Facultatea de Cibernetică, Statistică și Informatică Economică din cadrul ASE

Program de masterat - Statistică Aplicată și Data Science

Procese și Modele Stochastice

- Proiect –

|  |  |
| --- | --- |
| Profesor,  Conf. dr. Covrig Mihaela | Masterand,  Diana-Mihaela Dorobanțu |

2022

Cuprins

[Aplicabilitatea lanțurilor Markov în prognoza meteo 3](#_Toc92388346)

[Spațiul stărilor 5](#_Toc92388347)

[Matricea de trecere 5](#_Toc92388348)

[Graful orientat 6](#_Toc92388349)

[Clase de comunicare 7](#_Toc92388350)

[Clasificarea stărilor lanțului 8](#_Toc92388351)

[Descompunerea canonică a spaţiului stărilor 9](#_Toc92388352)

[Perioada fiecărei stări 9](#_Toc92388353)

[Ergodicitate 9](#_Toc92388354)

[Distribuţa staţionară 10](#_Toc92388355)

[Distribuția limită 11](#_Toc92388356)

[Simularea unei traiectorii 12](#_Toc92388357)

[Durata medie de întoarcere 13](#_Toc92388358)

[Bibliografie 15](#_Toc92388359)

# Aplicabilitatea lanțurilor Markov în prognoza meteo

De la apariția computerelor puternice, modelele numerice au fost folosite pentru a înțelege mai bine dinamica vremii.

Cele două obiective principale care se doresc a fi atinse cu aceste modele sunt predicțiile meteorologice îmbunătățite și o mai bună cunoaștere a sensibilității sistemului climatic al Pământului la influențele antropice. În prezent, modelele globale împreună cu datele din observațiile satelitare și de la sol sunt folosite pentru a conduce modele locale. (Jordan & Talkner, 2000)

Una dintre metodele de a face o prognoză meteo se numește metoda analogică. Este vorba despre examinarea stării trecute (care este aceeași cu starea noastră actuală) și prognozează vremea pe baza a ceea ce s-a întâmplat în trecut. Prin urmare, se poate folosi modelul Markov pentru a aplica metodele analogice de prognoză a vremii. (Ryan, 2020)

În teoria probabilității, un model Markov este un model stochastic utilizat pentru a modela sistemele care se schimbă aleator. Se presupune că stările viitoare depind doar de starea actuală, nu de evenimentele care au avut loc înaintea acesteia.

În cadrul acestui proiect, vom face observații asupra vremii din Seoul, capitala Coreei de Sud. Astfel, avem date asupra tipului de vreme din perioada 31.12.2019 - 20.04.2020. (Vishal, 2020)

În linii generale, clima Coreei este caracterizată de alternanța sezonieră a influenței continentale: rece și uscat – iarna, și a influenței musonului: umed și cald - vara. (koaha, 2020)

În Seoul, temperatura medie a lunii ianuarie variază între -7°C și 1°C, iar temperatura medie a lunii iulie variază între 22°C și 29°C.

Cantitatea de precipitații este concentrată între lunile de vară, din iunie până în septembrie. Coasta de sud este locul unde se produc taifunurile târzii, cu vânturi puternice și precipitații abundente. Precipitațiile medii anuale sunt cuprinse între 1.370mm/m2 în Seoul și 1.470mm/m2 în Busan (cu vârfuri de aproximativ 1.500 pe insula Jeju). (koaha, 2020) (Wikipedia, 2020)

Iarna începe în decembrie și durează până la mijlocul lunii martie. În regiunea de nord, unde se află și Seoul, iernile sunt dure cu temperaturi aproape întotdeauna sub 0°C, cu taote acestea ninsorile sunt relativ rare și nu sunt abundente. Principalul motiv pentru iernile extreme din Coreea este vântul siberian. În partea de sud a Coreei de Sud, iernile sunt blânde (Busan cu o temperatură medie de 5°C). (Climates to travel, -)

*Chart, line chart

Description automatically generatedTemperaturi medii lunare și vremea. Precipitații și ninsori lunare în*

*Seoul 2015 – 2021* (hikersbay, 2021)

