinio skyriaus) studentas

**STATISTINIŲ ĮRANKIŲ, PAGRĮSTŲ POKYČIŲ
STATISTIKA, PAKETO KŪRIMAS**

(Creation of statistical tools pack based on increment ratio statistic)

MAGISTRO DARBAS

Darbo vadovas

Vyrens. moksl. darb. dr. M. Vaičiulis

Recenzantai

Lektorius dr. V. Sirius

Dr. A. Klivečka

Šiauliai, 2007/2008 m.m.

Turinys

I. Įvadas.....	3
II. Teorinė dalis.....	4
1. Temos analizė.....	8
2. Darbo srities analizė.....	9
3. Darbinės srities modelis.....	15
III. Projektinė dalis.....	15
1. Įrankių ir priemonių pasirinkimo analizė.....	15
2. Projekto (darbo) vykdymo planas.....	16
3. Pradinis projekto aprašymas.....	18
IV. Darbo eigos aprašymas.....	22
1. Darbų eigos grafas.....	22
2. Galutinio projekto būklės aprašymas.....	23
3. Problemų ir jų sprendimų aprašymai ir pagrindimai.....	25
4. Patarimai, pastebėjimai, rekomendacijos.....	26
5. Darbo rezultatų analizė.....	26
V. Išvados.....	30
VI. Literatūros ir informacinių šaltinių sąrašai.....	30
VII. Anotacija.....	31
VIII. Kompaktinio disko turinys.....	31
IX. Priedai.....	32

I. Įvadas

Sparčiai vystantis informacinėms technologijoms, kompiuteriai tampa vis spartesni, jais sėkmingai galime atlikti įvairius modeliavimus, bei skaičiavimus su dideliais informacijos kiekiais. Kompiuterių gebėjimas atlikti matematinius skaičiavimus su milžiniškais duomenų kiekiais per gana trumpą laiką suteikia galimybę sukurti programą, kuri būtų naudojama pokyčių sankykio (IR) statistikoje.

Darbo tikslas – sukurti programą, kuri padėtų tęsti atminties parametro d įverčio, pagrįsto pokyčių statistika, sąvybių tyrimą.

Programa turi būti patogi ir suprantama vartotojui, bei greitai atlikti sudėtingus matematinius skaičiavimus.

Reikia išspręsti šiuos uždavinius:

1. Pasirinkti tinkamą programavimo kalbą, kuri tenkintų visus keliamus reikalavimus (detaliau žiūrėti 15 psl).
2. Sukurti programą, kuri padėtų tęsti atminties parametro d įverčio, pagrįsto pokyčių statistika, sąvybių tyrimą.
3. Sukurti vartotojo sąsają, kuri vartotojui būtų paprasta ir nesudėtinga neprogramuojančiam vartotojui.
4. Sukurtą programą visapusiškai ištestuoti su įvairiais duomenimis, jų kiekiais ir patikrinti programos veikimo stabilumą su populiariausiomis operacinėmis sistemomis.

Analogų analizė

Panašių programų sukurta nėra. Kol kas yra gauti tik teoriniai, liečiantys IR statistiką rezultatai, o realiems finansiniams ir ekonominiams duomenims vartotojui draugiškos programos sukurta nėra.

II. Teorinė dalis

IR statistika

Vienas iš pagrindinių laiko eilučių teorijos uždavinių yra: *turint stebinius* x_1, x_2, \dots, x_n *atlikti prognozes* $\hat{x}_{n+1}, \hat{x}_{n+2}, \dots, \hat{x}_{n+k}$. Tai pakankamai sudėtingas ir nevienareikšmiškai sprendžiamas uždavinys, kurio sudėtinės dalys yra:

- 1) įvairaus pobūdžio trendų, ciklinių bei periodinių dedamųjų šalinimas;
- 2) modelio (stochastinės lygties) parinkimas;
- 3) parametrų pasirinktame modelyje vertinimas;
- 4) modelio atitikimo turimiems stebiniams tikrinimas;
- 5) prognozavimas taikant pasirinktą modelį.

Antrajame etape tyrėjui reikia nuspręsti, kokia atmintimi (ilga, trumpa ar neigiama) pasižymintį modelį pasirinkti. Yra žinomi keli atminties parametro (jis dažniausiai žymimas d) įvertinimo metodai. Priminsime diskretaus laiko proceso, pasižyminčio ilga atmintimi, apibrėžimą.

Apibrėžimas. Yra sakoma, kad stacionarus atsitiktinis procesas $X_t, t \in \mathbb{Z}$ pasižymi ilga atmintimi, jei egzistuoja tokia konstanta $c > 0$ ir parametras $|d| < 1/2$, kad spektrinė tankio funkcija $f(x)$ tenkina sąryšį $f(x)/(cx^{-2d}) \rightarrow 1$, kai $x \rightarrow 0_+$ (x artėja į nulį iš dešinės).

Platų ilgos atminties procesų aprašymą galima rasti monografijoje [1], taip pat žiūrėkite [4].

Populiariausi iš jų yra lokalusis Whittle'o ir bangelių metodas. Vardan dėstymo pilnumo juos pateikiame žemiau. Šiame darbe mes sutelksime dėmesį ties pačiu naujausiu, paremtu IR statistika, atminties parametro d įvertinimo metodu. Pastarasis yra įvestas ir teoriškai pagrįstas straipsnyje [3], toliau tyrinėtas straipsnyje [2]. Trumpai perteiksime metodo esmę. Tarkime, kad turime stebinius x_1, x_2, \dots, x_n . Pokyčių santykio statistika yra:

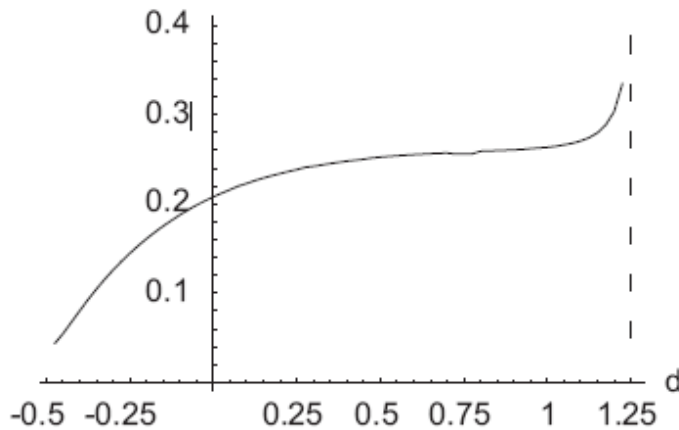
$$(1.1) \quad IR := \frac{1}{N - 3m} \sum_{k=0}^{N-3m-1} \frac{|\sum_{t=k+1}^{k+m} (X_{t+m} - X_t) + \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t)|}{|\sum_{t=k+1}^{k+m} (X_{t+m} - X_t)| + |\sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t)|},$$

čia $m \in \mathbb{N}$ yra taip vadinamas "lango" parametras, be to skaičiuojant šią statistiką galioja susitarimas, kad $\frac{0}{0} := 1$. Funkcija $\Lambda(d) = \Lambda_0(\rho(d))$, kur:

$$(1.2) \quad \Lambda_0(r) := \frac{2}{\pi} \arctan \left(\sqrt{\frac{1+r}{1-r}} \right) + \frac{1}{\pi} \left(\sqrt{\frac{1+r}{1-r}} \right) \log \left(\frac{2}{1+r} \right),$$

$$(1.3) \quad \rho(d) := \frac{-9^{d+0.5} + 4^{d+1.5} - 7}{2(4 - 4^{d+0.5})},$$

yra apibrėžta visiems realiems skaičiams $d \in (-\frac{1}{2}, \frac{5}{4})$, žiūrėkite jos grafiką pav.1.



1. pav. $\Lambda(d)$ grafikas

Straipsnyje [3] yra įrodyta, kad, jei stebiniai tenkina tam tikras "gausiškumo" sąlygas, tai:

$$(1.4) \quad \mathbf{E}IR \rightarrow \Lambda(d), \quad \mathbf{E}(IR - \Lambda(d))^2 \rightarrow 0,$$

kai $n \rightarrow \infty$ ir $m \rightarrow \infty$. Pirmasis (1.4) sąryšis reiškia, kad kai imtis IR sudaryta iš pakankamai didelio skaičiaus stebinių x_1, \dots, x_n , tai IR statistikos vidurkis $\mathbf{E}IR$ tenkina apytikslę lygybę $\mathbf{E}IR \approx \Lambda(d)$, o antrasis (1.4) sąryšis dar stipresnis, ir tvirtina kad minėta apytikslė lygybė (kai imtis x_1, \dots, x_n pakankamai didelė) tenkina ne tik IR statistikos vidurkis, bet ir pati IR statistika:

$$(1.5) \quad IR \approx \Lambda(d).$$

Vadinasi, norint įvertinti atminties parametą d , reikia duotai imčiai x_1, \dots, x_n apskaičiuoti IR statistiką bei skaitiškai išspręsti lygtį (1.5). Spręsdami šią lygtį remiamės funkcijos $\Lambda(d)$ monotoniniu didėjimu.

