

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Hordós Donát

MARKOV LÁNCOK RENDJÉNEK BECSLÉSE

BSc Szakdolgozat

Témavezető:

Csiszár Villő

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2017

Tartalomjegyzék

Köszönetnyilvánítás	4
1. Bevezetés	5
2. Diszkrét paraméterű Markov láncok	7
2.1. Markov-tulajdonság	7
2.2. Átmenet-valószínűség	8
2.2.1. Átmenet mátrix	8
2.2.2. Példa	9
2.3. Markov lánc rendje	11
2.3.1. Példa	11
2.3.2. Példa	12
3. A rend becslése	15
3.1. Likelihood módszer	15
3.1.1. Maximum likelihood becslés	16
3.1.2. Példa	18
3.2. Információelméleti módszer	20
3.2.1. Akaike információs kritérium (AIC)	20
3.2.2. Bayes-féle információs kritérium (BIC)	21
3.2.3. Összehasonlítva a két kritérium	21
3.3. Cross-validation	22
3.3.1. A módszer	22
3.4. Bayes módszer	24
3.4.1. A priori eloszlás	24
3.4.2. A normáló tényező	25

3.4.3. A posteriori	26
4. Módszerek használata elemzéshez	30
4.1. Az adatok	31
4.2. Likelihood módszer	31
4.3. Információs kritériumok	32
4.4. Bayes módszer	33
5. Összegzés	34
Irodalomjegyzék	35

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Csiszár Villónak, akihez bármikor fordulhattam segítségért, illetve végtelen szakmai tudásával és elengedhetetlen jó tanácsaival segítette a szakdolgozatom elkészülését.

1. fejezet

Bevezetés

Tekintsünk egy olyan irányított gráfot, ami véges sok csúcsponttal rendelkezik, minden csúcsból megy ki él (ezek lehetnek hurokélek is), valamint az éleken nemnegatív számok szerepelnek. Ezek a számok a valószínűségek, amelyek összege 1. Ezen a gráfon bolyongunk a valószínűségek szerint. Akkor beszélünk diszkrét paraméterű Markov láncról, ha egységnyi időközönként lépünk tovább valamely élen.

A Markov lánc egy olyan diszkrét sztochasztikus folyamatot jelent, amely Markov-tulajdonságú. A Markov-tulajdonság annyit tesz, hogy adott jelen mellett a jövő feltételesen független a múlttól. Vagyis a jelen leírása teljesen magába foglalja az összes olyan információt, ami befolyásolhatja a folyamat jövőbeli helyzetét. Semmi, ami a múltban történt, nem hat, nem ad előrejelzést a jövőre nézve. A jövőben minden lehetséges.

Felmerülhet a kérdés, hogy mi van ha nem független eseményeket szeretnénk vizsgálni. Ekkor már szóba jöhet a Markov lánc rendje, ami azt jelenti, hogy ha a jövő függ az előző k eseménytől, akkor a lánc k -ad rendű lesz. Arról van szó, hogy a valós folyamatnak van egy tényleges rendje, amit mi nem tudunk, ezért szeretnénk megbecsülni. Ha a lánc ténylegesen elsőrendű, akkor a jövőre nézve csak a jelen ad információt. Ez egy egyszerű modell, azonban a valós folyamatok gyakran nem ilyen egyszerűek, hanem hosszabb a memóriájuk, azaz a jövő több utolsó állapottól függ. Ha egy ilyen hosszabb memóriájú folyamatot elsőrendű láncként modelleznénk, azal hibát követnénk el, nem jól íránk le a valóságot. A valóságban természetesen olyan folyamatok is vannak, amelyek nem Markov láncok semmilyen rendre, hanem végtelen a memóriájuk. Ezeket is hasznos lehet azonban valamilyen magasabb rendű lánccal közelíteni.

Igen széleskörű az alkalmazása. Használható a fizikában, statisztikai folyamatokban, szerencsejáték modellezésében, gazdaságban, matematikai biológiában stb. De például egy hétköznapi társasjáték is reprezentálható Markov láncsal.

A szakdolgozatban a Markov láncok rendjéről lesz szó, ahol először be lesznek mutatva a diszkrét Markov láncok, majd, hogy mit is jelent az hogy egy Markov lánc k -ad rendű. Ezek után olyan módszerek lesznek bemutatva, amelyek nagyon gyakoriak a statisztikában, mint például a likelihood módszer, vagy a Bayes módszer. Ezután a megismert módszerekre alkalmazunk egy életszerű példát Párizs 2016-os időjárása alapján. Az elemzések során Python és R programnyelvek lesznek használva.

