

LATVIJAS UNIVERSITĀTE  
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE  
MATEMĀTIKAS NODAĻA

**AKCIJU TIRGUS CENU PROGNOZĒŠANA,  
IZMANTOJOT SLĒPTOS MARKOVA MODEĻUS**

BAKALAURA DARBS

Autors: **Laura Jozefa**  
Studenta apliecības Nr.: lj13026  
Darba vadītājs: Dr. math. Nataļja Budkina

RĪGA 2017

## ANOTĀCIJA

Bakalaura darba pētāmā problēma ir slēpto Markova modeļu pielietojums akciju tirgus cenu prognozēšanai.

Darbs ir veltīts slēpto Markova modeļu un to galveno pamatproblēmu izpētei un risināšanai, kā arī akciju tirgus cenu prognozēšanai trīs uzņēmumu akcijām, izmantojot gan slēptos Markova modeļus ar diviem slēptajiem stāvokļiem, gan ar trīs slēptajiem stāvokļiem. Katra uzņēmuma prognožu precizitātes ir salīdzinātas, izmantojot vidējo absolūto procentuālo kļūdu un izveidojot grafisku attēlojumu, salīdzinot reālās akciju cenas ar iegūtajām prognozētajām akciju cenām.

Datu analizēšanai un aprēķiniem izmantota programma Microsoft Excel, kā arī statistisko aprēķinu programma R.

Atslēgas vārdi: Slēptie Markova modeļi (SMM), akciju tirgus, akciju cenas, prognozes

## ANOTATION

Bachelor thesis research topic is application of hidden Markov models to stock price prediction.

Bachelor thesis is dedicated to hidden Markov models, to exploration and settlement of their key underlying problems, as well as the stock market price forecasting for three companies using both hidden Markov models with two hidden states and hidden Markov models with three hidden states. Each company's forecast accuracy is compared using mean absolute percentage error and reflected in a graphic representation comparing actual stock price with predicted stock price.

Data analysis and calculations are done using Microsoft Excel, as well as the statistical calculation program R.

Keywords: Hidden Markov models (HMM), stock market, stock prices, predictions

## SATURS

Ievads .....	5
1. Slēptais Markova modelis .....	7
1.1. Markova process .....	7
1.2. Teorētiskais apraksts par slēptiem Markova modeļiem .....	9
1.3. Slēpto Markova modeļu pamatproblēmas .....	10
2. Slēpto Markova modeļu pielietojums akciju cenu prognozēšanā .....	20
2.1. Modeļa pielietojums un nepieciešamo aprēķinu veikšana .....	20
2.2. Iegūto prognožu rezultāti un precizitāte .....	22
Secinājumi .....	25
Izmantotā literatūra un avoti .....	27
Pielikumi	
1.pielikums. Aprēķini programmā Microsoft Excel .....	29
2.pielikums. Izveidotais programmas R kods .....	30

## IEVADS

Akciju tirgus ir kā tīkls, kas nodrošina platformu gandrīz visiem galvenajiem saimnieciskajiem darījumiem pasaulē, kas ir balstīti uz akciju vērtībām un tirgus līdzsvaru. Akciju vērtību prognozēšana piedāvā milzīgas peļņas iespējas, kas ir liela motivācija veikt pētījumus šajā jomā. Iepriekš zināma akciju vērtība, kaut tikai sekundi, var izvērsties par lielu peļņu. Vilinājums atrast risinājumu ir piesaistījis pētniekus gan rūpniecības, gan akadēmiskajās nozarēs, lai varētu atrisināt tādas problēmas kā svārstīgumu, sezonalitāti, atkarību no laika, ekonomikas un pārējā tirgus. Agrāk tika izmantotas mašīnmācīšanās un mākslīgās izlūkošanas metodes - mākslīgie neironu tīkli, atbalsta vektoru mašīnas un citas.

Tikai nesen slēpto Markova modeļu pieeja tika piemērota šai problēmai. [1] Iemesls izmantot šo pieeju ir ļoti intuitīvs. Slēptie Markova modeļi agrāk bijuši piemēroti, lai analizētu un prognozētu laika atkarības fenomenu vai laikrindas. Tie tika izmantoti runas atpazīšanā, kā arī EKG analīzē. Akciju tirgus cenu prognozes problēma ir līdzīga tādā ziņā, ka arī tai ir saistība ar laiku. Slēpto Markova modeļu pamatā ir kopums, kas sastāv no nepārraudzītiem stāvokļiem, kuru starpā var notikt pārejas un kur katrs stāvoklis ir saistīts ar pamata novērojumu kopu. Akciju tirgū arī varam novērot ko līdzīgu – stāvokļi, kas nosaka akciju cenas vērtību, parasti investoram ir neredzami. Pārejas starp šiem stāvokļiem galvenokārt nosaka uzņēmuma politika, lēmumi un ekonomiskie apstākļi. Redzamais efekts, kas atspoguļo visu iepriekš minēto, ir akciju cenas vērtība.

Tāpēc atribūtu un līdzekļu izvēle ir ļoti nozīmīga šajā pieejā. Pēdējie mēģinājumi ir veikti, izmantojot tirdzniecības apjomu, akciju tirgus dinamiku, saistību ar tirgu un akciju vērtību svārstībām. Šajā darbā aprakstītajā modelī tiek izmantotas ikdienas izmaiņas akciju vērtībā. Ikdienas cenu izmaiņas ir nepieciešams izmantot, lai varētu izveidot nepieciešamo prognozi un novirzes noteikšana ir labs lielums, jo tas dod ieskatu, kāda kopumā ir virzība akciju vērtības svārstībām.

Šajā darbā tiek izmantotas trīs uzņēmumu tirgus akcijas, lai varētu labāk izvērtēt pieeju – Apple Inc, Microsoft Corporation un Samsung Group.

Darba mērķis ir izpētīt kā un cik precīzi iespējams prognozēt akciju tirgus cenas, izmantojot slēptos Markova modeļus. Mērķa sasniegšanai ir izvirzīti sekojoši darba uzdevumi:

1. Iepazīties ar slēptiem Markova modeļiem un to pamatproblēmām;
2. Iepazīties ar slēpto Markova modeļu pielietojumu akciju tirgus cenu prognozēšanai;
3. Izveidot programmas R kodu algoritma pielietojumam;

4. Pielietot aplūkoto teoriju un izveidoto programmu reālām uzņēmumu akcijām;
5. Izveidot secinājumus.

Izstrādātais darbs ir organizēts sekojoši: pirmajā nodaļā tiek apskatīti Markova procesi, lai varētu izprast, kā tie darbojas un lai varētu turpmāk strādāt ar Markova modeļiem, kā arī pirmajā nodaļā apskatīts vispārīgums tieši uz slēptajiem Markova modeļiem. Papildus pirmajā nodaļā ir apskatītas trīs slēpto Markova modeļu pamatproblēmas un atbilstošie algoritmi šo problēmu novēršanai, atrisināšanai. Otrajā nodaļā tiek sniegta informācija par izstrādāto pieeju akciju cenu prognozēšanā un prezentēti iegūtie akciju cenu prognožu rezultāti. Darba beigās ir apkopoti secinājumi un izteiktas rekomendācijas algoritma uzlabošanai. Pielikumā ir ievietots izstrādātais programmas R kods.

# 1. SLĒPĒTAIS MARKOVA MODELIS

## 1.1. Markova process

Šajā nodaļā tiks izmantoti avoti [2], [3], [4] un [5].

Jēdzieni „Markova procesi” un „Markova ķēdes” radās 1906. gadā, kad krievu matemātiķis Andrejs Markovs (dzimis 1856. gada 14. jūnijā, miris 1922. gada 20. jūlijā) nāca klajā ar pirmajiem teorētiskiem stohastisko procesu rezultātiem. Vēlāk viņa darbība tieši koncentrējās ap Markova ķēdēm un procesiem.

**Definīcija 1.** Stohastisks process  $X$  ir gadījuma lielumu saime  $\{X_t : t \in T\}$ , kur  $T$  ir indeksu kopa. Ja  $T = \{0, 1, 2, \dots\}$ , tad process ir diskrēta laika process, bet, ja  $T = \mathbb{R}$  vai  $T = [0, \infty)$ , tad tas ir nepārtraukta laika process.

Pieņemsim, ka laika momentā  $t$  procesa  $\{X_t : t \in T\}$  notikumu vēsture ir

$$\{X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}, X_t = x_t\}. \quad (1.1)$$

Tad nosacītā varbūtība

$$P(X_{t+1} = x_{t+1} | X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}, X_t = x_t) \quad (1.2)$$

atspoguļo varbūtību notikumam laika momentā  $t + 1$ , ja ir dota visa procesa vēsture līdz laika momentam  $t$ . Varbūtība

$$P(X_{t+1} = x_{t+1} | X_t = x_t) \quad (1.3)$$

ir nosacītā varbūtība nākotnes notikumam, ja ir dots tikai tagadnes jeb pašreizējais stāvoklis. Ja abas šīs varbūtības ir vienādas, tad izpildās Markova īpašība – ja ir zināma tagadnes  $X_t$  vērtība, tad iepriekšējās jeb pagātnes vērtības neietekmē nākotnes vērtības.