Detalus apytikslis lygties (1.5) sprendimas yra aptartas poskyriuose „Žingsnių skaičiavimas“, „Pirmojo žingsnio skaičiavimas“, „Antrojo žingsnio skaičiavimas“, „Aštuntojo žingsnio skaičiavimas“.

Lokalus Whittle statistinis įvertinimas

Apibrėžimas. Lokalus Whittle statistinis įvertinimas $\hat{\alpha}_{LW}$ buvo pasiūlytas Künsch (1987) ir toliau plėtotas Robinson (1995a). Jis yra apibūdinamas:

$$\hat{\alpha}_{LW} = \arg \min_{\alpha} G(\alpha, m) := \left\{ \ln \left(\frac{1}{m} \sum_{j=1}^m \frac{I_Y(\lambda_j)}{\lambda_j^{-\alpha}} \right) - \frac{\alpha}{m} \sum_{j=1}^m \ln(\lambda_j) \right\}$$

kur $I_Y(\lambda_j)$ yra periodograma, skaičiuojama parametro m Fourier dažniams $\lambda_j = \frac{2\pi j}{T}$, $j = 1, \dots, m \ll \lfloor \frac{T}{2} \rfloor$, kur $\lfloor \cdot \rfloor$ žymi sveiką dalį, lango parametras m krypsta į begalybę su duoto dydžio T , bet daug lėčiau nuo $\frac{1}{m} + \frac{m}{T} \rightarrow 0$ kai $T \rightarrow \infty$. Procesas neprivalo būti gausiškas, bet jų spektras yra diferencijuojamas netoli nulinio dažnio ir procesas turi judantį vidurkio vaizdavimą.

Charakteristikos. Lokalus Whittle statistinis įvertinimas turi asimptotinį pasiskirstymą:

$$\sqrt{m}(\hat{\alpha}_{LW} - \alpha) \xrightarrow{d} N(0, 1)$$

kur \xrightarrow{d} reiškia konvergavimą pagal pasiskirstymą.

Šis metodas yra išdėstytas [5].

Ilgos ir trumpos atminties generatoriai

Mes apibrėšime praktinius metodus, leidžiančius generuoti diskretaus laiko pačiu paprasčiausiu, pasižyminčiu ilga (trumpa) atmintimi procesu N ilgio trajektorijas.

FARIMA [0,d,0] ir AR(1) generavimo algoritmai kuriamoje programoje bus naudojami pradinių duomenų generavimui.

FARIMA [0,d,0] generatoriaus algoritmas

Algoritmas:

1. Apskaičiuoti seką:

$$a_k = a_{k-1} \cdot \frac{k-1+d}{k},$$

- reikšmė k kinta nuo 0 iki vartotojo pasirinktos
 - reikšmė d kinta nuo -0.5 iki 0.5
 - $a_0 = 1$.
2. Generuoti nepriklausomų standartinių normalių atsitiktinių dydžių seką η_{t-k} .
 - reikšmė k kinta nuo 0 iki vartotojo pasirinktos
 3. Skaičiuoti rezultatų seką:

$$X_t = \sum_{k=0}^{\infty} a_k \cdot \eta_{t-k},$$

Kadangi iki begalybės sumą skaičiuoti negalime, įsivedame nupjovimo aukščio parametą M ir aproksimuojame:

$$\tilde{X}_t = \sum_{k=0}^M a_k \cdot \eta_{t-k}$$

- reikšmė t kinta nuo 0 iki vartotojo pasirinktos
- reikšmė M kinta nuo 0 iki vartotojo pasirinktos

Šis metodas yra išdėstytas [6].

AR (1) generatoriaus algoritmas

Algoritmas:

AR(1) proceso trajektorijų generavimas

Stacionarus diskretaus laiko atsitiktinis procesas $X_t, t \in \mathbb{Z}$ yra vadinamas pirmos eilės autoregresiniu procesu (žymimas AR(1)), jei tenkina skirtuminę lygtį:

$$(1.6) \quad X_t = aX_{t-1} + \xi_t,$$

čia parametras a tenkina sąlygą $|a| < 1$, o $\{\xi_t, t \in \mathbb{Z}\}$ ir yra nepriklausomų standartinių normaliųjų atsitiktinių dydžių seka.

Generuoti atsitiktinio proceso AR(1) trajektoriją X_1, X_2, \dots, X_n yra paprasta. Tam pasirenkame $X_0 = 0$. Tada įstatę $t=1$ į lygtį (1.6) gauname:

$$X_1 = a\xi_1.$$

Vadinasi, sugeneravę ξ_1 reikšmę ir pasinaudoję paskutine lygybe gauname X_1 reikšmę. Toliau įstatome $t=2$ į lygtį (1.6):

$$X_2 = aX_1 + \xi_2.$$

Kadangi X_1 reikšmę jau turime apskaičiavę, tai sugeneruojame ξ_2 reikšmę ir pasinaudoję paskutine lygybe apskaičiuojame X_2 . Analogiškai tęsdami toliau randame likusias AR(1) proceso reikšmes X_3, X_4, \dots, X_n .

1. Temos analizė

Pagal teorijoje pateiktus algoritmus pirmiausia reikia realizuoti du pradinių duomenų generavimo algoritmus: AR ir FARIMA. Prieš pradėdant pradinių duomenų generavimą bet kuriuo iš šių generavimo algoritmų, turi būti galimi šių algoritmų pradiniai nustatymai. Pradinių nustatymų metu yra keičiami algoritmų kintamieji, kurie įtakoja rezultatus. Šiems algoritmams turi būti sukurta vartotojo sąsaja.

Po to reikia realizuoti IR statistikos skaičiavimo algoritmą. Prieš pradėdant IR statistikos skaičiavimus taip pat turi būti galimi pradiniai nustatymai. Šiam algoritmui taip pat turi būti sukurta vartotojo sąsaja.

IR algoritmo sudedamosios dalys:

1. IR statistikos skaičiavimas.
2. Pagal suskaičiuotą IR statistikos reikšmę reikia įvertinti atminties parametro reikšmę d pasirinktu tikslumu. Reikšmė d įvertinama atliekant žingsnių skaičiavimą, kurių metu apskaičiuojamos tam tikros reikšmės, kurios bus naudojamos sekančiame etape - įverčių skaičiavime. Kuo daugiau žingsnių atliekama, tuo tiksliau įvertinama reikšmė d . Turi būti realizuoti aštuoni žingsniai.
3. Įverčių skaičiavimo metu turi būti apskaičiuojamos reikšmės, kurios bus naudojamos paskutiniame skaičiavimų etape - reikšmių bias (poslinkis) ir mse (vidutinė kvadratinė paklaida) skaičiavime.
4. Suskaičiuotos reikšmės bias ir mse turi būti pateiktos kaip rezultatai - lentelė ir diagrama.
5. Algoritmas turi turėti galimybę atlikti skaičiavimus tiek kartų, kiek pasirenka vartotojas. Šių skaičiavimų metu yra keičiamas vienas parametras, kuris įtakoja rezultatus.

2. Darbo srities analizė

Šio darbo sudedamosios dalys:

- Pradinių duomenų generavimas:
 - AR duomenų generavimo algoritmas:
 - AR duomenų generavimo algoritmo pradiniai nustatymai.
 - AR duomenų generavimo algoritmo realizavimas.
 - FARIMA duomenų generavimo algoritmas:
 - FARIMA duomenų generavimo algoritmo pradiniai nustatymai.
 - FARIMA duomenų generavimo algoritmo realizavimas.
 - Abiejų šių algoritmų vartotojo sąsajos realizavimas.
- IR statistikos skaičiavimas:

- IR statistikos skaičiavimo algoritmo pradiniai nustatymai.
- IR statistikos skaičiavimo algoritmo realizavimas:
 - IR statistikos skaičiavimas.
 - Žingsnių skaičiavimas.
 - Įverčių skaičiavimas.
 - Reikšmių bias ir mse skaičiavimas.
 - Šio algoritmo vartotojo sąsajos realizavimas.
- Algoritmo skaičiavimų kartojimas, keičiant parametą.
- Rezultatų pateikimas.

Programinei realizacijai įrankiai nėra apibrėžti, todėl jų pasirinkimas yra laisvas.

IR statistikos skaičiavimas

IR statistika skaičiuojama pagal (1.1) formulę.

Žingsnių skaičiavimas

Kiekvieno žingsnio skaičiavimo metu yra skaičiuojama reikšmė Λ , kuri yra lyginama su prieš tai suskaičiuota IR statistikos reikšme. Kiek žingsnių bus skaičiuojama, pradiniuose nustatymuose pasirenka vartotojas. Kuo daugiau žingsnių skaičiuojama, tuo Λ reikšmė skaitiškai tampa panašesnė į IR statistikos reikšmę. Pirmojo žingsnio skaičiavimas atitinka vienos dešimtosios Λ reikšmės tikslumą nuo IR statistikos reikšmės. Atitinkamas tikslumas yra sekančiuose žingsniuose.

Λ reikšmė yra skaičiuojama pagal formulę (1.2). Šioje formulėje yra keičiama reikšmė r , kuri skaičiuojama pagal formulę (1.3), kaip reikšmė $\rho(d)$, jei tenkina tam tikras sąlygas.

(1.3) formulėje keičiant reikšmę d stengiamasi surasti reikšmę Λ , kuri kuo mažiau skirtųsi nuo reikšmės IR.