A szakdolgozatban főként Philipp Singer és munkatársai cikkeire támaszkodtam [6, 7, 8], valamint az elemzés során is az ő kódjait használtam fel, amit a <https://github.com/psinger/PathTools> oldalon lehet megtalálni.

2. fejezet

Diszkrét paraméterű Markov láncok

2.1. Markov-tulajdonság

Ebben a fejezetben az [1, 2, 3, 5] irodalmat használtam föl.

Definiáljuk pontosan a Markov láncot.

2.1.1. Definíció. Legyen adott az $\{X_n\}_{n \in \mathbb{N}}$ folyamat, ahol $X_n : \Omega \rightarrow I$ valószínűségi változók az (Ω, \mathcal{F}, P) valószínűségi mezőn, I pedig megszámlálható halmaz. A folyamatot diszkrét paraméterű homogén Markov láncnak nevezzük, ha

– A folyamat Markov tulajdonságú, azaz

$$P(X_m \in B | X_{n_1} = i_1, \dots, X_{n_k} = i_k) = P(X_m \in B | X_{n_k} = i_k)$$

ha $n_1 < \dots < n_k \leq m$.

– A Markov folyamat stacionárius (vagy homogén) átmenetvalószínűségű, azaz minden $i, j \in I$ esetén minden olyan n -re, melyre $P(X_n = i) > 0$,

$$P(X_{n+1} = j | X_n = i) = p_{ij},$$

n -től függetlenül.

Fontos következménye a definíciónak, hogy

$$P(X_{n+m} = j_m, \dots, X_{n+2} = j_2, X_{n+1} = j_1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) =$$

$$= P(X_{n+m} = j_m, \dots, X_{n+2} = j_2, X_{n+1} = j_1 \mid X_n = i).$$

Azt jelenti ez az egyenlőség, hogy ha ismerjük a jelent, ami az n indexnél van, és annak az értékét ($X_n = i$), akkor a Markov láncban a jövő ($X_{n+m} = j_m, \dots, X_{n+2} = j_2, X_{n+1} = j_1$) független a múlt eseményeitől ($X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0$).

2.2. Átmenet-valószínűség

Minden egyes pillanatban a rendszer az adott valószínűségi változó eloszlása alapján vagy megváltoztatja az állapotát a jelenbeli állapotától, vagy ugyanúgy marad. Az állapotváltozásokat átmenetnek nevezzük, és azokat a valószínűségeket, melyek a különböző állapotváltozásokra vonatkoznak, átmenet-valószínűségeknek nevezzük. Az átmenet-valószínűségeket pedig úgy kaphatjuk meg, hogy az addigi valószínűségeket összeszorozzuk.

Vagyis:

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) &= \\ &= P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0) \cdots P(X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = \\ &= p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \end{aligned}$$

2.2.1. Átmenet mátrix

A Markov láncokat az egylépéses átmenetvalószínűségek, valamint a kezdeti eloszlás határozza meg. Minden $i, j \in I$ esetén $p_{ij} = P(X_{n+1} = j \mid X_n = i)$, ami az i -ből j -be való egylépéses átmenet-valószínűség. Ezeket az átmenet-valószínűségeket egy $|I| \times |I|$ méretű mátrixban tároljuk, ahol a mátrix i -edik sorában és j -edik oszlopában a következő valószínűség áll:

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$

Ha I végtelen, akkor azt mondjuk, hogy $P = (p_{ij})_{i,j \in I}$ egy végtelen mátrix.

Ennek az lesz a következménye, hogy a Markov láncot egy irányított gráfként tudjuk reprezentálni. A gráfon való bolyongás úgy néz ki, hogy ha n -edik lépésben eljutottunk az i -edik állapotba, akkor annak a valószínűsége, hogy j -be megyünk tovább: p_{ij} . A P mátrix sztochasztikus mátrix ha $p_{ij} \geq 0$, valamint minden sor összege 1, azaz

$$\sum_{j \in I} p_{ij} = 1$$

minden i -re.