**Definīcija 2.** Stohastisku procesu  $\{X_t : t \in T\}$  ar stāvokļu kopu  $S$  sauc par Markova procesu, ja tam piemīt Markova īpašība:

$$P(X_{t+1} = x_{t+1} | X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}, X_t = x_t) = P(X_{t+1} = x_{t+1} | X_t = x_t) \quad (1.4)$$

visiem  $t \geq 1$  un visiem  $x_1, x_2, \dots, x_{t-1}, x_t \in S$ .

**Definīcija 3.** Markova ķēde ir Markova process ar diskrētu stāvokļu kopu, kas var būt gan saskaitāma, gan galīga.

**Definīcija 4.** Ja  $X_1, X_2 \dots$  ir Markova ķēde un  $S_i, S_j \in S$ , tad varbūtību

$$a_{ij}(t) = P(X_{t+1} = S_j | X_t = S_i) \quad (1.5)$$

sauc par pārejas varbūtību no stāvokļa  $i$  uz stāvokli  $j$  laika momentā  $t$ . Sistēma ne vienmēr pāriet jaunā stāvoklī – tā var nākamajā momentā palikt tajā pašā stāvoklī kā iepriekšējā momentā un tas notiek ar varbūtību  $a_{ii}$ .

**Definīcija 5.** Markova ķēdi  $X$  sauc par homogēnu attiecībā pret laiku, ja pārejas varbūtības ir neatkarīgas no laika:

$$P(X_{t+1} = j | X_t = i) = P(X_1 = j | X_0 = i). \quad (1.6)$$

Tādā gadījumā pārejas varbūtības pieraksta vienkārši kā  $a_{ij}$ .

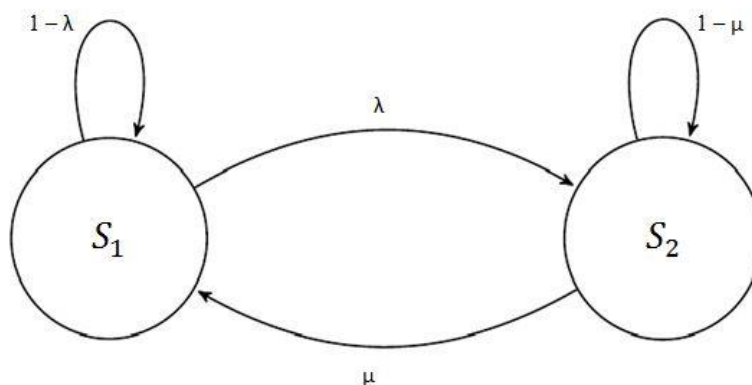
Pieņemsim, ka stāvokļu kopa  $S$  ir ierobežota un tā sastāv no  $N$  stāvokļiem, tātad varam to apzīmēt ar  $S = S_1, S_2, \dots, S_N$ . Parasti pārejas varbūtības pieraksta matricas formā:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix}$$

Matricu  $A$  sauc par pārejas varbūtību matricu, retāk par stohastisko matricu. Katrā matricas rindā atrodas visas pārejas varbūtības, kas atbilst stāvoklim  $i$ . Matricā katras rindas elementu summa ir vienāda ar 1:

$$\sum_{j=1}^N a_{ij} = 1. \quad (1.7)$$

Lai labāk varētu izprast Markova ķēdi un pārejas no viena stāvokļa uz citu, Markova ķēdes tiek uzskatāmi attēlotas grafā, kur katra virsotne atbilst stāvoklim no Markova ķēdes stāvokļu kopas  $S$ , savukārt virsotnes tiek savienotas ar bultiņām, uz tām norādot pārejas varbūtību vērtību starp attiecīgajiem stāvokļiem.



1.1. att. Stāvokļu pārejas grafs Markova ķēdei ar stāvokļiem  $S_1$  un  $S_2$

Pārejas varbūtību matrica šai Markova ķēdei izskatās šādi:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 - \lambda & \lambda \\ \mu & 1 - \mu \end{pmatrix}$$

## 1.2. Teorētiskais apraksts par slēptiem Markova modeļiem

Slēptā Markova modelī (SMM), atšķirībā no vienkāršā Markova procesa, kur visi stāvokļi ir zināmi un novērotājam redzami atklātā veidā, stāvokļi nav konkrēti redzami, bet ir redzami novērojumi, kas atkarīgi no šiem slēptajiem stāvokļiem. Šis process, ko sauc par SMM, ir stohastisks process, kas tiek balstīts uz kādu citu novērotājam apslēptu stohastisku procesu. Katram no šiem apslēptajiem stāvokļiem praktiski ir būtiska nozīme.

**Definīcija 6.** Par slēpto Markova modeli sauc modeli  $\lambda = \{A, B, \pi\}$ , kur :

1.  $S = \{S_1, S_2, \dots, S_N\}$  ir stāvokļu kopa, kur  $N$  ir stāvokļu skaits modelī;
2.  $V = \{V_1, V_2, \dots, V_M\}$  ir novērojuma simbolu kopa, kur  $M$  ir novērojuma simbolu izmērs;
3.  $Q = q_1 q_2 \dots q_T$  ir fiksēta garuma stāvokļu virkne;
4.  $O = O_1 O_2 \dots O_T$  ir stāvokļu virknei  $Q$  atbilstošā novērojumu virkne;
5.  $A = \{a_{ij}\}$  ir pārejas varbūtību sadalījums, kur  $a_{ij}$  ir varbūtība, ka modelis pāriet no stāvokļa  $i$  uz stāvokli  $j$ . Pārejas varbūtības nav atkarīgas no laika  $t$ :

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i). \quad (1.8)$$

Ja jebkuru stāvokli ir iespējams sasniegt no ikviena cita stāvokļa, tad  $a_{ij} > 0$  visiem  $i$  un visiem  $j$ .

6.  $B = \{b_j(k)\}$  ir novērojumu varbūtību matrica, kur  $b_j(k)$  ir varbūtība, ka novērojums  $k$ , neatkarīgi no laika  $t$ , parādīsies stāvoklī  $S_j$ :

$$b_j(k) = P(O_t = V_k | q_t = S_j). \quad (1.9)$$

7.  $\pi = \{\pi_i\}$  ir sākuma stāvokļu varbūtību sadalījums, kur

$$\pi_i = P(q_1 = S_i). \quad (1.10)$$

Pilnīgai slēpto Markova modeļu specifikācijai būtu jānorāda arī modeļa parametrus  $N$  un  $M$ , novērojumu simbolu alfabētu, kā arī trīs varbūtības  $A, B$  un  $\pi$ . Lai būtu ērtāk apzīmēt modeļa parametru kopu, darbā izmantots saīsinātais pieraksts:  $\lambda = \{A, B, \pi\}$ .

Slēptiem Markova modeļiem ir raksturīgas divas ļoti svarīgas īpašības:

1. Markova īpašība: katrs stāvoklis modelī ir atkarīgs tikai un vienīgi no iepriekšējā stāvokļa:

$$P(q_t = S_t | q_{t-1} = S_j, q_{t-2} = S_k, \dots) = P(q_t = S_t | q_{t-1} = S_j). \quad (1.11)$$

2. Ikviens novērojums modelī ir atkarīgs tikai un vienīgi no pašreizējā stāvokļa – tas nav atkarīgs ne no iepriekšējiem stāvokļiem, ne no iepriekšējiem novērojumiem modelī.

### 1.3. Slēpto Markova modeļu pamatproblēmas

Slēptiem Markova modeļiem eksistē trīs problēmas, kuras ir nepieciešams atrisināt, lai tie būtu izmantojami praktisku problēmu atrisināšanai:

- Novērtēšanas problēma: Ja ir dots modelis  $\lambda = \{A, B, \pi\}$  un novērojumu virkne  $O = O_1 O_2 \dots O_T$ , kāda ir varbūtība, ka šī novērojumu virkne  $O$  ir iegūta ar modeli  $\lambda$ ?
- Šifrēšanas problēma: Ja ir dots modelis  $\lambda = \{A, B, \pi\}$  un novērojumu virkne  $O = O_1 O_2 \dots O_T$ , kāda ir vistīcāmākā stāvokļu virkne  $Q = q_1 q_2 \dots q_T$ , kas izskaidro novērojumus?
- Apmācības problēma: Ja ir dots modelis  $\lambda = \{A, B, \pi\}$  un novērojumu virkne  $O = O_1 O_2 \dots O_T$ , kā izvēlēties modeļa parametrus  $(A, B, \pi)$ , lai maksimizētu varbūtību  $P(O|\lambda)$ ?

Apskatīsim sīkāk un analizēsim katru no trīs pamatproblēmām:

#### 1. Novērtēšanas problēma

Nepieciešams atrast varbūtību  $P(O|\lambda)$  novērojumu virknei  $O = O_1 O_2 \dots O_T$ , ņemot vērā, ka modelis ir  $\lambda = \{A, B, \pi\}$ . Šī varbūtība rāda to, cik labi jeb precīzi modelis paredz doto novērojumu virkni, tāpēc ar tās palīdzību var izvēlēties labāko no vairākiem modeļiem.