Pirmojo žingsnio skaičiavimas

Pirmiausia yra vykdomas ciklas, kuris yra atliekamas 20 kartų, pradedant nuo reikšmės -4 iki reikšmės 15. Kintant šiai reikšmei kinta ir reikšmė η_1 , t.y. η_1 reikšmė kinta nuo -4 iki 15. Su kiekviena η_1 reikšme yra skaičiuojama reikšmė Λ , kuri įstatoma į (1.3) formulę vietoj reikšmės d :

$$\text{artinys} = \frac{\text{eta1}}{10}$$

Priklausomai nuo reikšmės artinys, yra skaičiuojama reikšmė r. Jei reikšmė artinys lygi reikšmei 5, tai reikšmė r skaičiuojama:

$$r = \frac{16 \cdot \log 4 - 9 \log 9}{-8 \log 4}$$

Jei reikšmė artinys nelygi reikšmei 5, tai reikšmė r skaičiuojama pagal formulę (1.3).

Kiekvieno žingsnio metu yra skaičiuojama reikšmė artinys, kuri yra įstatoma į (1.3) formulę vietoj reikšmės d, o tada apskaičiuojama r reikšmė. Ši reikšmė r yra įstatoma į formulę (1.2) ir taip yra suskaičiuojama reikšmė Λ , kuri yra lyginama su anksčiau, pagal (1.1) formulę suskaičiuota, IR statistikos reikšme.

Jei IR reikšmė yra mažesnė už reikšmę Λ , tai pirmojo žingsnio skaičiavimo ciklas yra nutraukiamas, o reikšmė eta1 išsaugoma, t.y. kintamasis sis1 yra prilyginamas reikšmei eta1. Reikšmė sis1 bus naudojama tolimesniuose skaičiavimuose. Jei reikšmė IR nėra mažesnė už reikšmę Λ , tai ciklas yra kartojamas su kiekviena, vienetu didesne eta1 reikšme. Vėl atlikus skaičiavimus, Λ reikšmė yra lyginama su IR statistikos reikšme.

Antrojo žingsnio skaičiavimas

Šis žingsnis yra skaičiuojamas panašiai kaip ir pirmasis, tačiau šio žingsnio skaičiavimo ciklas yra atliekamas 10 kartų, t.y. nuo reikšmės 1 iki reikšmės 10, atitinkamai eta1 reikšmė kinta nuo 1 iki 10. Reikšmė eta2 skaičiuojama taip:

$$\text{eta2} = \text{sis1} - 1$$

sis1 reikšmė yra apskaičiuojama pirmojo žingsnio metu. Reikšmė artinys skaičiuojama taip:

$$\text{artinys} = \frac{\text{eta2}}{10} + \frac{\text{eta1}}{100}$$

Visa kita skaičiuojama kaip ir pirmąjame žingsnyje, o skaičiavimų pabaigoje yra išsaugoma reikšmė eta1, t.y. kintamasis sis2 yra prilyginamas reikšmei eta1. Reikšmė sis2 taip pat bus naudojama tolimesniuose skaičiavimuose.

Trečiojo - septintojo žingsnių skaičiavimai yra panašūs, skiriasi tik reikšmės artinys skaičiavimas, priklausomai nuo žingsnio numerio.

Aštuntojo žingsnio skaičiavimas

Šio žingsnio skaičiavimas panašus į prieš tai skaičiuotus žingsnius. Reikšmė artinys skaičiuojama taip:

$$\text{artinys} = \frac{\text{eta2}}{10} + \frac{\text{eta3}}{100} + \frac{\text{eta4}}{1000} + \frac{\text{eta5}}{10000} + \frac{\text{eta6}}{100000} + \frac{\text{eta7}}{1000000} + \frac{\text{eta8}}{10000000} + \frac{\text{eta1}}{100000000}$$

Visa kita skaičiuojama analogiškai kaip ir ankstesniuose žingsniuose. Skaičiavimų pabaigoje yra išsaugoma reikšmė eta1, t.y. kintamasis sis8 yra prilyginamas reikšmei eta1. Reikšmės sis1, sis2, ..., sis8 bus naudojamos įverčių skaičiavimui.

Skaičiuojamų žingsnių skaičius yra toks, kokį nustato vartotojas, nustatydamas skaičiuojamos d reikšmės tikslumą. Jei vartotojas nustato pvz. 5 skaičių po kablelio tikslumą, tai kiekvieno skaičiavimo metu bus skaičiuojami penki žingsniai nuo pirmojo iki penktojo, nes kiekvieno žingsnio skaičiavimo rezultatai yra naudojami sekančio žingsnio skaičiavime.

Įverčių skaičiavimas

Skaičiuojamų įverčių skaičius yra lygus skaičiuojamų žingsnių skaičiui, t.y. kiekvienam žingsniui yra skaičiuojamas įvertis.

Pirmojo įverčio skaičiavimas

Pirmiausia yra skaičiuojama reikšmė sis11:

$$\text{sis11} = \text{sis1} - 1$$

Reikšmė sis1 yra apskaičiuojama pirmajame žingsnyje. Po to yra skaičiuojama reikšmė ivertis[g]:

$$\text{ivertis}[g] = \frac{\text{sis11}}{10}$$

Reikšmė g atitinka reikšmę, kuri yra lygi duomenų bylos numeriui, t.y. bylos numeriui, su kurios duomenimis yra atliekami skaičiavimai. Reikšmės ivertis[g] bus naudojamos tolimesniuose skaičiavimuose.

Antrojo įverčio skaičiavimas

Analogiškai kaip ir pirmajame žingsnyje, pirmiausia yra skaičiuojama reikšmė sis11:

$$\text{sis11}=\text{sis1}-1$$

Reikšmė sis1 yra apskaičiuojama pirmajame žingsnyje. Po to skaičiuojama reikšmė sis22:

$$\text{sis22}=\text{sis2}-1$$

Reikšmė sis2 yra apskaičiuojama antrajame žingsnyje. Tada skaičiuojama reikšmė ivertis[g]:

$$\text{ivertis}[g] = \frac{\text{sis11}}{10} + \frac{\text{sis22}}{100}$$

Trečiojo - septintojo įverčių skaičiavimai yra panašūs, priklausomai nuo skaičiuojamo įverčio numerio.

Aštuntojo įverčio skaičiavimas

Atliekamų skaičiavimų eiliškumas:

sis11=sis1-1; Reikšmė sis1 yra apskaičiuojama pirmajame žingsnyje.

sis22=sis2-1; Reikšmė sis2 yra apskaičiuojama antrajame žingsnyje.

sis33=sis3-1; Reikšmė sis3 yra apskaičiuojama trečiajame žingsnyje.

sis44=sis4-1; Reikšmė sis4 yra apskaičiuojama ketvirtajame žingsnyje.

sis55=sis5-1; Reikšmė sis5 yra apskaičiuojama penktajame žingsnyje.

sis66=sis6-1; Reikšmė sis6 yra apskaičiuojama šeštajame žingsnyje.

sis77=sis7-1; Reikšmė sis7 yra apskaičiuojama septintajame žingsnyje.

sis88=sis8-1; Reikšmė sis8 yra apskaičiuojama aštuntajame žingsnyje.

Reikšmės ivertis[g] skaičiavimas:

$$\text{ivertis}[g] = \frac{\text{sis11}}{10} + \frac{\text{sis22}}{100} + \frac{\text{sis33}}{1000} + \frac{\text{sis44}}{10000} + \frac{\text{sis55}}{100000} + \frac{\text{sis66}}{1000000} + \frac{\text{sis77}}{10000000} + \frac{\text{sis88}}{100000000}$$

Skaičiuojamų įverčių skaičius yra toks, kokį nustato vartotojas, nustatydamas skaičiuojamos d reikšmės tikslumą. Jei vartotojas nustato pvz. 2 skaičių po kablelio tikslumą, tai kiekvieno skaičiavimo metu bus skaičiuojamas tik antras įvertis.

Reikšmių bias ir mse skaičiavimas

Suskaičiavus įverčių reikšmes, skaičiuojamos reikšmės bias (poslinkis) ir mse (vidutinė kvadratinė paklaida).

Reikšmės bias skaičiavimas

Remiantis (1.4) formulės pirmuoju sąryšiu skaičiuojame bias reikšmę pagal formulę:

$$\text{bias} = \frac{\sum_{i=1}^n x_i}{n} - a$$

Pirmiausia apskaičiuojame visų įverčių sumą ir ją padaliname iš pradinių duomenų bylų skaičiaus. Taip yra randamas įverčių vidurkis:

$$\text{vidurkis} = \frac{\sum_{g=1}^n \text{ivertis}[g]}{n}$$

Tada skaičiuojama reikšmė bias. Ji randama iš įverčių vidurkio atėmus reikšmę d:

$$\text{bias} = \text{vidurkis} - d$$

Reikšmės mse skaičiavimas

Remiantis (1.4) formulės antruoju sąryšiu skaičiuojame mse reikšmę pagal formulę:

$$\text{mse} = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2$$

Mūsų atveju reikšmė mse skaičiuojama taip:

$$\text{mse} = \sqrt{\frac{\sum_{g=1}^n (\text{ivertis}[g] - d)^2}{n}}$$

čia n - pradinių duomenų bylų skaičius.