2.2.2. Példa

Tegyük fel, hogy a 12 éves Móricka hazaérve az iskolából bekapcsolja a rádióját, ahol egy olyan zenét hall, amelyben az énekes csak 4 különböző szót használ. Móricka egyből tollat és papírt vesz maga elé és próbálja lejegyzetelni a szavakat. Minden szót elnevezett egy-egy számként, hogy könnyítse a dolgát. Sajnos a zene végére ért oda ezért nem sok számot tudott leírni. A következő számsorozatot írta le maga elé:

132431231322314144123

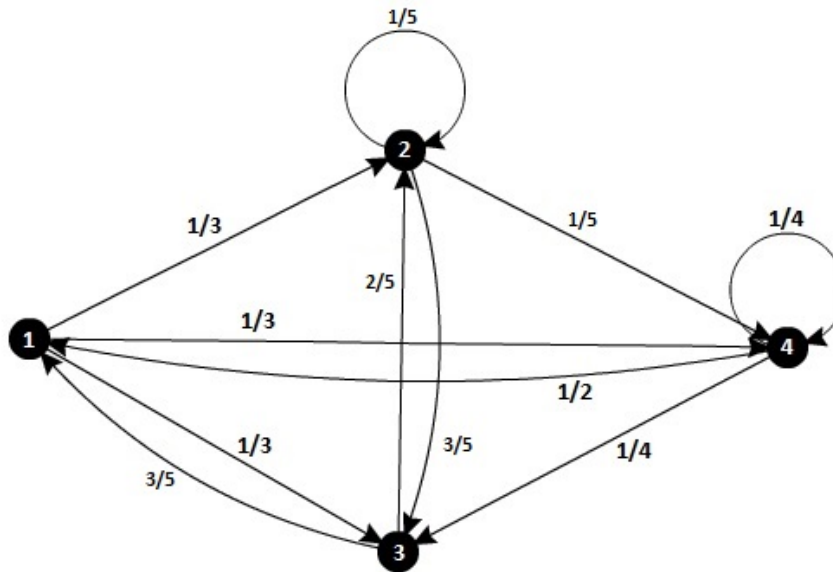
A megadott számsorozat alapján a következőképpen tudunk becsülni az egyes szavak átmenet-valószínűségeire: Először megszámloljuk, hogy az 1-es, 2-es, 3-as és 4-es szám után milyen gyakorisággal, mely számok következnek. Vagyis megnézzük az összes lehetséges állapot relatív gyakoriságát.

Nézzük meg például az 1-es és a 2-es számokat. Az 1-es számból a minta alapján összesen 6-szor indultunk ki, amelyből 2-szer a 2-es számba, 2-szer a 3-as számba, valamint 2-szer a 4-es számba érkeztünk. Ha a 2-es számot nézzük, akkor azt láthatjuk, összesen 5 alkalommal léptünk tovább a 2-as állapotból, amely az alábbiakban oszlik el: 3-szor a 3-asba tértünk, 1-szer a 4-esbe és 1-szer visszatért önmagába, a 2-esbe. Ezt folytatjuk minden további számmal és ezáltal fel tudjuk írni az összes állapotváltozásra a becsült valószínűségeket. Tehát annak a becsült valószínűsége, hogy az 1-esből a 2-esbe megyünk $2/6$, 3-asba szintén $2/6$ stb.

Ezeket a becsült valószínűségeket fel tudjuk írni egy átmenet-valószínűség mátrixba, ahol az i -edik sor j -edik eleme azt jelenti, hogy az i állapotból milyen becsült valószínűséggel juthatunk el j csúcsba.

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/5 & 3/5 & 1/5 \\ 3/5 & 2/5 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/4 \end{pmatrix}$$

A könnyebb megértéshez, az átláthatóságához fel tudunk rajzolni egy irányított gráfot is, amely az alábbi módon ábrázolható:



2.1. ábra. A feladat irányított gráfja

Nézzünk még egy példát ehhez a feladathoz.

Becsüljük meg, hogy mi annak a valószínűsége, hogy a következő sorrendben elhangzottak az alábbi szavak a zeneszámban, feltéve, hogy 1-es szó elhangzott: 12322314?