Varbūtība, ka tiks novērota virkne  $O$  pie fiksētas stāvokļu virknes  $Q$ , ir

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda) = b_{q_1}(O_1)b_{q_2}(O_2) \dots b_{q_T}(O_T) \quad (1.12)$$

un varbūtība stāvokļu virknei ir

$$P(Q|\lambda) = \pi_{q_1}a_{q_1q_2}a_{q_2q_3} \dots a_{q_{T-1}q_T}. \quad (1.13)$$

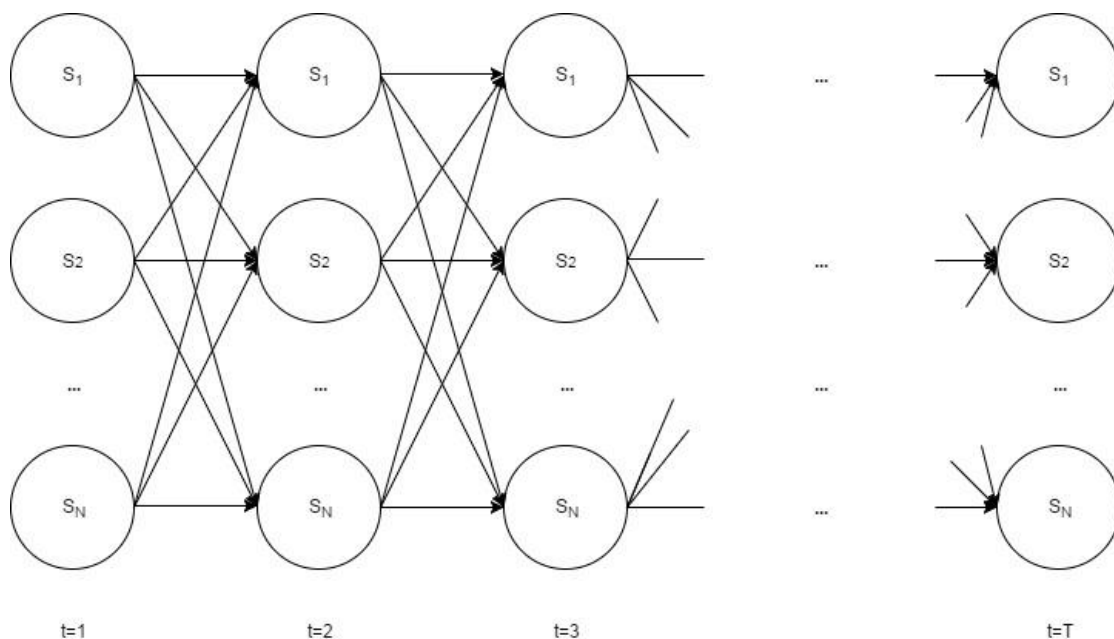
Kopējā varbūtība notikumiem  $Q$  un  $O$  (varbūtība, ka vienlaicīgi notiek gan notikums  $Q$ , gan notikums  $O$ ) ir vienādojumu (1.12) un (1.13) reizinājums:

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda). \quad (1.14)$$

Tātad varbūtība, ko dod modelis  $\lambda$  novērojumu virknei  $O$ , ir visu iespējamo stāvokļu virkņu  $Q$  varbūtību summa:

$$P(O|\lambda) = \sum_Q P(O|Q, \lambda)P(Q|\lambda) = \sum_{q_1 \dots q_T} \pi_{q_1}b_{q_1}(O_1)a_{q_1q_2}b_{q_2}(O_2) \dots a_{q_{T-1}q_T}b_{q_T}(O_T) \quad (1.15)$$

Ar formulu (1.15) ir iespējams aprēķināt varbūtību novērojumu virknei, taču tiešs aprēķins būtu ļoti neefektīvs, jo darbību skaits būtu  $2TN^T$ . Aprēķināt varbūtību pēc šīs formulas nebūtu praktiski pat tad, ja  $N$  un  $T$  vērtības būtu salīdzinoši mazas. Vērtīgi būtu atpazīt liekos aprēķinus un tos noteiktā veidā sagrupēt, lai varētu atvieglot uzdevumu.



1.2. att. Forward algoritma vizualizācija

Apskatām režģi (*attēls 1.2*), kur katra virsotne attēlo stāvokli laikā  $t$  un bultas attēlo pāreju uz jaunu stāvokli nākamajā laika momentā  $t + 1$ . Nepieciešams saglabāt informāciju režģī katrā laika momentā  $t$ , saglabāto informāciju katrā stāvoklī aprēķinot kā summu pa visiem stāvokļiem iepriekšējā laika momentā  $t - 1$ . Definēsim *turpejošo* (angļu val. – *Forward*) mainīgo  $\alpha_t(i)$  kā varbūtību stāvoklim  $S_i$  laika momentā  $t$  un daļējai stāvokļu virknei  $O_1 O_2 \dots O_t$  pie modeļa  $\lambda$ :

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = S_i | \lambda). \quad (1.16)$$

Lai būtu iespējams tikt cauri režģim, kas sastāv no  $\alpha_t(i)$  vērtībām, nepieciešamas iterācijas skaitā  $T$ . Iegūtās pēdējās režģa kolonnas elementu summa veido novērojuma virknes varbūtību. Šāds algoritms tiek saukts par *Forward* algoritmu.

**Definīcija 7.** *Forward* algoritms ir definēts sekojoši:

1. Priekšnosacījumi:

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \quad (1.17)$$

2. Indukcija:

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \sum_{i=1}^N \alpha_t(i) a_{ij}, 1 \leq t \leq T - 1, 1 \leq j \leq N \quad (1.18)$$

3. Iznākums:

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (1.19)$$

Galvenā algoritma priekšrocība ir tā, ka nepieciešams mazāks darbību skaits salīdzinājumā ar pilno pārlassi. Atpazīstot liekos aprēķinus un iegūstot  $\alpha_t(i)$  vērtības, *Forward* algoritms spēj samazināt aprēķinu sarežģītību no  $2TN^T$  uz  $N^2T$ .

Analoģiski *Forward* algoritmam varam definēt *atpakaļ ejošo* (angļu val. – *Backward*) algoritmu, kas darbojas ļoti līdzīgi, tikai pretējā virzienā, izmantojot *atpakaļ ejošo* jeb *Backward* mainīgo  $\beta_t(i)$ :

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = S_i, \lambda), \quad (1.20)$$

kas ir varbūtība fiksētai stāvokļu virknei no  $t + 1$  līdz  $T$ , kur  $S_i$  ir pašreizējais stāvoklis.

**Definīcija 8.** *Backward* algoritms ir definēts sekojoši:

1. Priekšnosacījumi:

$$\beta_t(i) = 1, 1 \leq i \leq N \quad (1.21)$$

2. Indukcija:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), 1 \leq i \leq N, t = T-1, T-2, \dots, 1 \quad (1.22)$$

3. Iznākums:

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(O_1) \beta_1(j). \quad (1.23)$$

Katram slēptajam Markova modelim  $\lambda$  jādod vienāds rezultāts varbūtībai  $P(O|\lambda)$  gan ar *Forward*, gan ar *Backward* algoritmu:

$$P(O|\lambda) = \sum_{i=1}^N P(O_1, O_2, \dots, O_t, q_t = S_i | \lambda) = \sum_{i=1}^N \alpha_t(i) \beta_t(i). \quad (1.24)$$

## 2. Šifrēšanas problēma

Nepieciešams atrast vistīcāmāko apslēpto stāvokļu virkni zināmajai novērojumu virknei. Viens veids, kā atrast vistīcāmāko ceļu, būtu veikt pilnu pārslāpī, atlasot varbūtības katram iespējamam ceļam caur režģi un izvēloties labāko no tiem, taču eksistē efektīvāki algoritmi, kā šo problēmu risināt – iespējams izmantot *Viterbi* algoritmu, lai atrastu vienu labāko stāvokļu virkni, kas dotu jau zināmo novērojumu virkni. *Viterbi* algoritms ir vēl viens režģa algoritms, kas ir ļoti līdzīgs jau iepriekš minētajam *Forward* algoritmam, izņemot to, ka šajā algoritmā pārejas varbūtības katrā posmā tiek maksimizētas, nevis summētas, kā to dara *Forward* algoritmā.

Lai atrastu vienu labāko stāvokļu virkni  $Q = q_1 q_2 \dots q_T$  dotajai novērojumu virknei  $O = O_1 O_2 \dots O_T$ , nepieciešams definēt mainīgo, kas dod vislielāko varbūtību  $\delta_t(i)$ , kas ir lielākā varbūtība vienam ceļam laika momentā  $t$ , kas veido pirmos  $t$  novērojumus un beidzas stāvoklī  $S_i$ :

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t = S_i, O_1, O_2, \dots, O_t | \lambda) \quad (1.25)$$

Pēc indukcijas iespējams secināt:

$$\delta_{t+1}(j) = b_j(O_{t+1}) \max_{i=1, \dots, N} \delta_t(i) a_{ij}, 1 \leq j \leq N, 1 \leq t \leq T-1, \quad (1.26)$$

kur  $\delta_1(j) = \pi_j b_j(O_1)$ ,  $1 \leq j \leq N$ . Lai aprēķinātu ticamāko stāvokļu virkni, nepieciešams sākt ar  $\delta_T(j)$ ,  $1 \leq j \leq N$ . Ja tiek meklēts maksimums, tad vienmēr atzīmē “uzvaras” stāvokli. Rezultātā tiek iegūts stāvoklis  $j^*$ :

$$j^* = \arg \max_{i=1, \dots, N} \delta_T(j) \quad (1.27)$$

Darbības tiek sāktas no šī stāvokļa un atmuguriski tiek atrasta stāvokļu virkne, ko sauc par atzīmju virkni un ko uzrāda katrs “uzvaras” stāvoklis – tādā veidā tiek iegūta nepieciešamā stāvokļu kopa.