Reikšmės bias ir mse yra pateikiamos kaip rezultatai programos vartotojo sąsajoje lentelė, o reikšmės mse yra vaizduojamos diagrama.

Skaičiavimų kartojimas, keičiant parametro reikšmę

IR statistikos, žingsnių, įverčių, reikšmių bias ir mse skaičiavimas gali būti kartojamas, (1.1) formulėje keičiant parametro m reikšmę. Parametras yra keičiamas pradinuose nustatymuose. Parametras gali būti keičiamas tokiais būdais:

- Pradinė reikšmė - reikšmė, nuo kurios bus pradami skaičiavimai.
- Žingsnis - reikšmė, kuria sekanti m reikšmė bus didesnė už buvusią.
- Reikšmių skaičius - su kiek m reikšmių bus atliekami skaičiavimai.

3. Darbinės srities modelis

Modelis apima šiuos uždavinius:

- Pradinių duomenų generavimo ir IR statistikos algoritmų realizavimas;
- Šių algoritmų pradinių nustatymų realizavimas;
- Programos vartotojo sąsajos realizavimas, skirtas pradinių duomenų generavimui ir IR statistikos skaičiavimui;
- Rezultatų pateikimas lentelė ir diagrama.

III. Projektinė dalis

1. Įrankių ir priemonių pasirinkimo analizė

Darbo įrankių ir priemonių pasirinkimo kriterijai:

- Universalumas – pasirinkto įrankio suderinamumas su populiariausiomis operacinėmis sistemomis. Šiuo įrankiu sukurtas produktas taip pat turi veikti populiariausių operacinių sistemų aplinkoje ir nereikalauti papildomų programų ar komponentų.

- Patikimumas - pasirinktas įrankis turi stabiliai atlikti sudėtingus matematinius skaičiavimus su dideliais duomenų kiekiais.
- Našumas - sudėtingi matematiniai skaičiavimai su dideliais duomenų kiekiais turi būti atlikti per kuo trumpesnį laiką.
- Vartotojo sąsajos kūrimo galimybė.

Programinė įranga

Šios programos kūrimui tinka daugelis programavimo kalbų, nes pagrindinį keliamą reikalavimą, t.y. sudėtingi matematiniai skaičiavimai, tenkina daugelis programavimo kalbų. Buvo tiriamos ir analizuojamos šios programavimo kalbos, kurios atitiko keliamus reikalavimus:

- Delphi
- C++
- Java

O pasirinkta Borland C++ Builder 5 programavimo kalba, dėl jau turimos programavimo patirties šia kalba.

2. Projekto (darbo) vykdymo planas

Etapas	Atliekami darbai	Rizikos	Pateikiami rezultatai
Pradinė temos analizė	Susipažinimas su tema, jai keliamais reikalavimais	Nesuprantama temos esmė, dėl ko gali užtrukti pradinės analizės vykdymo laikas	Parengtas pirminis temos realizavimo planas
Detali temos analizė	Informacijos apie temą paieška ir analizė	Informacijos stoka gali įtakoti projekto vykdymo laiką	Paruošta teorinė medžiaga, sukauptos žinios, priimtas projekto realizavimui reikalingas sprendimas

Įrankių parinkimas	Remiantis sukauptomis žiniomis, parinkti tinkamus įrankius projekto realizavimui	Tema gali reikalauti specifinių įrankių kuriuos sudėtinga rasti	Parinkti įrankiai projekto realizavimui
Projektavimas	Remiantis sukauptomis žiniomis, išsiaiškinti programinio produkto sudedamąsias dalis	Dėl įrankių netinkamumo gali užsitęsti jų parinkimas ir diegimas	Sukurta modulinė programinio įrankio schema
Programavimas	Programos sudedamųjų komponentų programavimas, algoritmo realizavimas	Dėl kompetencijos stokos gali užsitęsti vykdymo laikas. Gali iškilti problemų dėl kurių gali tekti gilinti žinias	Sukurta programinis produktas
Testavimas	Sukurto programinio produkto visapusiškas testavimas	Dėl programinio produkto veikimo nesklaidumų gali tekti ieškoti klaidų bei jas ištaisyti	Visapusiškai ištestuotas programinis produktas
Dokumentavimas	Galutinis darbo aprašymas, išvadų pateikimas		Atliktas pilnas dokumentacijos aprašymas

1 lentelė. Projekto (darbo) vykdymo planas

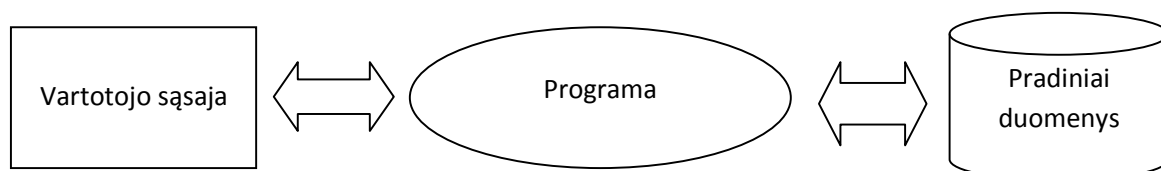
Projekto (darbo) planuojama vykdymo eiga

Projekto vykdymo planas																	
Etapas	I semestras				II semestras				III semestras				IV semestras				
	Spalis	Lapkritis	Gruodis	Sausis	Vasaris	Kovas	Balandis	Gegužė	Birželis	Spalis	Lapkritis	Gruodis	Sausis	Vasaris	Kovas	Balandis	Gegužė
Temos pradinė analizė	■	■															
Detali temos analizė			■	■													
Įrankių pasirinkimas					■												
Projektavimas						■	■	■									
Programavimas								■	■	■	■						
Testavimas												■	■				
Dokumentavimas															■	■	■

Pav. 2. Projekto (darbo) planuojama vykdymo eiga

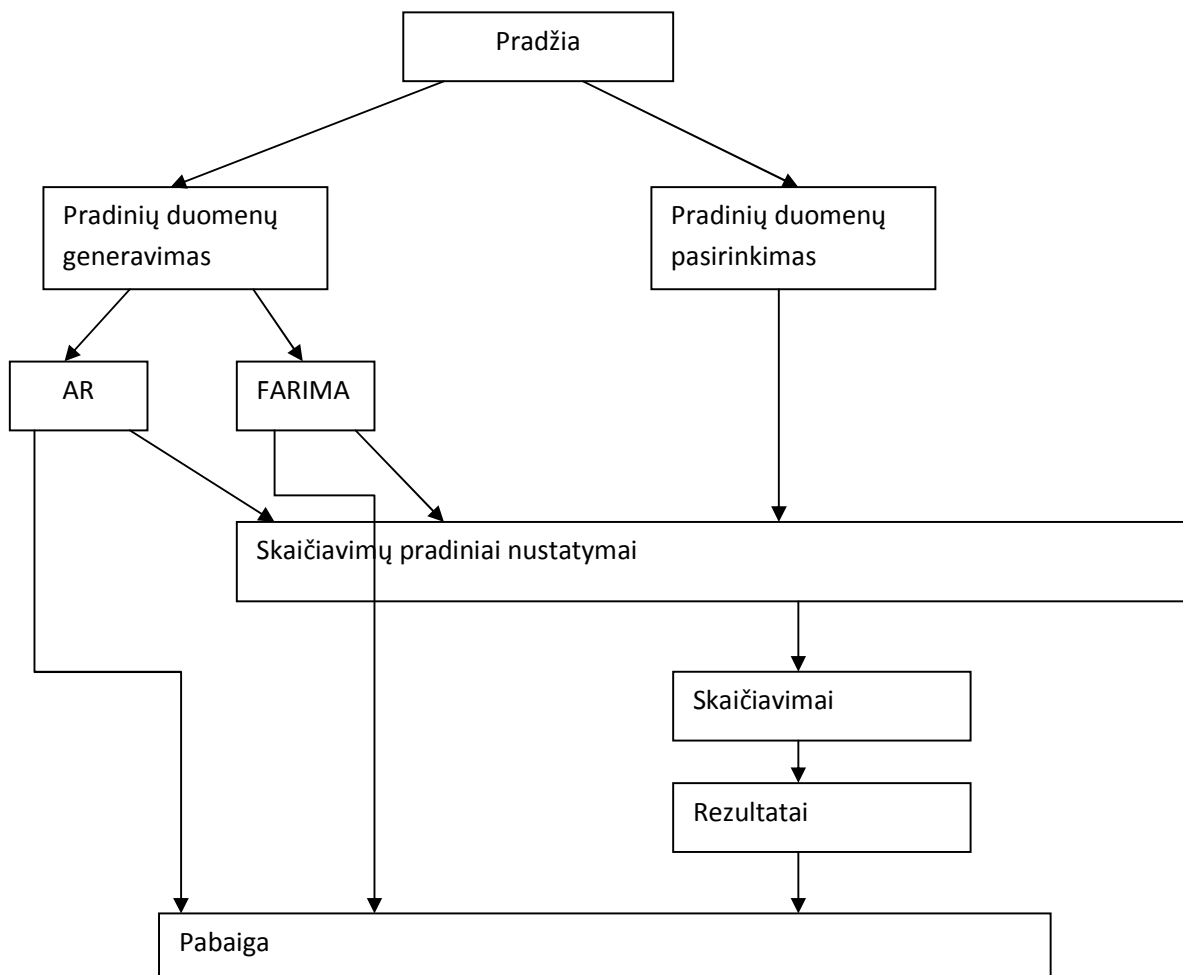
3. Pradinis projekto aprašymas

Pagrindinis projekto tikslas – pasirinkta programavimo kalba sukurti programą, kuri būtų naudinga tolimesniems IR statistikos tyrimams bei realių duomenų testavimui. Projektavimas buvo pradėtas nuo programos modelio:



Pav. 3 Programos modelis

Toliau buvo išsiaiškinta kokia eilės tvarka kuriamame produkte galima atlikti veiksmus. Tai vaizduoja sekanti schema:

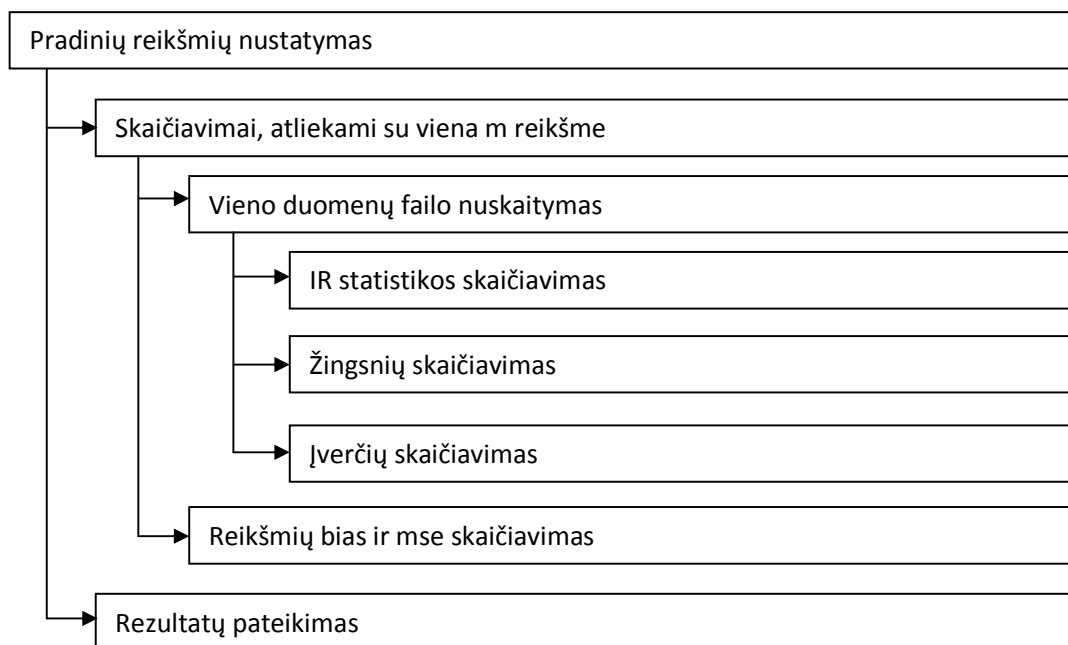


Pav. 4 Atliekamų veiksmų eiliškumo schema

4 paveikslėlyje pavaizduota dalis „skaičiavimai“ atitinka IR statistikos skaičiavimo algoritmą.

Programos realizavimo schema

Programos realizavimo schema parodo kaip realizuojamas IR statistikos skaičiavimo algoritmas, šio algoritmo pradiniai nustatymai ir rezultatų pateikimas.



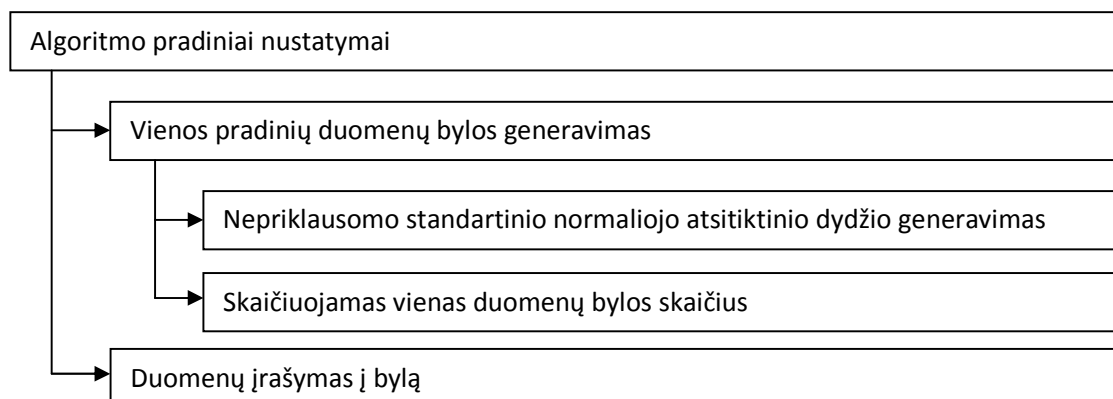
Pav. 5 Programos realizavimo schema

Pradinių duomenų generavimo algoritmai

Pradiniai duomenys gali būti generuojami dviem algoritmais. Pirmasis yra AR, antrasis – FARIMA. Pagal teorinėje dalyje pateiktus algoritmus buvo sudarytos šių algoritmų realizavimo schemas, o pagal šias schemas, remiantis algoritmais, realizuoti patys algoritmai.

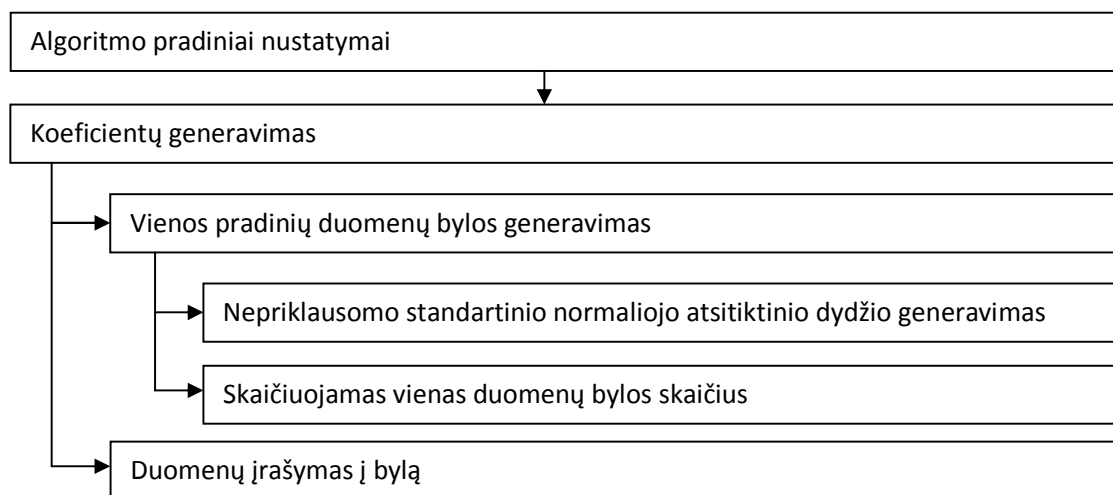
Abu generavimo algoritmai sugeneruotus pradinis duomenis išsaugo vartotojo pasirinktame kataloge.

AR duomenų generavimo algoritmo realizavimo schema



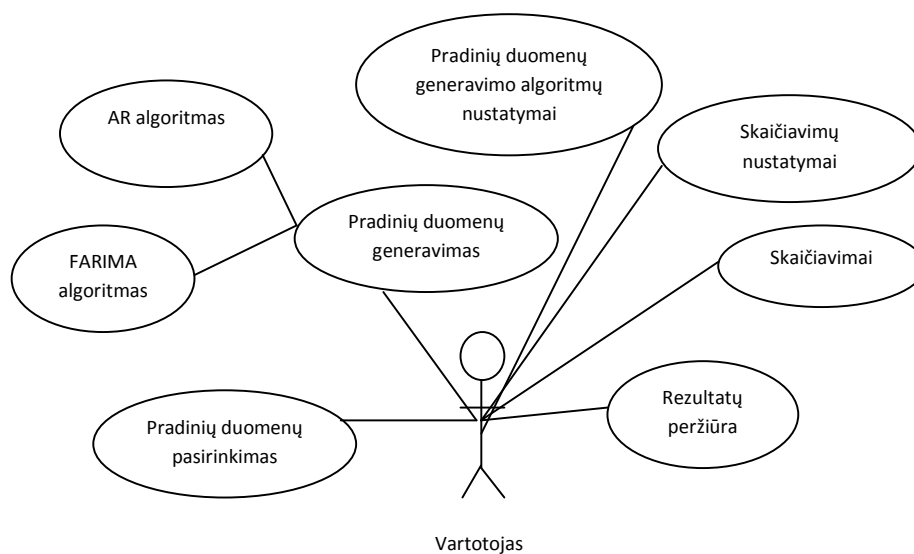
Pav. 6 AR duomenų generavimo algoritmo realizavimo schema

FARIMA duomenų generavimo algoritmo realizavimo schema



Pav. 7 FARIMA duomenų generavimo algoritmo realizavimo schema

Galutinis vartotojas su šia programa galės atlikti tokius veiksmus:



Pav. 8 Programos panaudojimo atvejai

IV. Darbo eigos aprašymas

1. Darbų eigos grafas

Planuojama darbo vykdymo eiga beveik nesiskyrė nuo realių darbų eigos.