Megoldás:

A megoldást úgy becsülhetjük meg, hogy ha az átmenet-valószínűségeket összeszorozzuk, azaz

$$\begin{aligned}
 &P(X_1 = 2, X_2 = 3, X_3 = 2, \dots, X_6 = 1, X_7 = 4 | X_0 = 1) = \\
 &= P_{(12322314)} = 1/3 \cdot 3/5 \cdot 2/5 \cdot 1/5 \cdot 3/5 \cdot 3/5 \cdot 1/3 = 0.00192
 \end{aligned}$$

2.3. Markov lánc rendje

Most kicsit általánosítjuk a Markov lánc fogalmát, valamint megmutatjuk, hogy egy magasabb rendű Markov lánc visszavezethető első rendű Markov láncre.

2.3.1. Definíció. Azt mondjuk, hogy az $\{X_n\}_{n \in \mathbb{N}}$ folyamat I állapotterű m -edrendű Markov lánc ($m \geq 1$), ha minden $n = 0, 1, \dots$ és $i_k \in I, k = 0, 1, \dots$ mellett teljesül

$$\begin{aligned} P(X_{n+m} = i_{n+m} \mid X_{n+m-1} = i_{n+m-1}, \dots, X_0 = i_0) &= \\ &= P(X_{n+m} = i_{n+m} \mid X_{n+m-1} = i_{n+m-1}, \dots, X_n = i_n). \end{aligned}$$

Meg lehet mutatni, hogy egy magasabb m -ed rendű ($m > 1$) Markov lánc visszavezethető egy m dimenziós $Y = \{Y_0, Y_1, \dots\}$ folyamatra, amely elsőrendű Markov lánc.

Legyen $I' = \{(k_1, \dots, k_m) : (k_1, \dots, k_m) \in I\}$ állapotterű

$$Y_n = (X_{n+m-1}, \dots, X_n), \quad n = 0, 1, \dots$$

sztochasztikus folyamat. Ekkor

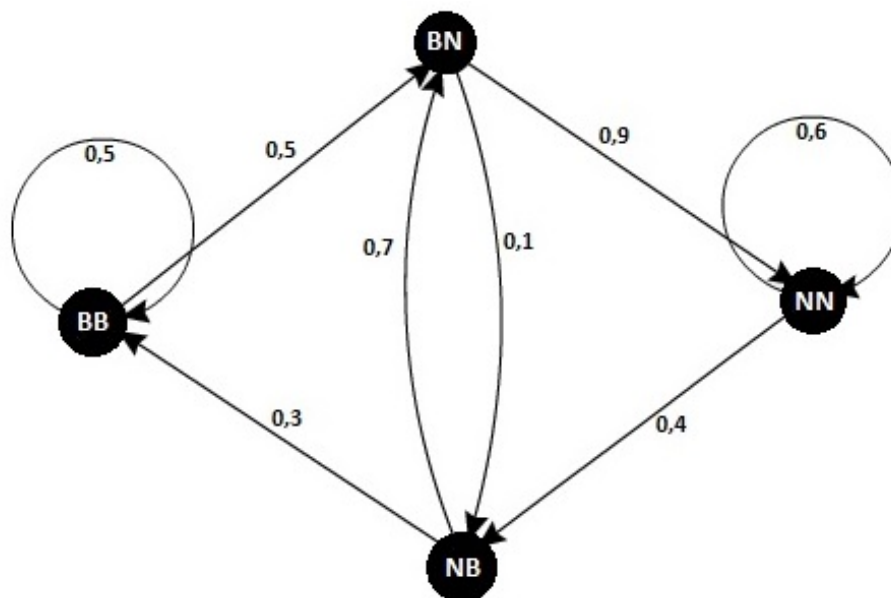
$$\begin{aligned} P(Y_{n+1} = (i_{n+m}, \dots, i_{n+1}) \mid Y_n = (i_{n+m-1}, \dots, i_n), \dots, Y_0 = (i_{m-1}, \dots, i_0)) &= \\ &= P(X_{n+m} = i_{n+m}, \dots, X_{n+1} = i_{n+1} \mid X_{n+m-1} = i_{n+m-1}, \dots, X_0 = i_0) = \\ &= P(X_{n+m} = i_{n+m}, \dots, X_{n+1} = i_{n+1} \mid X_{n+m-1} = i_{n+m-1}, \dots, X_n = i_n) = \\ &= P(Y_{n+1} = (i_{n+m}, \dots, i_{n+1}) \mid Y_n = (i_{n+m-1}, \dots, i_n)). \end{aligned}$$

Így beláttuk, hogy az újonnan bevezetett $Y = \{Y_0, Y_1, \dots\}$ folyamat elsőrendű Markov lánc.