Lai iegūtu stāvokļu secību, nepieciešams katram stāvoklim režģī atcerēties argumentu, kas maksimizē vienādību (1.25), tāpēc tiek izmantota atpakaļceļa norāde, kas tiek apzīmēta ar  $\psi_t(j)$ .

**Definīcija 9.** Viterbi algoritms ir definēts sekojoši:

1. Priekšnosacījumi:

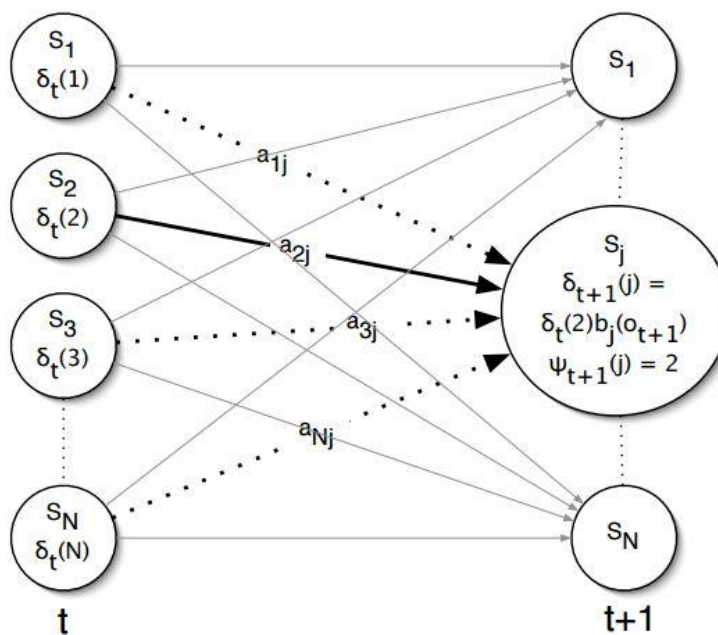
$$\delta_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \quad (1.28)$$

$$\psi_1(i) = 0 \quad (1.29)$$

2. Rekursija:

$$\delta_t(j) = \max_{i=1,\dots,N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), 1 \leq j \leq N, 2 \leq t \leq T \quad (1.30)$$

$$\psi_t(j) = \operatorname{argmax}_{i=1,\dots,N} [\delta_{t-1}(i) a_{ij}], 1 \leq j \leq N, 2 \leq t \leq T \quad (1.31)$$



1.3. att. Viterbi algoritma rekursijas solis

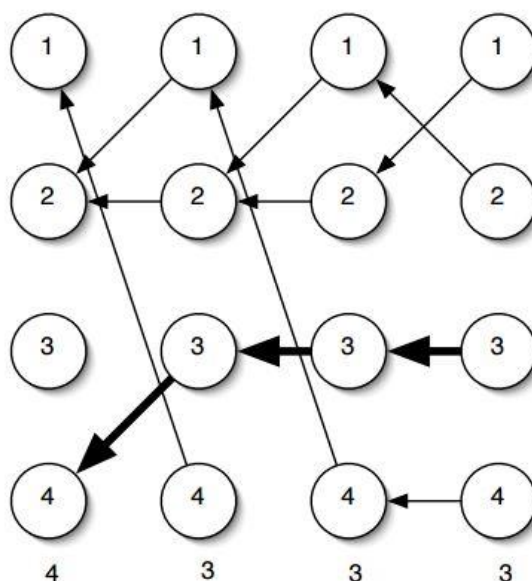
3. Iznākums:

$$P^* = \max_{i=1,\dots,N} [\delta_T(i)] \quad (1.32)$$

$$q_T^* = \operatorname{argmax}_{i=1,\dots,N} [\delta_T(i)] \quad (1.33)$$

#### 4. Atpakaļceļš (angļu val. – *path backtracking*)

$$q_t^* = \psi_{t+1}(i)q_{t+1}^* \quad (1.34)$$



1.4. att. Viterbi algoritma atpakaļceļa solis

Rekursijas solī, attēlā 1.3, galvenā atšķirība no *Forward* algoritma ir tā, ka tiek maksimizētas, nevis summētas pārejas varbūtības un tiek uzglabāts tas stāvoklis, kurš tika izvēlēts kā maksimums un tiek izmantots kā atpakaļceļa norāde. Atpakaļceļa solis, attēls 1.4, no rekursijas solī iegūtajām un uzglabātajām atpakaļceļa norādēm ļauj atrast vienu labāko stāvokļu virkni, bet jāatzīmē, ka nav zināms viegls veids, kā atrast otru labāko stāvokļu virkni.

*Viterbi* algoritma priekšrocības:

- algoritms nodrošina labu doto novērojumu interpretāciju;
- tiek samazināta aprēķinu sarežģītība, izmantojot rekursiju. Ieguvums salīdzinājumā ar tiešo aprēķinu ir analogs *Forward* un *Backward* algoritmiem – sarežģītība samazinās no  $2TN^T$  uz  $N^2T$ .

### 3. Apmācības problēma

Trešās un arī sarežģītākās slēpto Markova modeļu problēmas mērķis ir novērtēt un uzlabot modeļa parametrus: stāvokļu sākuma varbūtības, stāvokļu pārejas varbūtības, kā arī novērojumu varbūtības tā, lai maksimizētu dotās novērojumu virknes varbūtību.

Modeļa parametrus ir iespējams novērtēt ar vislielākās ticamības (angļu val. – *maximum likelihood*) metodi. Modeļa ticamību neatkarīgu novērojumu virknei  $O$  var aprēķināt šādi:

$$P(O|\lambda) = \prod_{t=1}^T P(O_t|\lambda) \quad (1.35)$$

Tā kā vienādojums (1.35) ir nelineāra parametru  $\lambda$  funkcija, tad analītiskā ceļā nav iespējams atrast vislielāko ticamību, bet, izmantojot iteratīvu atkārtotās novērtēšanas metodi, kā piemēram *Baum-Welch* algoritmu, ir iespējams atrast lokālo maksimumu varbūtībai  $P(O|\lambda)$ .

*Baum-Welch* metodes pamatideja ir sekojoša: no sākotnējā modeļa  $\lambda$  iegūt tādu novērtēto modeli  $\bar{\lambda}$ , kam izpildās  $P(O|\bar{\lambda}) \geq P(O|\lambda)$ . Process tiek atkārtots tik ilgi līdz varbūtība  $P(O|\bar{\lambda})$  konverģē.

Nepieciešams definēt lielumu  $\xi_t(i, j)$ , kas ir varbūtība atrasties stāvoklī  $S_i$  pie laika  $t$ , un atrasties stāvoklī  $S_j$  pie laika  $t + 1$ , ja ir dots modelis  $\lambda$  un novērojumu virkne  $O$ :

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_{i+1} | O, \lambda) \quad (1.36)$$

No *forward* un *backward* mainīgo definīcijām seko, ka  $\xi_t(i, j)$  var izteikt sekojoši:

$$\begin{aligned} \xi_t(i, j) &= \frac{P(q_t = S_i, q_{t+1} = S_{i+1} | O, \lambda)}{P(O|\lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} = \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (1.37)$$

Definējam  $\gamma_t(i)$  kā varbūtību laika momentā  $t$  atrasties stāvoklī  $S_i$ , ja ir dota novērojumu virkne  $O$  un modelis  $\lambda$ :

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \quad (1.38)$$

Tad šo varbūtību var izteikt kā  $\xi_t(i, j)$  summu visiem  $j$ :

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (1.39)$$

Saskaitot visus  $\gamma_t(i)$  laikā  $t$ , iegūstam lielumu, kas rāda to, cik reizes ir nonākts stāvoklī  $S_i$  jeb veikto pāreju skaitu no stāvokļa  $S_i$ , bet, saskaitot visus  $\xi_t(i, j)$  laikā  $t$ , iegūstam pāreju skaitu starp stāvokļiem  $S_i$  un  $S_j$ :

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{sagaidāmais pāreju skaits no stāvokļa } S_i \quad (1.40)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{sagaidāmais pāreju skaits starp stāvokļiem } S_i \text{ un } S_j \quad (1.41)$$

Ar šo formulu palīdzību iespējams iegūt metodi parametru atkārtotai novērtēšanai (angļu val. – *reestimation*), lai novērtētu  $A, B$  un  $\pi$ :

$$\begin{aligned}\bar{a}_{ij} = P(S_i|S_j) &= \frac{\text{sagaidāmais pāreju skaits starp stāvokļiem } S_i \text{ un } S_j}{\text{sagaidāmais pāreju skaits no stāvokļa } S_i} = \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}\end{aligned}\quad (1.42)$$

$$\begin{aligned}\bar{b}_j(k) = P(V_k|S_i) &= \\ &= \frac{\text{sagaidāmais reižu skaits, kad novērojam simbolu } v_k \text{ stāvoklī } S_j}{\text{sagaidāmais reižu skaits, kad modelis ir stāvoklī } S_j} = \\ &= \frac{\sum_{t=1}^T \gamma_t(j), \text{ kur } O_t = v_k}{\sum_{t=1}^T \gamma_t(j)}\end{aligned}\quad (1.43)$$

$$\bar{\pi}_i = P(S_1) = \text{sagaidāmais atrašanās biežums stāvoklī } S_i \text{ laikā } (t = 1) = \gamma_1(i) \quad (1.44)$$

Ja pašreizējais modelis ir  $\lambda = \{A, B, \pi\}$ , tad no vienādojumiem (1.42), (1.43) un (1.44) var iegūt novērtēto modeli  $\bar{\lambda} = \{\bar{A}, \bar{B}, \bar{\pi}\}$ . Šādā gadījumā pastāv divas iespējas:

1. Sākotnējais modelis  $\lambda$  ir kritiskais punkts ticamības funkcijā, tātad  $\bar{\lambda} = \lambda$ ;
2. Modelis  $\bar{\lambda}$  ir ticamāks par sākotnējo modeli  $\lambda$  (izpildās  $P(O|\bar{\lambda}) > P(O|\lambda)$ ), kas nozīmē, ka ir atrasts jauns modelis  $\bar{\lambda}$ , kas ir ticamāks, balstoties uz doto novērojumu virkni.