Projekto vykdymo planas																	
Etapas	I semestras				II semestras				III semestras				IV semestras				
	Spalis	Lapkritis	Gruodis	Sausis	Vasaris	Kovas	Balandis	Gegužė	Birželis	Spalis	Lapkritis	Gruodis	Sausis	Vasaris	Kovas	Balandis	Gegužė
Temos pradinė analizė	■	■															
Detali temos analizė			■	■													
Įrankių pasirinkimas					■												
Projektavimas						■	■	■	■								
Programavimas									■	■	■	■	■	■			
Testavimas														■			
Dokumentavimas															■	■	■

Pav. 9 Darbų eigos grafas

2. Galutinio projekto būklės aprašymas

Projekte buvo sukurta:

1. Pradinių duomenų generavimo algoritmai:
 - AR
 - FARIMA
2. Pradinių duomenų apdorojimo ir rezultatų pateikimo algoritmai;
3. Visiems algoritmams sukurta vartotojo sąsaja.

Pagrindinės funkcijos

- AR(..) - generuoja pradinis duomenis pagal AR generavimo algoritimą.
- FARIMA(..) - generuoja pradinis duomenis pagal FARIMA generavimo algoritimą.
- IR(..) - skaičiuoja IR statistiką.
- Skaiciuoti(..) - atlieka žingsnių ir įverčių skaičiavimą.
- Iverciai(..) - atlieka galutinių rezultatų skaičiavimą ir jų pateikimą lentele bei diagrama.

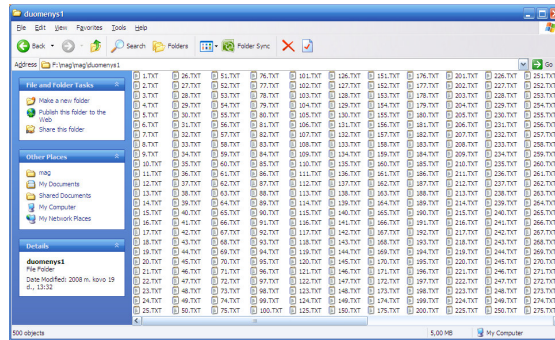
Pradinių duomenų aprašymas

Pradiniai duomenys yra tekstinės bylos (plėtinys *.txt), kuriose yra saugoma daugybė skaičių. Šios bylos privalo būti saugomos viename kataloge. Katalogas su tekstinėmis bylomis ir yra pradinis duomuo. Pradinius duomenis vartotojas programai gali "pateikti" pats arba juos sugeneruoti vienu iš dviejų duomenų generavimo algoritmų.

Reikalavimai pradiniam duomenims

Tekstinėse bylose yra saugomi skaičiai. Vienas skaičius yra saugomas vienoje eilutėje. Standartiškai vienoje tekstinėje byloje yra saugoma 1000 skaičių, t.y. vienoje byloje yra 1000 eilučių. Kiekviena byla turi būti pavadinta skaičiumi. Bylų vardai pradedami žymėti nuo skaičiaus 1 iki skaičiaus N, iš eilės, didėjančia tvarka. N - bylų skaičius.

Pradinių duomenų katalogo pavyzdys:



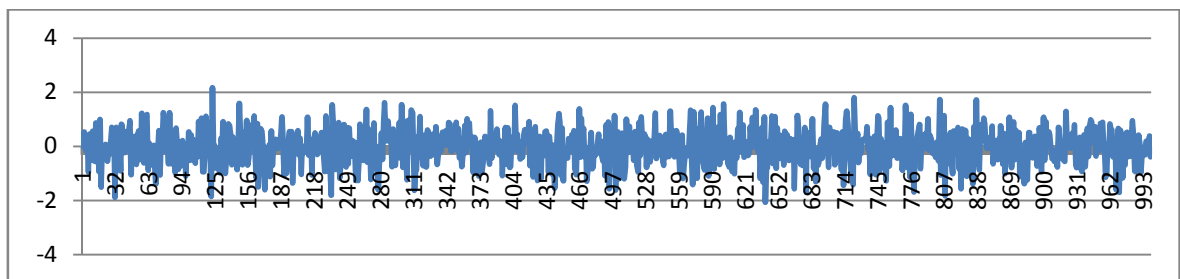
Pav. 10 Pradinių duomenų katalogo pavyzdys

Pradinių duomenų saugojimas

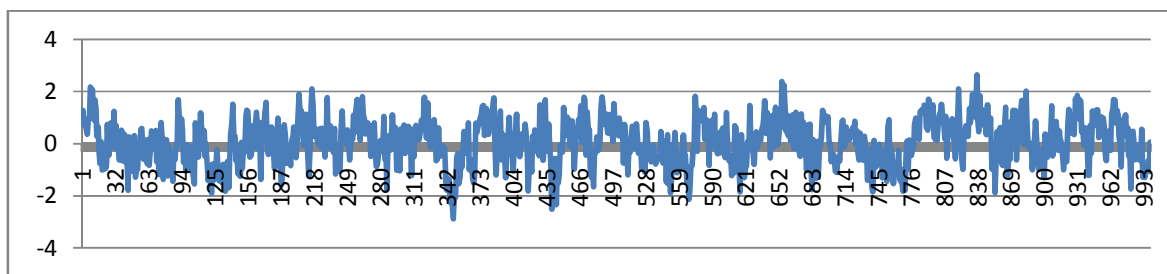
Pradiniai duomenys, kurie generuojami programa, yra išsaugomi pasirinktame kataloge. Katalogas turi būti sukurtas prieš pradėdant duomenų generavimą. Sugeneruotų duomenų katalogas gali būti naudojamas tolimesniuose skaičiavimuose.

Pradinių duomenų grafinis vaizdas

11 ir 12 paveikslėliuose pateikta vienoje sugeneruotoje duomenų byloje esančių skaičių reikšmių diagrama. Diagrama nubrėžta pasinaudojant Microsoft Office Excel programa. Horizontali diagramos ašis žymi duomenų byloje esančio skaičiaus numerį, o vertikali – šio skaičiaus reikšmę.



Pav. 11 AR generavimo algoritmo sugeneruotų duomenų grafinis vaizdas



Pav. 12 FARIMA generavimo algoritmo sugeneruotų duomenų grafinis vaizdas

Optimizavimas

Kuriant programą stengtasi kiek įmanoma ją optimizuoti, kad skaičiavimai būtų atliekami per kuo trumpesnę laiką. Programai atliekant duomenų generavimą ar skaičiavimus, tarpiniai skaičiavimai nėra pateikiami. IR statistikos skaičiavimo metu yra pateikiami tik rezultatai, t.y. kai tik su viena m reikšme yra suskaičiuojamos reikšmės bias ir mse, jos iškart yra parodomos vartotojo sąsajos lentelėje. Tik tokių tarpinių rezultatų pateikimas vartotojui yra nepatogus, nes jis negali matyti ką programa atlieka konkrečiu skaičiavimo laiko momentu, bei negali žinoti kiek ilgai programa dar atliks skaičiavimus. Tačiau bet koks tarpinių skaičiavimų pateikimas vartotojo sąsajoje pastebimai sulėtina skaičiavimų atlikimo spartą. Tad tarpinių rezultatų pateikimo vartotojo sąsajoje atsisakyta. Vartotojui yra pateikiami tik galutiniai rezultatai. Kiek laiko programa atliks duomenų generavimą ar skaičiavimus vartotojas gali orientotis pagal duomenų kiekio ir skaičiavimo (generavimo) laiko lentelę, kuri yra pateikta darbo rezultatų analizėje.

3. Problemų ir jų sprendimų aprašymai ir pagrindimai

Kadangi į kuriamą programą panašių nėra, tad ir literatūros šia tema nėra daug. Visa rasta literatūra yra parašyta “techninio” pobūdžio anglų kalba, tad buvo daug sąvokų, kurias teko analizuoti.

Tarpinių duomenų pateikimas vartotojo sąsajoje skaičiavimo ar duomenų generavimo metu pastebimai sulėtino skaičiavimo laiką, todėl to buvo atsisakyta.

Programos atliekamų skaičiavimų teisingumas buvo tikrinamas atliekant daug testavimų.

4. Patarimai, pastebėjimai, rekomendacijos

Kai programa atlieka skaičiavimus ar duomenų generavimą, patariama, kad kitos veikiančios programos būtų išjungtos – skaičiavimai bus atliekami greičiau.

Kuo didesnis pradinių duomenų kiekis ir/ar žingsnių skaičius pasirenkamas, tuo tikslesni rezultatai gaunami.

Nors visus iškeltus uždavinius realizuoti pavyko, tačiau šią programą dar galima tobulinti, jei skaičiavimai būtų atliekami su daug didesniais duomenų kiekiais. Tokiu atveju šią programą būtų galima pritaikyti prie naujausių kompiuterių, kurie turi keletą procesorių arba jų procesoriai yra sudaryti iš kelių branduolių. Tokiu atveju, skaičiavimo metu, kompiuterio resursai būtų pilnai išnaudojami, o skaičiavimo laikas būtų daug trumpesnis.