2.3.1. Példa

Tegyük fel, hogy hosszú időn keresztül folyamatosan feljegyeztük az időjárást. Megfigyeltük, hogy az alábbi szabályok szerint alakul az időjárás: Ha adott nap borús idő volt, valamint előző nap is borús idő volt, akkor 50% eséllyel másnap is borús nap lesz. Ha egy adott nap napos idő volt, valamint előző nap is napos volt, akkor 60% az esély, hogy a következő nap is napos lesz. Ha aznap napos volt de előtte borús, akkor 10%, hogy megint borús lesz. Valamint, ha aznap borús volt, de előtte napos, akkor 70%, hogy megint napos lesz.

Ha a napok külön-külön lennének leírva akkor lenne a lánc másodrendű. Viszont ha minden i -edik és $(i + 1)$ -edik tagot 1 tagként nézünk, majd erre csináljuk meg a modellt, akkor elsőrendű modellt kapunk. Itt a párokban a k -edik pár második eleme megegyezik a $(k+1)$ -edik tag első elemével. A feladat elsőrendű modelljének a gráfja:



2.2. ábra. Az elsőrendű modell gráfja

2.3.2. Példa

Tegyük fel, hogy vannak adataink arról, hogy milyen időjárás volt egy adott helyen, adott intervallum alatt. Nézzük azt, hogy milyen láncot fogunk kapni, hogy ha a holnapi időjárásra vagyunk kíváncsiak annak függvényében, hogy ma milyen nap volt.

Tekintsük az $\{X_n, n = 1, 2, \dots\}$ sztochasztikus folyamatot, ahol

$$X_n = \begin{cases} 0 & \text{ha az } n\text{-edik napon esik} \\ 1 & \text{ha az } n\text{-edik napon nem esik} \end{cases}$$

Tegyük fel, hogy a holnapi nap csak a mától függ, ekkor a láncunk elsőrendű lesz. Ezért:

$$P(X_{n+1}|X_n, X_{n-1}, \dots, X_1) = P(X_{n+1}|X_n)$$

Ekkor az átmenet-valószínűség mátrix a következő lesz:

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} \\ P_{1,0} & P_{1,1} \end{pmatrix},$$

ahol $P_{0,0} = P(\text{holnap esik} \mid \text{ma esik}) = \alpha$. Így $P_{0,1} = 1 - \alpha$.

$P_{1,0} = P(\text{holnap esik} \mid \text{ma nem esik}) = \beta$, valamint $P_{1,1} = 1 - \beta$.

Most tegyük fel, hogy másodrendű a láncunk, vagyis nem csak a mától függ holnapi időjárás hanem a tegnaptól is. Nézzük meg ebben az esetben hogyan alakul át a feladat.

Mivel már két csúcstól is függ, ezért:

$$P(X_{n+1} \mid X_n, X_{n-1}, \dots, X_1) = P(X_{n+1} \mid X_n, X_{n-1})$$

Viszont ez nem egy elsőrendű, ezért jelölje Y_n a következőt:

$$Y_n = \begin{cases} 0 & \text{ha } X_n = 0, X_{n-1} = 0, \text{ EE} \\ 1 & \text{ha } X_n = 0, X_{n-1} = 1, \text{ NE} \\ 2 & \text{ha } X_n = 1, X_{n-1} = 0, \text{ EN} \\ 3 & \text{ha } X_n = 1, X_{n-1} = 1, \text{ NN} \end{cases}$$

ahol

$$\begin{aligned} P(Y_{n+1} \mid Y_n, Y_{n-1}, \dots) &= P(X_{n+1}, X_n \mid X_n, X_{n-1}, \dots) = P(X_{n+1}, X_n \mid X_n, X_{n-1}) = \\ &= P(Y_{n+1} \mid Y_n). \end{aligned}$$

Így már egy elsőrendű modellt kaptunk a másodrendűből. Tehát $\{Y_n, n = 1, 2, 3, \dots\}$ egy elsőrendű Markov lánc, aminek az átmenet-valószínűség mátrixa a következőképpen fog kinézni:

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} & P_{0,2} & P_{0,3} \\ P_{1,0} & P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,0} & P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,0} & P_{3,1} & P_{3,2} & P_{3,3} \end{pmatrix}$$

ahol $P_{0,1} = P(Y_{n+1} = 1|Y_n = 0) = P(X_{n+1} = 0, X_n = 1|X_n = 0, X_{n-1} = 0) = 0$.