Iepriekš aprakstītajā procedūrā modelis  $\lambda$  tiek iteratīvi aizstāts ar modeli  $\bar{\lambda}$ , uzlabojot varbūtību, ka tiks novērota virkne  $O$ , līdz brīdim, kad to tālāk uzlabot nav iespējams. Šīs procedūras gala rezultāts tiek saukts par slēpto Markova modeļu maksimālās ticamības novērtējumu. Jāņem vērā tas, ka ar atkārtotās novērtēšanas algoritmu iespējams atrast tikai lokālos maksimuma punktus, taču nereti praktiskajos uzdevumos optimizācijas problēmai eksistē vairāki lokālie maksimumi.

Iepriekš minētās atkārtotās novērtēšanas formulas (1.42) – (1.44) var tikt iegūtas arī tad, ja, izmantojot standarta optimizācijas metodes, tiek maksimizēta Brauna palīgfunkcija

$$Q(\lambda, \bar{\lambda}) = \sum_Q P(Q|O, \lambda) \log[P(Q, O|\bar{\lambda})] \quad (1.45)$$

priekš modeļa  $\bar{\lambda}$ .

Ir pierādīts, ka maksimizējot šo palīgfunkciju  $Q(\lambda, \bar{\lambda})$ , tiek iegūta lielāka ticamība:

$$\max_{\bar{\lambda}} Q(\lambda, \bar{\lambda}) \Rightarrow P(O|\bar{\lambda}) \geq P(O|\lambda), \quad (1.46)$$

tāad šī ticamības funkcija konverģē uz kritisko punktu.

Atkārtotās novērtēšanas procedūras laikā katrā atkārtotās novērtēšanas iterācijā stohastiskie ierobežojumi slēpto Markova modeļu parametriem izpildās automātiski:

$$\sum_{i=1}^N \bar{\pi}_i = 1 \quad (1.47)$$

$$\sum_{j=1}^N \bar{a}_{ij} = 1, 1 \leq i \leq N \quad (1.48)$$

$$\sum_{k=1}^M \bar{b}_j(k) = 1, 1 \leq j \leq N \quad (1.49)$$

Ja parametru novērtēšanas problēma tiek apskatīta kā varbūtības  $P(O|\lambda)$  optimizācijas problēma ar ierobežojumiem (1.47) – (1.49), tad, izmantojot Lagranža reizinātāju metodi, iespējams atrast tādas parametru  $\pi_i$ ,  $a_{ij}$  un  $b_j(k)$  vērtības, kas maksimizē  $P(O|\lambda)$ . Veicot standarta Lagranža optimizāciju un izmantojot Lagranža reizinātāju metodi, iespējams pierādīt, ka varbūtība  $P(O|\lambda)$  ir maksimizēta, ja:

$$a_{ij} = \frac{a_{ij} \frac{\partial P}{\partial a_{ij}}}{\sum_{k=1}^N a_{ik} \frac{\partial P}{\partial a_{ij}}} \quad (1.50)$$

$$b_j(k) = \frac{b_j(k) \frac{\partial P}{\partial b_j(k)}}{\sum_{l=1}^M b_j(l) \frac{\partial P}{\partial b_j(l)}} \quad (1.51)$$

$$\pi_i = \frac{\pi_i \frac{\partial P}{\partial \pi_i}}{\sum_{k=1}^N \pi_k \frac{\partial P}{\partial \pi_k}} \quad (1.52)$$

Vienādojumu (1.50) – (1.52) labās puses iespējams pārveidot tā, lai tās sakristu ar vienādojumu (1.42) – (1.44) labajām pusēm, tādā veidā parādot, ka atkārtotās novērtēšanas formulas kritiskajos punktos ir patiesas. Šī metode var tikt interpretēta kā EM (angļu val. – *expectation-maximisation*) algoritms, kur E (angļu val. – *expectation*) solis ir Brauna

palīgfunkcijas aprēķināšana, savukārt M (angļu val. – *maximisation*) solis ir funkcijas maksimizēšana modelim  $\bar{\lambda}$ .

## 2. SLĒPTO MARKOVA MODEĻU PIELIETOJUMS AKCIJU CENU PROGNOZĒŠANĀ

### 2.1. Modeļa pielietojums un nepieciešamo aprēķinu veikšana

Šajā nodaļā tiks izmantots avots [6].

Lai varētu strādāt ar izvēlēto modeli, tika izvēlēti trīs uzņēmumi, kas šobrīd ir aktīvi un kuru akcijas ir akciju tirgū: *Apple Inc*, *Microsoft Corporation* un *Samsung Group*. Tiks izmantots slēptais Markova modelis, lai prognozētu akciju cenas un salīdzinātu šīs prognozes ar patiesajām akciju cenām.

Prognozēšanas process var tikt sadalīts trīs soļos:

1. Notiek parametru kalibrēšana un tiek aprēķināta modeļa varbūtība;
2. Tiek atrasta diena pagātnē, kurai modeļa varbūtība ir līdzīga varbūtībai, ko dod modelis pašreizējai dienai;
3. Tiek izmantota atšķirība akciju cenās starp pašreizējo dienu un dienu, kurā modeļa varbūtība ir līdzīga, lai prognozētu nākotnes akciju cenas.

Šī prognozēšanas pieeja ir balstīta uz indiešu matemātikas profesoru Rafiqul Hassan un Bijan Nath darbu, tomēr darbā apskatītā procedūra atšķiras no profesoru metodes ar to, ka mēs ņemsim vērā arī varbūtību atšķirības zīmi, lai prognozētu akciju cenas. Mēs izmantosim vienu novērojumu – slēgšanas cenu (angļu val. – *close price*) attiecīgajā dienā, lai prognozētu nākamās dienas slēgšanas cenu.

Pieņemsim, ka vēlamies prognozēt rītdienas slēgšanas cenu kāda uzņēmuma akcijām  $A$ . Pirmajā solī tiek izvēlēta datu kopa ar  $T$  slēgšanas cenām ( $O = \{O_t, t = T - 99, t - 98, \dots, T\}$ ), lai kalibrētu slēptā Markova modeļa parametrus,  $\lambda$ , un tad aprēķinātu novērojumu varbūtību  $P(O|\lambda)$ .

Iesākumā nepieciešams iegūt ieejas parametrus modelim  $\lambda = \{A, B, \pi\}$ . Tos iegūstam, izmantojot programmu Microsoft Excel, kur sākumā ielasām datus, kurus ieguvu interneta vietnē. [7] Tālāk atrodam slēgšanas cenu izmaiņu starp laika momentiem  $t$  un  $t - 1$ . Pēc tam, izmantojot funkciju  $IF$ , katram novērojumam tiek piekārtots attiecīgais stāvoklis:

- Stāvoklis “A” – akciju cena laika momentā  $t$  aug, salīdzinājumā ar akciju cenu laika momentā  $t - 1$ .
- Stāvoklis “D” – akciju cena laika momentā  $t$  dilst, salīdzinājumā ar akciju cenu laika momentā  $t - 1$ .
- Stāvoklis “N” – akciju cena nemainās.

Šeit jāpiemin, ka darbā izmantotajā laika periodā uzņēmuma *Apple Inc* akciju cena nevienu dienu nebija nemainīga, tātad, prognozējot šī uzņēmuma akcijas, tika izmantoti nevis trīs, bet divi stāvokļi, attiecīgi stāvoklis “A” un stāvoklis “D”. Lai iegūtu pārejas varbūtību matricu  $A$ , iesākumā ar funkciju *COUNTIF* nepieciešams saskaitīt, cik reizes novērojumu virknē parādās katrs stāvoklis. Tad ar funkciju (*IF(AND)*) iespējams noskaidrot, vai pēc stāvokļa “A” seko stāvoklis “A” vai stāvoklis “D” vai stāvoklis “N”. Tāpat iespējams noskaidrot, vai pēc stāvokļa “D” seko stāvoklis “A” vai “D” vai “N”, tādas pašas darbības tiek veiktas arī stāvoklim “N”. Tiek iegūts skaits, cik reizes veidojas kombinācija “AA”, “AD”, “AN”, “DA”, “DD”, “DN”, “NA”, “ND”, “NN”. No šiem iegūtajiem rezultātiem viegli izveidot sākuma pārejas varbūtību matricu  $A$ . Sākuma novērojumu varbūtību matricu  $B$  varam iegūt, izpētot, cik novērojumos katrs stāvoklis izdod konkrētu simbolu. Izmantojam šādus simbolus:

1, ja slēgšanas cenu izmaiņa starp laika momentiem  $t$  un

$$t - 1 \geq \text{vidējā cenu izmaiņa} \quad (2.1)$$

2, ja slēgšanas cenu izmaiņa starp laika momentiem  $t$  un

$$t - 1 < \text{vidējā cenu izmaiņa} \quad (2.2)$$

3, ja slēgšanas cenu izmaiņa starp laika momentiem  $t$  un

$$t - 1 = 0 \text{ (tātad cena starp attiecīgajiem laika momentiem nemainās)} \quad (2.3)$$

Sākuma varbūtības atrasties stāvoklī “A”, stāvoklī “D” vai stāvoklī “N” pieņemu vienādas, t.i. katram stāvoklim tā ir  $\frac{1}{3}$ , kas arī veido matricu  $\pi$ .