5. Darbo rezultatų analizė

Rezultatai yra pateikiami vartotojo sąsajoje lentele ir diagrama. Pirmoji reikšmė m lentelėje ir diagramoje rodoma ta, kurią pradiniuose nustatymuose prieš skaičiavimus pasirenka vartotojas. Sekanti reikšmė m didėja reikšme, kurią taip pat pasirinko vartotas prieš skaičiavimus. Lentelėje ir diagramoje yra tiek m reikšmių, kiek pasirinko vartotojas.

Rezultatų lentelė

Lentelėje pateikti rezultatai parodo reikšmes m, bias ir mse. Kadangi su kiekviena m reikšme yra skaičiuojamos reikšmės bias ir mse, tai lentelėje parodyta, kaip nuo reikšmės m priklauso reikšmės bias ir mse.

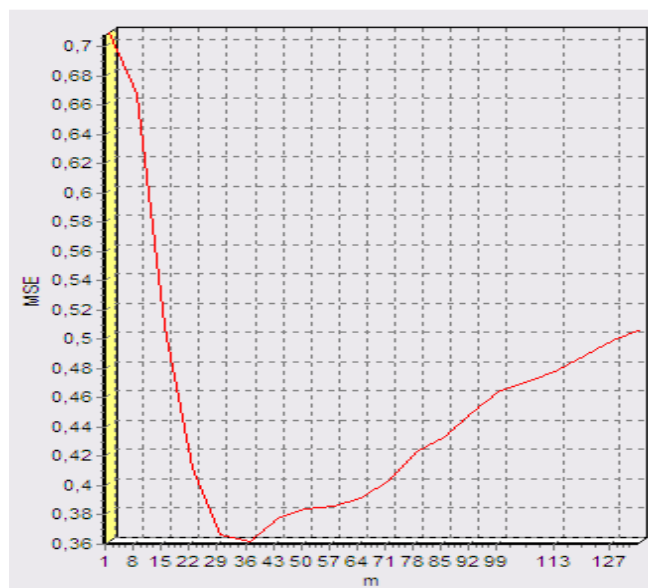
Rezultatų lentelės pavyzdys:

m	BIAS	MSE
1	0,705635726451874	0,707771897315979
8	0,655397951602936	0,665563702583313
15	0,473644584417343	0,506853401660919
22	0,33562645316124	0,409426033496857
29	0,25163608789444	0,363878548145294
36	0,194992765784264	0,359535485506058
43	0,152392461895943	0,375419586896896
50	0,118090443313122	0,381624668836594
57	0,0965715125203133	0,383961945772171
64	0,0841420292854309	0,38901075720787
71	0,0796489790081978	0,401204794645309
78	0,0791947022080421	0,420655518770218
85	0,071734607219696	0,431543469429016
92	0,0712810382246971	0,447673290967941
99	0,0650312453508377	0,462867766618729
106	0,0645901337265968	0,469003766775131
113	0,0676763281226158	0,47728243470192
120	0,0756223052740097	0,487327396869659
127	0,0829277783632278	0,497198134660721
134	0,086947999894619	0,504707217216492

Pav. 13 Rezultatų lentelė

Rezultatų diagrama

Kad vartotojui būtų patogiau, rezultatai yra pateikiami ir diagrama. Šioje diagramoje parodyta, kaip nuo reikšmės m priklauso reikšmė mse . Rezultatų diagramos pavyzdys:



Pav. 14 Rezultatų diagrama

Sukurto produkto panaudojimas

Sukurta produktas tinkamas šioms praktinėms reikmėms:

1. Mokslo tyrinėjimams.
2. Praktinėms reikmėms.

Testavimas

Sukurta produktas buvo testuojamas su populiariausiomis operacinėmis sistemomis. Jis veikė šiose Windows šeimos operacinėse sistemose: Windows 98, Windows 2000, Windows XP ir Windows Vista operacinėse sistemose.

Duomenų kiekio ir skaičiavimo (generavimo) laiko lentelės

Kompiuterio, su kuriuo buvo atliekami testavimai techninės charakteristikos:

- Procesorius – Mobile DualCore Intel Pentium M, 1,6 GHz;
- Operatyvioji atmintis – 1 GB;
- Operacinė sistema - Windows XP Media Center Edition.

Lentelė parodo kaip priklauso AR duomenų generavimo algoritmo generuojamų duomenų kiekis nuo jų generavimo laiko.

Duomenų kiekis (bylų skaičius)	Byloje esančių duomenų kiekis	Generavimo laikas
100	1000	1 s.
1000	1000	4 s.
10000	1000	40 s.
50000	500	3 min. 35 s.
100000	1000	12 min. 30 s.

2 lentelė. AR duomenų generavimo algoritmo sugeneruotų duomenų kiekio ir generavimo laiko priklausomybės lentelė

Lentelė parodo kaip priklauso FARIMA duomenų generavimo algoritmo generuojamų duomenų kiekis nuo jų generavimo laiko.

Duomenų kiekis (bylų skaičius)	Byloje esančių duomenų kiekis	Generavimo laikas
100	500	1 s.
1000	1000	10 s.
10000	1000	1 min. 50 s.
50000	1000	10 min. 25 s.

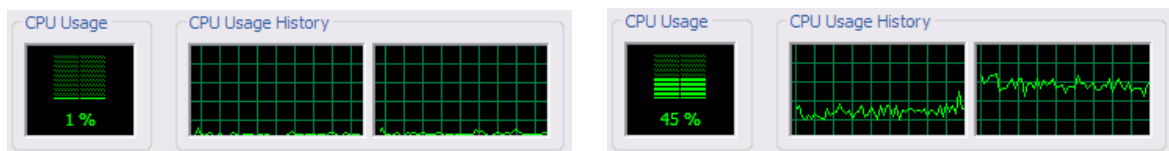
3 lentelė. FARIMA duomenų generavimo algoritmo sugeneruotų duomenų kiekio ir generavimo laiko priklausomybės lentelė

Lentelė parodo kaip priklauso IR statistikos skaičiavimo algoritmo atliekamų skaičiavimo laikas nuo pradinių duomenų kiekio.

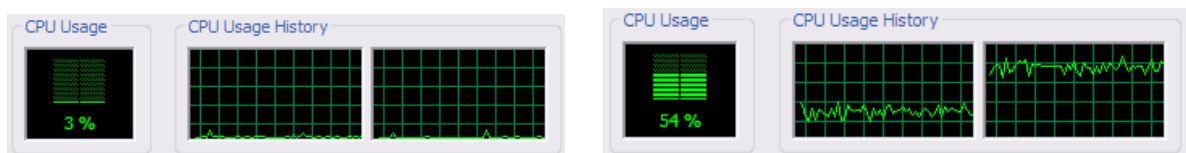
m reikšmių skaičius	Duomenų kiekis (bylų skaičius)	Byloje esančių duomenų kiekis	Skaičiavimo laikas
5	100	1000	2 s.
20	1000	1000	1 min. 4 s.
50	1000	1000	2 min. 25 s.
30	2000	1000	3 min. 4 s.
50	2000	1000	5 min. 42 s.

4 lentelė. Duomenų kiekio ir skaičiavimo laiko lentelė

Atliekant pradinių duomenų generavimą ir IR statistikos skaičiavimą buvo tikrinama kompiuterio apkrova.



Pav. 15 Kompiuterio apkrova prieš atliekant duomenų generavimą ir generavimo metu



Pav. 16 Kompiuterio apkrova prieš atliekant IR statistikos skaičiavimus ir skaičiavimo metu

Paveikslėliai parodo, kad pradinių duomenų generavimo ir IR statistikos skaičiavimo metu yra išnaudojamas tik vienas iš dviejų procesoriaus branduolių, tad šią programą dar galima optimizuoti, kad kompiuterio resursai būtų labiau išnaudojami.

V. Išvados

Darbo pradžioje kelti reikalavimai buvo sėkmingai įvygdyti. Sukurta programa, dviem algoritmais generuojanti pradinius duomenis ir skaičiuojanti IR statistiką.

- Norint suskaičiuoti tikslesnę IR statistikos reikšmę reikia pasirinkti arba sugeneruoti daugiau pradinių duomenų ir/arba pasirinkti didesnę žingsnių skaičių.
- Sukurta programa veikė daugelyje Windows šeimos operacinių sistemų aplinkų.
- Realizuotą programą galima naudoti tolimesniems IR statistikos tyrimams.
- Sukurtą programinį produktą galima tobulinti.

VI. Literatūros ir informacinių šaltinių sąrašai

[1] Beran J.(1998) *Statistics for long-memory processes*, Chapman and Hall/CRC.

[2] Bardet J.M. and Surgailis D. (2007) Measuring roughness of random paths by increment ratios, *Stochastic Process. Appl.* (to appear).

[3] Surgailis D., Teyssiére G. and Vaičiulis M. (2008) The increment ratio statistic, *J. Multivariate Anal.*, 99, pp.510-541.

[4] Teyssiére G. and Abry P. (2007) Wavelet analysis of financial time series, *Long memory in economics*, Eds. Teyssiére G. and Kirman A.P., pp.172-238.