$P_{0,3} = P(Y_{n+1} = 3|Y_n = 0) = P(X_{n+1} = 1, X_n = 1|X_n = 0, X_{n-1} = 0) = 0$.

Hasonlóképpen, $P_{1,1} = P_{1,3} = 0, P_{2,0} = P_{2,2} = 0, P_{3,0} = P_{3,2} = 0$.

Ekkor a mátrix a következőre módosul:

$$P = \begin{pmatrix} P_{0,0} & 0 & P_{0,2} & 0 \\ P_{1,0} & 0 & P_{1,2} & 0 \\ 0 & P_{2,1} & 0 & P_{2,3} \\ 0 & P_{3,1} & 0 & P_{3,3} \end{pmatrix}$$

ahol a valószínűségek a következőket mutatják:

$P_{0,0} = P(\text{holnap fog esni} \mid \text{ma esik, tegnap esett})$.

$P_{1,0} = P(\text{holnap fog esni} \mid \text{ma esik, tegnap nem esett})$.

$P_{2,1} = P(\text{holnap fog esni} \mid \text{ma nem esik, tegnap esett})$.

$P_{3,1} = P(\text{holnap fog esni} \mid \text{ma nem esik, tegnap nem esett})$.

3. fejezet

A rend becslése

Ebben a fejezetben a [2, 3, 9, 10] irodalmat használtam föl.

A Markov láncok rendjének meghatározása előnyös lehet, ahogy a bevezetésben is elhangzott, hogy a valós folyamatnak van egy tényleges rendje, amit mi nem tudunk, ezért szeretnénk megbecsülni. Ha a lánc ténylegesen elsőrendű, akkor a jövőre nézve csak a jelen ad információt. Ez egy egyszerű modell, azonban a valós folyamatok gyakran nem ilyen egyszerűek, hanem hosszabb a memóriájuk, azaz a jövő több utolsó állapottól függ. A Markov láncok rendjét többféleképpen is meg lehet becsülni, de közülük csak néhány kerül megemlítésre a dolgozatban, mint például a likelihood hanyadospróba, Bayes-becsléses módszer, információ elméleti módszer, valamint a cross-validation módszer.

3.1. Likelihood módszer

Talán az egyik legelterjedtebb statisztikai becslés a *maximum likelihood* becslés. Központi eleme a statisztikának. A *likelihood* kifejezést az alkotó, R. A. Fisher népszerűsítette az 1920-as években. A maximum likelihood módszer célja, hogy adott mérési értékekhez, az ismeretlen paramétereknek olyan becslését adja meg, amely mellett az adott érték a legnagyobb valószínűséggel következik be. Az eljárás a likelihood függvény maximalizálásával történik.

3.1.1. Maximum likelihood becslés

Legyen egy elsőrendű Markov láncunk, ahol jelölje $D = (X_1, \dots, X_n)$ a megfigyelésünket. Legyen a $\vartheta = P$ az ismeretlen paraméter, vagyis az átmenetmátrix. Valamint legyen adott a $p(x_1) = 1$ kezdeti eloszlás az egyszerűség végett. Az I pedig legyen a véges állapottér. Ekkor a *likelihood* függvény a következő lesz:

$$\begin{aligned} L(\vartheta) &= P(D|\vartheta) = p(x_n|x_{n-1})p(x_{n-1}|x_{n-2}) \dots p(x_2|x_1)p(x_1) = \\ &= p(x_1) \prod_i \prod_j p_{ij}^{n_{ij}} \end{aligned}$$

ahol az n_{ij} az átmenetek száma i csúcsból j csúcsba a D sorozatban.