Microsoft Excel izveidoto failu skatīt pielikumā nr. 1.

Tālāk statistisko aprēķinu programmā  $R$  tiek ievadīti nepieciešamie dati un iesākumā tiek izveidots slēptais Markova modelis, izmantojot pakotnes (angļu val. – *package*) “*HMM*” funkciju “*initHMM*”, pēc tam tiek veikta modeļa parametru kalibrēšana, izmantojot šīs pašas pakotnes funkciju “*baumWelch*”. Uzlabotie modeļa parametri tālāk tiek izmantoti kā ieejas parametri, lai aprēķinātu modeļa varbūtību  $P(O|\lambda)$ . Ar funkciju “*forward*” tiek iegūtas “uz priekšu” varbūtības un, saskaitot pēdējās kolonnas visus elementus, iegūstam meklēto modeļa varbūtības vērtību.

Otrajā solī datu kopa tiek pabīdīta atpakaļ pa vienu dienu, lai iegūtu jaunu novērojumu datu kopu  $O^{jauna} = \{O_t, t = T - 100, T - 99, \dots, T - 1\}$  un atkal ar Microsoft Excel un programmu  $R$  aprēķinātu varbūtību  $P(O^{jauna}|\lambda)$ . Tā turpinām nobīdīt datu kopu atpakaļ pa

vienai dienai līdz atrodam datu kopu  $O^* = \{O_t, t = T^* - 99, T^* - 98, \dots, T^*\}$  tādu, kam izpildās  $P(O^*|\lambda) \approx P(O|\lambda)$ .

Pēdējā solī mēs aprēķinām akciju slēgšanas cenu laikā  $T + 1$ ,  $O_{T+1}$ , izmantojot formulu

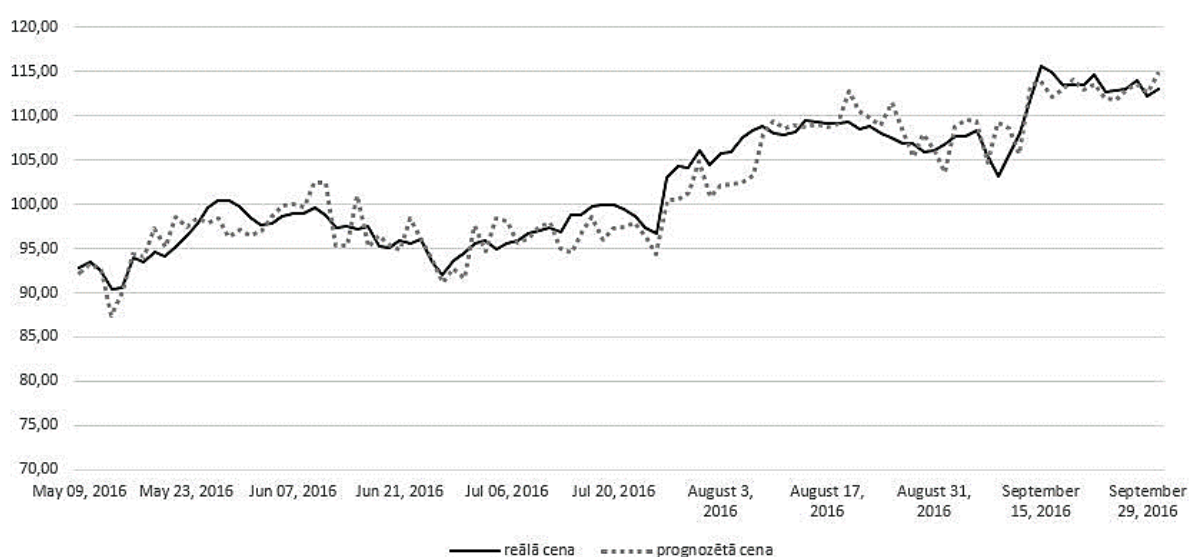
$$O_{T+1} = O_T + (O_{T^{**+1}} - O_{T^*}) \cdot \text{zīme}(P(O^*|\lambda) - P(O|\lambda)) \quad (2.4)$$

Pēc pirmās prognozēšanas akciju slēgšanas cenai laikā  $T + 1$  mēs Microsoft Excel tabulā atjaunojam novērojumu datu kopu  $O$ , nobīdot to vienu dienu uz priekšu,  $O = \{O_t, t = T - 98, T - 97, \dots, T + 1\}$ , lai atkal atkārtotu visas nepieciešamās darbības un prognozētu akciju slēgšanas cenu nākamajā dienā (laikā  $T + 2$ ). Kalibrētie slēptā Markova modeļa parametri pirmajā prognozē tiek izmantoti kā sākuma parametri otrajai prognozei. Turpinām atkārtot šo trīs soļu prognozēšanu otrajai prognozei un tā tālāk.

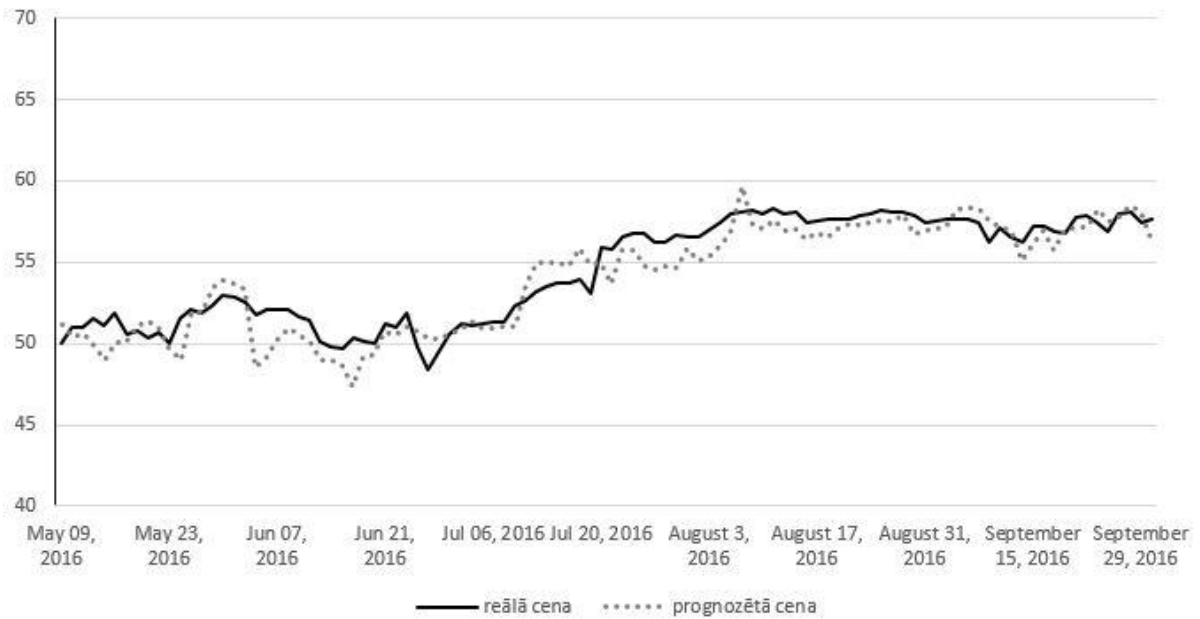
Izveidoto programmas R kodu skatīt 2. pielikumā.

## 2.2. Iegūto prognožu rezultāti un precizitāte

Tiks apskatīti iegūtie rezultāti, izmantojot divu stāvokļu slēpto Markova modeli uzņēmuma *Apple Inc* akcijām un izmantojot trīs stāvokļu slēpto Markova modeli uzņēmumu *Microsoft Corporation* un *Samsung Group* akcijām. Akciju cenas tika prognozētas 100 dienām un iegūtie rezultāti ir attēloti attēlos 2.1 - 2.3.