[5] Teyssiére G. and Abry P. Wavelet Analysis of Nonlinear Long-Range Dependent Processes. Applications to Financial Time Series financial time series, pp.183-238.

[6] Bardet J.M., Lang G., Oppenheim G., Philippe A., S. Taqqu M. Generators of Long-Range Dependent Processes: A Survey, pp.590-624.

VII. Anotacija

Šiame magistro darbe nagrinėjama IR statistika bei FARIMA ir AR duomenų generavimo algoritmai. Darbe pateikta programinė IR statistikos skaičiavimo bei FARIMA ir AR duomenų generavimo algoritmų realizacija. Visi algoritmai visapusiškai ištestuoti bei jiems sukurta vartotojo sąsaja.

IR statistika dar nėra pilnai įrodyta, todėl šis darbas galės būti naudojamas tolimesniems IR statistikos tyrimams.

Annotation

IR statistics, FARIMA and AR data production algorithms are analyzed in this master's degree study. I present IR statistics calculating, FARIMA and AR data production algorithms software realization. All algorithms are universally tested and has user interface.

IR statistics is not fully proven, that's why this study could be used in further IR statistics researches.

VIII. Kompaktinio disko turinys

Pateikto kompaktinio disko turinys:

- C++ programos išeities bylų katalogas;
- IR.exe pagrindinė paleidžiamoji programa;
- Duomenys pradiniai duomenų katalogas;
- Darbas.docx magistrinio darbo aprašymas;
- Darbas.pdf magistrinio darbo aprašymas;
- CD turinys kompaktinio disko turinio aprašymas.

IX. Priedai

AR duomenų generavimo algoritmo vartotojo sąsajos aprašymas

AR duomenų generavimo algoritmo vartotojo sąsaja yra gan paprasta vartotojui. Pirmiausia reikia pasirinktoje kompiuterio kietojo disko vietoje susikurti katalogą laisvai pasirinktu vardu. Po to šį katalogą reikia nurodyti šio duomenų generavimo algoritmo vartotojo sąsajoje. Belieka nustatyti pradines reikšmes ir spausti mygtuką “Generuoti”. Po tam tikro laiko (priklausomai nuo pasirinktų reikšmių) bus parodytas pranešimas, kad duomenų generavimas yra baigtas. Šiame kataloge bus sugeneruotas pasirinktas duomenų bylų skaičius. Šis katalogas gali būti naudojamas tolimesniuose skaičiavimuose.

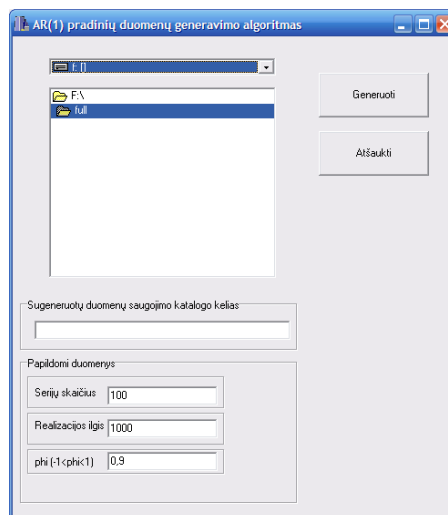
AR duomenų generavimo algoritmo pradiniai nustatymai

Serių skaičius – norimų sugeneruoti duomenų bylų skaičius.

Realizacijos ilgis – vienoje duomenų byloje norimų sugeneruoti skaičių kiekis.

Reikšmė phi – ši reikšmė galima tik intervale nuo -1 iki 1.

AR duomenų generavimo algoritmo vartotojo sąsaja



Pav. 17 AR duomenų generavimo algoritmo vartotojo sąsaja

FARIMA duomenų generavimo algoritmo vartotojo sąsajos aprašymas

FARIMA duomenų generavimo algoritmo vartotojo sąsaja yra gan paprasta vartotojui. Pirmiausia reikia pasirinktoje kompiuterio kietojo disko vietoje susikurti katalogą laisvai pasirinktu vardu, po to šį katalogą nurodyti šio duomenų generavimo algoritmo vartotojo sąsajoje. Belieka nustatyti pradines reikšmes ir spausti mygtuką “Generuoti”. Po tam tikro laiko (priklausomai nuo pasirinktų reikšmių) bus parodytas pranešimas, kad duomenų generavimas yra baigtas. Šiame kataloge bus sugeneruotas pasirinktas duomenų bylų skaičius. Šis katalogas gali būti naudojamas tolimesniuose skaičiavimuose.

FARIMA duomenų generavimo algoritmo nustatymai

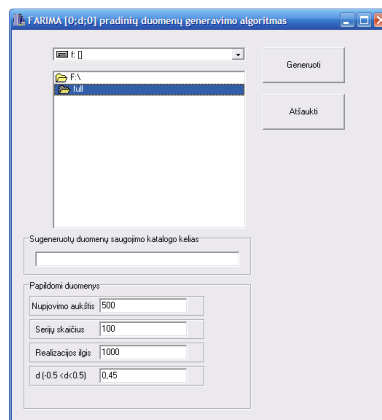
Nupjovimo aukštis - kintamasis, kuris įtakoja rezultatus.

Serių skaičius – norimų sugeneruoti duomenų bylų skaičius.

Realizacijos ilgis – duomenų byloje norimų sugeneruoti skaičių kiekis.

Reikšmė d – ši reikšmė galima tik intervale nuo -0.5 iki 0.5.

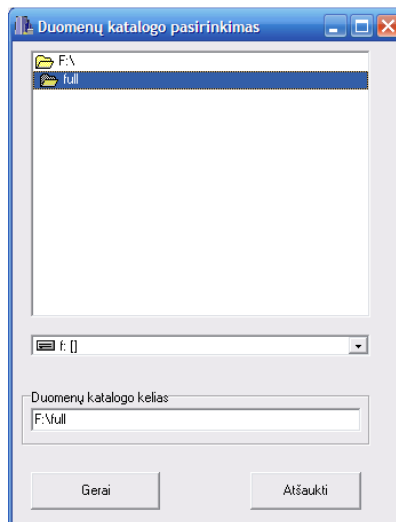
FARIMA duomenų generavimo algoritmo vartotojo sąsaja



Pav. 18 FARIMA duomenų generavimo algoritmo vartotojo sąsaja

Duomenų katalogo pasirinkimo vartotojo sąsaja

Šioje vartotojo sąsajoje tereikia nurodyti norimą pradinių duomenų katalogą ir spausti mygtuką “Gerai”. Duomenų katalogą taip pat galime sugeneruoti naudodami šią programą, vienu iš pasirinktų duomenų generavimo algoritmų, t.y. FARIMA arba AR.



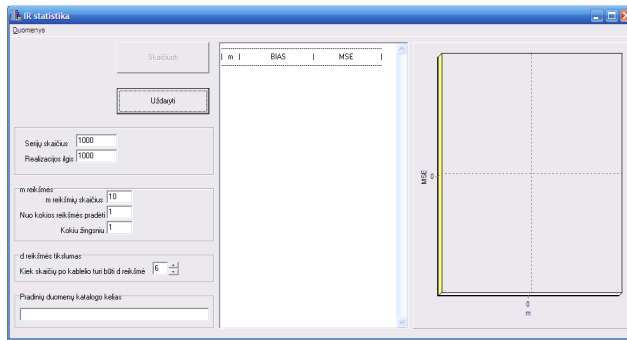
Pav. 19 Duomenų katalogo pasirinkimo vartotojo sąsaja

Programos aprašymas

Norint pradėti skaičiavimus, pirmiausia reikia pasirinkti pradinių duomenų katalogą. Po to reikia atlikti pradinius programos atliekamų skaičiavimų nustatymus, t.y. nurodyti kokios turi būti N ir m reikšmės. Tada spaudžiame mygtuką "skaičiuoti" ir po tam tikro laiko (priklausomai nuo pradinių duomenų ir m reikšmių kiekio) parodomas pranešimas, kad programa baigė skaičiavimus. Skaičiavimų metu yra rodomi ir tarpiniai skaičiavimų rezultatai, t.y. parodomas suskaičiuotos bias ir mse reikšmės su kiekviena m reikšme. Diagrama yra parodoma tik baigus visus skaičiavimus.

Programos vartotojo sąsaja

Grafinę vartotojo sąsają stengtasi sukurti kuo paprastesnę ir suprantamesnę net ir eiliniam vartotojui. Yra sukurtas privalomas veiksmų eiliškumas, t.y. skaičiavimo mygtukas bus neaktyvus tol, kol nebus pasirinktas pradinių duomenų katalogas.



Pav. 20 Programos vartotojo sąsaja

Pradiniai programos nustatymai

Serių skaičius N – kiek pradinių duomenų bylų bus naudojama skaičiavimuose;

Reikšmė m :

- m reikšmių skaičius – kiek m reikšmių bus naudojama skaičiavimuose;
- nuo kokios reikšmės pradėti – m reikšmė, nuo kurios bus pradedami skaičiavimai;
- koku žingsniu – reikšmė, kuria sekanti m reikšmė bus didesnė už prieš tai buvusią;

d reikšmės tikslumas – kiek skaičių po kablelio turės d reikšmė (didžiausias tikslumas - 8 skaičiai po kablelio).