Institutul Coreean de Sisteme de Predicție Atmosferică ( Korea Institute of Atmospheric Prediction Systems - KIAPS) a început un proiect național de dezvoltare a unui nou sistem global de modele atmosferice în 2011. Scopul final al acestui proiect de 9 ani este de a înlocui modelul operațional actual de la Administrația Meteorologică din Coreea (Korea Meteorological Administration - KMA), care a fost adoptat de la Modelul Unificat (Unified Model - UM) al Oficiului Meteorologic al Marii Britanii din 2010. (Hong, et al., 2018)

Modelul Coreean Integrat (Korean Integrated Model - KIM) a fost implementat într-un cadru de prognoză în timp real cu un pachet de parametrizare a fizicii avansat. Noua generație de model global KIM a fost lansat operațional în aprilie 2020. (Korea Meteorological Administration, 2020)

# Spațiul stărilor

Pentru a putea preciza care este spațiul stărilor, am importat în R fișierul “weather\_seoul” în format csv care conține descreierea vremii (pe zile) în Seoul, Coreea de Sud, în perioada 31.12.2019 - 20.04.2020.

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generatedAstfel, folosind funcția unique, am determinat spațiul stărilor.

Se poate observa că avem 6 stări, lucru pe care l-am verificat folosind și funcția nrow căreia i-am plasat argumentul stari\_col.

Putem concluziona că spațiul stărilor este:

S = {clear-day, partly-cloudy-day, rain, cloudy, snow, wind}

# Matricea de trecere

Pentru această cerintă am utilizat software-ul EXCEL.

Table

Description automatically generated with medium confidence

Astfel, pe o coloană avem șirul de evenimente , iar pe alta șirul de evenimente (pentru o mai ușoară înțelegere a problemei).

Graphical user interface, application, table

Description automatically generated

Am notat stările cu numere de la 1 la 6 (deci avem starea 1, starea 2, ..., starea 6) pentru a putea completa coloana de tranziții într-un mod facil cu formula: \* 10 + .

**Graphical user interface, application, table, Excel

Description automatically generated**Matricea de trecere conține probabilitățile (probabilitățile de trecere din starea i în starea j într-un singur pas). Pentru a calcula acești termeni, am realizat o coloană de tranziții (luate pe linii din matricea de trecere), alături am creeat coloana frecvențelor absolute pe care am completat-o utilizând funcția COUNTIF, iar în final am determinat probabilitățile, adică frecvențele relative.

Astfel, am introdus matricea de trecere în R:

Text

Description automatically generated with medium confidence

# Graful orientat

Orice lanț Markov poate fi descris printr-un graf direcționat numit graful de tranziție al lanțului Markov, unde vârfurile sau nodurile sunt stările lanțului, iar între perechile de stări probabilitățile de trecere sunt marcate lângă arce (muchii).

Diagram

Description automatically generated

# Clase de comunicare

Pentru a putea scrie mai ușor clasele de comunicare, vom alege să lucrăm cu notațiile claselor cu cifrele de la 1 la 6.

Diagram, schematic

Description automatically generated

Avem astfel:

Text, letter

Description automatically generatedDiagram

Description automatically generated

# Clasificarea stărilor lanțului

Pentru un lanț Markov finit ireductibil, toate stările sunt recurente.

Cu toate acestea, am ales să demonstrez:

A picture containing table

Description automatically generated

De aici rezultă că starea “clear” este o stare recurentă.

Text, letter

Description automatically generatedCum stările unei clase de comunicare sunt fie toate recurente, fie toate tranziente și tocmai am arătat că starea “clear” este o stare recurentă toate stările lanțului Markov din clasa {clear-day, partly-cloudy-day, rain, cloudy, snow, wind} sunt recurente.

Daca sistemul trece intr-o stare din care nu mai poate trece in alta, atunci aceasta este denumita "stare absorbanta" => o stare 𝑖 este absorbantă dacă = 1.

Putem afirma că lanțul nostru Markov nu are nicio stare absorbantă => lanţul nu este absorbant. Singura probabilitate egală cu 1 este cea din starea (5) “snow” în starea “clear” (1), însă din starea “clear” se poate ajunge în celelalte stări. Doar în cazul în care aveam, de exemplu, = 1 am fi vorbit de o stare absorbantă.