2.1. att. Divu stāvokļu SMM prognoze uzņēmuma *Apple Inc* akciju slēgšanas cenām no 2016. gada 9. maija līdz 2016. gada 30. septembrim



**2.2. att. Trīs stāvokļu SMM prognoze uzņēmuma *Microsoft Corporation* akciju slēgšanas cenām no 2016. gada 9. maija līdz 2016. gada 30. septembrim**



**2.3. att. Trīs stāvokļu SMM prognoze uzņēmuma *Samsung Group* akciju slēgšanas cenām no 2016. gada 9. maija līdz 2016. gada 30. septembrim**

Lai varētu izdarīt secinājumus ne tikai no iegūtajām līniju diagrammām, tika aprēķināta vidējā absolūtā procentuālā kļūda, MAPE:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|M_i - P_i|}{M_i}, \quad (2.5)$$

kur  $N$  – prognozēto novērojumu skaits;

$M$  – reālā cena;

$P$  – prognozētā cena.

Iegūtās vidējās absolūtās procentuālās kļūdas apkopotas tabulā 2.1.

2.1. tabula

**MAPE salīdzinājums akciju cenu prognozēšanai uzņēmumiem *Apple Inc*, *Microsoft Corporation* un *Samsung Group***

Uzņēmums, kura akciju cenas tika prognozētas	MAPE
Apple Inc	0,017
Microsoft Corporation	0,018
Samsung Group	0,020

No iegūtajiem rezultātiem iespējams redzēt, ka akciju tirgus cenas ir iespējams prognozēt ar samērā labu precizitāti. Iespējams secināt, ka no darbā izvēlētajām uzņēmumu akcijām vislabāk slēptais Markova modelis strādā uz kompānijas *Apple Inc* akciju cenām.

## SECINĀJUMI

Akciju cenu svārstības ir svarīgs rādītājs akciju tirgus aktivitātes veidošanā, kā arī tas atspoguļo visu ekonomiku kopumā. Ir daudz faktoru, kas ietekmē akciju cenu izmaiņas augšup vai lejup.

Šajā darbā ir izmantoti slēptie Markova modeļi, SMM, lai prognozētu akciju tirgus cenas trim uzņēmumiem. Tika izvēlēti trīs uzņēmumi, lai būtu iespējams izdarīt korektākus secinājumus un būtu vairāk informācijas, uz kā balstīt secinājumus. Akciju cenu prognozes tika ievietotas diagrammā, attēlojot prognozēto akciju cenu salīdzinājumā ar reālo akciju cenu noteiktajā laika momentā, kā arī tika izrēķināta vidējā absolūtā procentuālā kļūda, lai varētu spriest, cik korekti šīs akciju cenas tiek prognozētas ar slēptiem Markova modeļiem.

Rezultāti parāda, ka vislabāk izdevies prognozēt uzņēmuma *Apple Inc* akciju cenas, ko, iespējams, ietekmē izvēlētais stāvokļu skaits modelī, nākamās labākās prognozes ir iegūtas uzņēmuma *Microsoft Corporation* akciju cenām, taču nedaudz nesekmīgāk izdevies prognozēt uzņēmuma *Samsung Group* akciju cenas.

Šajā pieejā, izmantojot slēptos Markova modeļus, lai prognozētu akciju tirgus cenas, tiek pieņemts, ka viena uzņēmuma akcijas ir neatkarīgas no cita uzņēmuma akcijām, taču realitātē ir ļoti iespējams, ka noteiktu uzņēmumu akcijas savā starpā korelē, varbūt pat cieši. Skatoties nākotnē, būtu jāmēģina izveidot modelis, kas šīs korelācijas ņem vērā. Uzlabot varētu arī pašu novērojumu biežumu, ņemot akciju cenas stundu pa stundai, vai pat minūti pa minūtei. Tas, saprotams, būtu sarežģītāks process, bet intuitīvi tas noteikti dotu precīzākus prognozēšanas rezultātus.

Lai darbu varētu uzlabot, visus sākuma aprēķinus būtu nepieciešams veikt automatizēti, piemēram, statistisko aprēķinu programmā R, kur tas noteikti būtu ne tik laikietilpīgi kā to darīt Microsoft Excel vidē. Būtu nepieciešams sīkāk izpētīt, kā stāvokļu skaits modelī ietekmē prognozēšanas precizitāti, kā arī noskaidrot, vai to ietekmē novērojumu skaits vienā laika momentā, t.i. vai ir pietiekami, ka par vienas dienas novērojumu tiek izmantota tikai slēgšanas cena, vai tomēr labāk izmantot vairākus novērojumus, piemēram, slēgšanas cenu, atvēršanas cenu, maksimālo cenu dienas laikā, kā arī minimālo cenu dienas laikā, tomēr pēc šī darba rezultātiem iespējams secināt, ka prognozēšana, izmantojot slēptos Markova modeļus, ir potenciāla metode, kā noteikt akciju tirgus cenas nākotnei, balstoties uz dažādiem rādītājiem un var redzēt, ka rezultāti pietiekami labi atspoguļo akciju tirgus cenu tendences. Visticamāk, ka, papildinot modeli ar vairākiem novērojumiem vienā dienā vai pamainot slēpto stāvokļu

skaitu, akciju tirgus cenas būtu iespējams prognozēt vēl precīzāk un tas būtu ļoti interesējoši visiem uzņēmumiem un jebkuram aktīvam ekonomikas tirgus dalībniekam.

## IZMANTOTĀ LITERATŪRA UN AVOTI

- [1] A.S. Weigend and S. Shi. “Predicting daily probability distributions of s&p500 returns”. Journal of Forecasting, pages 375–392, 2000.  
[http://onlinelibrary.wiley.com/doi/10.1002/1099-131X\(200007\)19:4%3C375::AID-FOR779%3E3.0.CO;2-U/epdf](http://onlinelibrary.wiley.com/doi/10.1002/1099-131X(200007)19:4%3C375::AID-FOR779%3E3.0.CO;2-U/epdf), skatīts 11.05.2017.
- [2] G. L. Kouemou. History and Theoretical Basics of Hidden Markov Models, 2010  
<http://cdn.intechweb.org/pdfs/15369.pdf>, skatīts 13.05.2017.
- [3] Geoffrey Grimmet and David Stirzaker. One Thousand Exercises in Probability. Oxford University Press, 2001.  
[https://eclass.uop.gr/modules/document/file.php/TST244/%5BGeoffrey\\_Grimmett,\\_David\\_Stirzaker%5D\\_One\\_Thousand\\_\(BookZZ.org\).pdf](https://eclass.uop.gr/modules/document/file.php/TST244/%5BGeoffrey_Grimmett,_David_Stirzaker%5D_One_Thousand_(BookZZ.org).pdf), skatīts 13.05.2017.
- [4] Lawrence Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77, 1989  
<http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/tutorial%20on%20hmm%20and%20applications.pdf>, skatīts 11.05.2017.
- [5] P. Blunsom. Hidden Markov Models, 2004.  
<http://digital.cs.usu.edu/~cyan/CS7960/hmm-tutorial.pdf>, skatīts 23.05.2017.
- [6] Nguyet Nguyen. Stock Price Prediction using Hidden Markov Model. Youngstown State University, Ohio, USA, 2016  
[https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=SILC2016&paper\\_id=38](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SILC2016&paper_id=38), skatīts 09.05.2017.
- [7] <https://www.investing.com/equities/> , skatīts 14.05.2017.