# Descompunerea canonică a spaţiului stărilor

Spațiul finit al stărilor unui lanț Markov poate fi partiționat în stări tranziente și recurente: S = T R1 R2 ... Rm, unde:

T este mulțimea tuturor stărilor tranziente;

R1, R2, Rm sunt clase închise de comunicare ce conțin stări recurente.

S = T R1 R2 ... Rm se numește descompunerea canonică a lui S.

Având în vedere că, în cazul nostru, lanțul Markov este ireductibil, deci avem o singură clasă (recurentă), spațiul stărilor nu poate fi scris în altă formă.

# Perioada fiecărei stări

Pentru un lanț Markov cu matricea de trecere P, perioada stării i S, denumită d(i), este c.m.m.d.c. al tuturor numerelor care arată numărul posibil de pași de întoarcere în starea i:

d(i) = c.m.m.d.c. {n > 0: > 0}

Stările unei clase de comunicare au toate aceeași perioadă.

d(“clear”) = d(1) = c.m.m.d.c. {1, 2, 3, ...} = 1 => starea “clear” este aperiodică

1 – 1 => 1 pas

1 – 2 – 1 => 2 pași

1 – 2 – 2 – 1 => 3 pași

...

Putem afirma, că toate stările din spațiul stărilor sunt aperiodice.

Astfel, lanțul Markov este aperiodic întrucât este ireductibil și toate stările au perioada egală cu 1.

# Ergodicitate

Un lanț Markov se numește ergodic daca este ireductibil, apreiodic și toate stările au timp mediu finit de întoarcere.

Se mai poate afirma și că un lanț Markov este ergodic dacă și numai dacă matricea de trecere este regulară.

În cazul nostru, lanțul respectă propozițiile => lanțul este ergodic.

Text

Description automatically generated

# Distribuţa staţionară

A piece of paper with writing

Description automatically generated with medium confidence

A piece of paper with writing

Description automatically generated with medium confidence

Dacă lanțul Markov este ergodic, atunci există o distribuție staționară care este și unica distribuție limită a lanțului.

Am determinat distribuția staționară în R, utilizând funcția stationary:

Text

Description automatically generated



# Distribuția limită

Având în vedere că matricea de trecere P este regulară

=> distribuția limită = distribuția staționară



Text

Description automatically generated

Distribuția limtă nu depinde de starea ințială a lanțului.

Pe termen lung:

• probabilitatea ca vremea să se afle în starea “clear” (indiferent de starea de pornire) este de 0.3451 => 34.51%;

• probabilitatea ca vremea să se afle în starea “partly-cloudy” (indiferent de starea de pornire) este de 0.2920 => 29.20%;

• probabilitatea ca vremea să se afle în starea “rain” (indiferent de starea de pornire) este de 0.0619 => 6.19%;

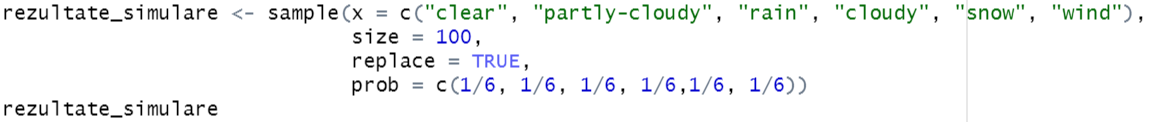
• probabilitatea ca vremea să se afle în starea “cloudy” (indiferent de starea de pornire) este de 0.2566 => 25.66%;

• probabilitatea ca vremea să se afle în starea “snow” (indiferent de starea de pornire) este de 0.0088 => 0.88%;

• probabilitatea ca vremea să se afle în starea “wind” (indiferent de starea de pornire) este de 0.0354 => 3.54%.

# Simularea unei traiectorii

Am propus o distribuţie iniţială (1/6, 1/6, 1/6, 1/6, 1/6, 1/6), iar apoi am simulat o traiectorie a lanţului pe un anumit număr de 100 paşi.



Am determinat distribuţia de frecvenţe absolute şi relative a întregii traiectorii.

Text

Description automatically generated

Vom compara distribuţia de frecvenţe relative a întregii traiectorii simulate cu distribuţia staţionară = distribuția limită.

Distribuția de frecvențe relative, în acest caz, este:



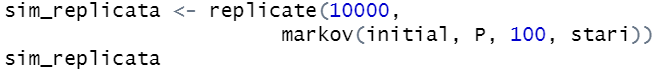
Amintim că distribuţia staţionară este:

clear partly-cloudy rain cloudy snow wind



Observam că frecvențele relative diferă foarte mult de rezultatul distribuției staționare. Astfel, putem crește numărul de replicări pentru a observa că, la un număr foarte mare al acestora, distribuția de frecvențe relative tinde către distribuţia staţionară.

Folosindu-ne de funcțiile markov și replicate, replicăm 10000 de traiectorii:



În acest caz, distribuția de frecvențe relative este:



Remarcăm faptul că rezultatele sunt foarte asemănătoare și, desigur, la un număr și mai mare de pași vor tinde să fie perfect egale.

|  |  |  |
| --- | --- | --- |
| Stări | Frecvențe relative | Distribuția staționară |
| clear | 0.34477228 | 0.345132743 |
| partly-cloudy | 0.28866634 | 0.292035398 |
| rain | 0.06374257 | 0.061946903 |
| cloudy | 0.25505941 | 0.256637168 |
| snow | 0.01037327 | 0.008849558 |
| wind | 0.03738614 | 0.035398230 |

# Durata medie de întoarcere

Pentru lanţul nostru Markov ireductibil, vom calcula durata medie de întoarcere în starea din care a plecat, adică numărul mediu de tranziţii sau paşi în care lanțul revizitează o anumită stare.

Pentru aceste calcule com utiliza formula:

, unde este durata medie de întoarcere în starea j, iar este unica distribuție staționară.

# Bibliografie

Climates to travel, -. *Climate - South Korea.* [Interactiv]   
Available at: https://www.climatestotravel.com/climate/south-korea  
[Accesat 05 January 2022].

hikersbay, 2021. *What is the weather like in Seoul? Seoul South Korea weather. Climate and weather in Seoul 2021.* [Interactiv]   
Available at: http://hikersbay.com/climate/southkorea/seoul?lang=en  
[Accesat 05 January 2022].

Hong, S.-Y.și alții, 2018. *The Korean Integrated Model (KIM) System for Global Weather Forecasting.* [Interactiv]   
Available at: https://www.researchgate.net/publication/326397032\_The\_Korean\_Integrated\_Model\_KIM\_System\_for\_Global\_Weather\_Forecasting  
[Accesat 05 January 2022].

Jordan, P. & Talkner, P., 2000. *A seasonal Markov chain model for the weather in the central Alps.* [Interactiv]   
Available at: https://www.tandfonline.com/doi/pdf/10.3402/tellusa.v52i4.12274  
[Accesat 05 January 2022].

koaha, 2020. *Coreea.* [Interactiv]   
Available at: https://koaha.org/wiki/Corea#Clima  
[Accesat 05 January 2022].

Korea Meteorological Administration, 2020. *Numerical Weather Prediction (NWP).* [Interactiv]   
Available at: https://web.kma.go.kr/eng/biz/forecast\_02.jsp  
[Accesat 05 January 2022].

Ryan, M., 2020. *A Very Simple Method of Weather Forecast Using Markov Model Lookup Table.* [Interactiv]   
Available at: https://towardsdatascience.com/a-very-simple-method-of-weather-forecast-using-markov-model-lookup-table-f9238e110938  
[Accesat 05 January 2022].

Vishal, V., 2020. *Historical Daily Weather Data 2020.* [Interactiv]   
Available at: https://www.kaggle.com/vishalvjoseph/weather-dataset-for-covid19-predictions  
[Accesat 29 December 2021].

Wikipedia, 2020. *Coreea de Sud.* [Interactiv]   
Available at: https://ro.wikipedia.org/wiki/Coreea\_de\_Sud#Clim%C4%83  
[Accesat 05 January 2022].