## **PIELIKUMI**

Pielikums nr. 1  
Aprēķini programmā Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	cieta	izmaiņa	stavoklis	abs	simboli					A-1	D-1				AA	AD	AN		DD	DA	DN		NN	NA	ND
2	May 09, 2016	92,79								0	0				0	0	0		0	0	0		0	0	0
3	May 10, 2016	93,42	0,63 A	0,63	2		54 <-cik A			0	0				0	1	0		0	0	0		0	0	0
4	May 11, 2016	92,51	-0,91 D	0,91	2					0	0				0	0	0		1	0	0		0	0	0
5	May 12, 2016	90,34	-2,17 D	2,17	1					0	1				0	0	0		0	1	0		0	0	0
6	May 13, 2016	90,52	0,18 A	0,18	2					0	0				1	0	0		0	0	0		0	0	0
7	May 16, 2016	93,88	3,36 A	3,36	1					1	0				0	1	0		0	0	0		0	0	0
8	May 17, 2016	93,49	-0,39 D	0,39	2					0	0				0	0	0		0	1	0		0	0	0
9	May 18, 2016	94,56	1,07 A	1,07	1		46 <-cik D			1	0				0	1	0		0	0	0		0	0	0
10	May 19, 2016	94,20	-0,36 D	0,36	2					0	0				0	0	0		0	1	0		0	0	0
11	May 20, 2016	95,22	1,02 A	1,02	1					1	0				1	0	0		0	0	0		0	0	0
12	May 23, 2016	96,43	1,21 A	1,21	1					1	0				1	0	0		0	0	0		0	0	0
13	May 24, 2016	97,90	1,47 A	1,47	1					1	0				1	0	0		0	0	0		0	0	0
14	May 25, 2016	99,62	1,72 A	1,72	1					1	0				1	0	0		0	0	0		0	0	0
15	May 26, 2016	100,41	0,79 A	0,79	2		0 <-cik N			0	0				0	1	0		0	0	0		0	0	0
16	May 27, 2016	100,35	-0,06 D	0,06	2					0	0				0	0	0		1	0	0		0	0	0
17	May 31, 2016	99,86	-0,49 D	0,49	2					0	0				0	0	0		1	0	0		0	0	0
18	Jun 01, 2016	98,46	-1,4 D	1,4	1					0	1				0	0	0		1	0	0		0	0	0
19	Jun 02, 2016	97,72	-0,74 D	0,74	2					0	0				0	0	0		0	1	0		0	0	0
20	Jun 03, 2016	97,92	0,2 A	0,2	2					0	0				1	0	0		0	0	0		0	0	0
21	Jun 06, 2016	99,63	0,71 A	0,71	2					0	0				1	0	0		0	0	0		0	0	0
22	Jun 07, 2016	99,03	0,4 A	0,4	2					0	0				0	1	0		0	0	0		0	0	0
23	Jun 08, 2016	98,94	-0,09 D	0,09	2					0	0				0	0	0		0	1	0		0	0	0
24	Jun 09, 2016	99,65	0,71 A	0,71	2					0	0				0	1	0		0	0	0		0	0	0
25	Jun 10, 2016	98,83	-0,82 D	0,82	2					0	0				0	0	0		1	0	0		0	0	0
26	Jun 13, 2016	97,34	-1,49 D	1,49	1					0	1				0	0	0		0	1	0		0	0	0
27	Jun 14, 2016	97,46	0,12 A	0,12	2					0	0				0	1	0		0	0	0		0	0	0
28	Jun 15, 2016	97,14	-0,32 D	0,32	2					0	0				0	0	0		0	1	0		0	0	0
29	Jun 16, 2016	97,55	0,41 A	0,41	2					0	0				0	1	0		0	0	0		0	0	0
30	Jun 17, 2016	95,33	-2,22 D	2,22	1					0	1				0	0	0		1	0	0		0	0	0
31	Jun 20, 2016	95,10	-0,23 D	0,23	2					0	0				0	0	0		0	1	0		0	0	0
32	Jun 21, 2016	95,57	0,47 A	0,47	2					0	0				0	0	0		0	0	0		0	0	0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
19	August 26, 2016	106,94	-0,63 D	0,63	2					0	0				0	0	0		1	0	0		0	0	0
20	August 29, 2016	106,82	-0,12 D	0,12	2					0	0				0	0	0		1	0	0		0	0	0
21	August 30, 2016	106,00	-0,82 D	0,82	2					0	0				0	0	0		0	1	0		0	0	0
22	August 31, 2016	106,10	0,1 A	0,1	2					0	0				1	0	0		0	0	0		0	0	0
23	September 1, 2016	106,73	0,63 A	0,63	2					0	0				1	0	0		0	0	0		0	0	0
24	September 2, 2016	107,73	1 A	1	1					1	0				0	1	0		0	0	0		0	0	0
25	September 6, 2016	107,70	-0,03 D	0,03	2					0	0				0	0	0		0	1	0		0	0	0
26	September 7, 2016	108,36	0,66 A	0,66	2					0	0				0	1	0		0	0	0		0	0	0
27	September 8, 2016	105,32	-2,84 D	2,84	1					0	1				0	0	0		1	0	0		0	0	0
28	September 9, 2016	103,13	-2,39 D	2,39	1					0	1				0	0	0		0	1	0		0	0	0
29	September 12, 2016	105,44	2,31 A	2,31	1					1	0				1	0	0		0	0	0		0	0	0
30	September 13, 2016	107,95	2,51 A	2,51	1					1	0				1	0	0		0	0	0		0	0	0
31	September 14, 2016	111,77	3,82 A	3,82	1					1	0				1	0	0		0	0	0		0	0	0
32	September 15, 2016	115,37	3,8 A	3,8	1					1	0				0	1	0		0	0	0		0	0	0
33	September 16, 2016	114,92	-0,65 D	0,65	2					0	0				0	0	0		1	0	0		0	0	0
34	September 19, 2016	113,38	-1,34 D	1,34	1					0	1				0	0	0		1	0	0		0	0	0
35	September 20, 2016	113,57	-0,01 D	0,01	2					0	0				0	0	0		1	0	0		0	0	0
36	September 21, 2016	113,55	-0,02 D	0,02	2					0	0				0	0	0		0	1	0		0	0	0
37	September 22, 2016	114,62	1,07 A	1,07	1					1	0				0	1	0		0	0	0		0	0	0
38	September 23, 2016	112,71	-1,91 D	1,91	1					0	1				0	0	0		0	1	0		0	0	0
39	September 26, 2016	112,88	0,17 A	0,17	2					0	0				1	0	0		0	0	0		0	0	0
40	September 27, 2016	113,09	0,21 A	0,21	2					0	0				1	0	0		0	0	0		0	0	0
41	September 28, 2016	113,95	0,86 A	0,86	2					0	0				0	1	0		0	0	0		0	0	0
42	September 29, 2016	112,13	-1,77 D	1,77	1					0	1				0	0	0		0	1	0		0	0	0
43	September 30, 2016	113,05	0,87 A				vidējā cenu izmaiņa	0,9639	TOTAL	22	13			TOTAL	29	25	0	TOTAL	21	25	0	TOTAL	0	0	0
44	September 30, 2016	583								A	D	N													
45										a1/01/n1	22	13	0												0
46										a2/02/n2	32	33	0												0
47										a3/03/n3	0	0	1												0
48															AA	AD	AN		DD	DA	DN		NN	NA	ND

```
install.packages("HMM")
library(HMM)
install.packages("data.table")
library(data.table)

getwd()
setwd("D:/Desktop/bak_dati")

#iespējamās vērtības stāvoklī
Si <- c("A","D")
Si1 <- as.matrix(Si)
Si <- Si1

#iespējamie simboli stāvoklī
vk <- c(1,2)
vk1 <- as.matrix(vk)
vk <- vk1

#sākuma varbūtības
pi <- c(1/2,1/2)
pi2 <- as.matrix(pi)
pi <- pi2

#matrica A
A <- matrix(c(21/48,27/52,27/48,25/52),
nrow=2,ncol=2)
```

```

#matrica B
B <- matrix(c(20/48,19/52,28/48,33/52),
nrow=2,ncol=2)

#inithmm
modelis <- initHMM(Si,vk,pi,A,B)

#BAUM-WELCH, uzlabojam parametrus
SV <- read.table("apple_simb.txt")
SV2 <- as.vector(SV$V1)

bw <- baumWelch(modelis,SV2,100)
print(bw)

#izmantojam uzlabotos parametrus par sākuma parametriem
A1 <- bw$hmm$transProbs
B1 <- bw$hmm$emissionProbs
pi1 <- bw$hmm$startProbs

#FORWARD - varbūtības, atrasties Si-tajā stāvoklī laikā k
#HMM ar kalibrētiem parametriem
modelis1 <- initHMM(Si,vk,pi1,A1,B1)
logforwardprobab <- forward(modelis1,SV2)
forwardprobab <- (exp(logforwardprobab))
forwardprobab

varb <- sum(forwardprobab [,100])

```

varb

```
#kad ir atrasta līdzīga varbūtība, tad darām tālāko
```

```
#sākuma varbūtības
```

```
pi <- c(1/2,1/2)
```

```
pi2 <- as.matrix(pi)
```

```
pi <- pi2
```

```
#matricaA
```

```
A <- matrix(c(27/54,26/46,27/54,20/46),
```

```
nrow=2,ncol=2)
```

```
#matrica B
```

```
B <- matrix(c(24/54,17/46,30/54,29/46),
```

```
nrow=2,ncol=2)
```

```
#inithmm
```

```
modelis <- initHMM(Si,vk,pi,A,B)
```

```
#BAUM-WELCH, uzlabojam parametrus
```

```
SV <- read.table("apple_simb2.txt")
```

```
SV2 <- as.vector(SV$V1)
```

```
bw <- baumWelch(modelis,SV2,125)
```

```
print(bw)
```

```
#izmantojam uzlabotos parametrus par sākuma parametriem
```

```
A1 <- bw$hmm$transProbs
```

```
B1 <- bw$hmm$emissionProbs
```

```
pi1 <- bw$hmm$startProbs
```

```
#FORWARD - varbūtības, atrasties Si-tajā stāvoklī laikā k
```

```
#HMM ar kalibrētiem parametriem
```

```
modelis1 <- initHMM(Si,vk,pi1,A1,B1)
```

```
logforwardprobab <- forward(modelis1,SV2)
```

```
forwardprobab <- (exp(logforwardprobab))
```

```
forwardprobab
```

```
varb1 <- sum(forwardprobab [,100])
```

```
varb1
```

```
starp <- abs(varb1-varb)
```

```
starp
```

```
starp<10e-10
```

```
#Prognoze
```

```
cena_rit <- 92.72+(94.20-98.72)*(varb1-varb)
```

```
cena_rit
```

Bakalaura darbs „Akciju tirgus cenu prognozēšana, izmantojot slēptos Markova modeļus”  
izstrādāts LU Fizikas un matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie  
informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Laura Jozefa

---

*(personiskais paraksts)*

Rekomendēju darbu aizstāvēšanai

Vadītāja: Dr. math. Nataļja Budkina

---

*(paraksts)*

05.06.2017.

Recenzents: Lektora p.i. Leonora Pahirko

Darbs iesniegts Matemātikas nodaļā \_\_.06.2017.

Dekāna pilnvarotā persona: vecākā metodiķe Dzintra Holsta

Darbs aizstāvēts Valsts pārbaudījuma komisijas sēdē

\_\_\_ 06.2017. prot. Nr. \_\_\_\_, vērtējums \_\_\_\_

Komisijas sekretāre: asociētā profesore Ingrīda Uljane