Ezután azokra a ϑ értékekre vagyunk kíváncsiak, amelyekkel maximalizálni tudjuk a *likelihood* függvényt. Vagyis

$$\hat{\vartheta}_{MLE} = \arg \max_{\vartheta} P(D|\vartheta).$$

Ha a függvényt lehet deriválni ϑ szerint, akkor az első és a második deriváltak segítségével megkereshetjük a függvény maximumhelyét. Viszont ha a feladatunkban n -szeres szorzatot kell deriválnunk, amelyeknek minden tagjában ott van az a változó ami szerint deriválni szeretnénk, akkor megkönnyíthetjük a helyzetünket, ha vesszük a *likelihood* függvény logaritmusát, majd ennek a függvénynek keressük a maximumhelyét. Ezt megtehetjük, mivel a logaritmus függvény monoton növekvő függvény.

Ezáltal azt kapjuk, hogy:

$$\begin{aligned} \mathcal{L}(P(D|\vartheta)) &= \log \left(p(x_1) \prod_i \prod_j p_{ij}^{n_{ij}} \right) \\ &= \log p(x_1) + \sum_i \sum_j n_{ij} \log p_{ij} \end{aligned}$$

Tudjuk, hogy minden i -re:

$$\sum_{j=1}^m p_{ij} = 1,$$

ahol $I = \{1, \dots, m\}$. Ebből az következik, hogy:

$$\sum_j n_{ij} \log p_{ij} = \sum_{j=1}^{m-1} n_{ij} \log p_{ij} + n_{im} \log p_{im} =$$

$$= \sum_{j=1}^{m-1} n_{ij} \log p_{ij} + n_{im} \log \left(1 - \sum_{j=1}^{m-1} p_{ij} \right)$$

Deriváljuk le p_{ij} szerint. Így

$$\frac{\partial}{\partial p_{ij}} \left(\sum_{j=1}^{m-1} n_{ij} \log p_{ij} + n_{im} \log p_{im} \right) = \frac{n_{ij}}{p_{ij}} + \frac{n_{im}}{p_{im}} \cdot (-1)$$

Majd a kapott kifejezést tegyük egyenlővé 0-val:

$$\frac{n_{ij}}{p_{ij}} = \frac{n_{im}}{p_{im}}, \quad j = 1, \dots, m-1$$

Amiből következik, hogy a $\frac{p_{ij}}{n_{ij}}$ hányados egy konstans, ami nem függ j -től.

$$p_{ij} = c \cdot n_{ij}$$

Mivel minden sor összege 1, ezért próbáljuk a c -t is úgy választani, hogy az 1 legyen, így:

$$1 = \sum_{k=1}^m p_{ik} = \sum_{k=1}^m c \cdot n_{ik} = c \sum_{k=1}^m n_{ik},$$

azaz

$$c = \frac{1}{\sum_{k=1}^m n_{ik}}.$$

Így azt kapjuk, hogy a p_{ij} maximum likelihood becslése a következő:

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_k n_{ik}}.$$

Tehát megkaptuk p_{ij} -re a *maximum likelihood* becslést, ami a relatív gyakorisággal egyenlő.

Viszont nekünk nem elegendő meghatározni az elsőrendű láncban az átmenet-mátrix *maximum likelihood* becslését. Minél jobban meg szeretnénk becsülni a lánc megfelelő rendjét. Az alacsonyabb rendű modellek be vannak építve a magasabb rendű modellekbe, valamint a magasabb rendű Markov lánc jobb becslést tud

adni számunkra. Ugyancsak oda kell figyelniük a túltanulás esetére is.

A továbblépéshez szükséges bevezetni a *likelihood* hányados tesztet, amellyel két különböző modell jóságát lehet összehasonlítani. Az egyik a *null – modell* (ami esetünkben az a rendű lesz), a másik pedig az *alternatív – modell* (b), ahol $b > a$. Ezzel a próbával azt tudjuk eldönteni, hogy az adatunkra szignifikánsan jobban illeszkedik-e az alternatív modell, mint a null modell. Jelölje a próbastatisztikát ${}_a\eta_b$. Így:

$${}_a\eta_b = 2 \left(\mathcal{L}(P(\mathcal{D} | \vartheta_b)) - \mathcal{L}(P(\mathcal{D} | \vartheta_a)) \right)$$

A modell jósága abból is eredhet, hogy ha megnő a paraméterek száma. Így nem fogunk jó eredményeket kapni, ezért alkalmaznunk kell a modellünkre egy szignifikancia tesztet, ezzel elkerülvén a túlillesztés lehetőségét. Ha a null modellünk be van ágyazva az alternatív modellbe, akkor általános eredmények szerint a próbastatisztika eloszlása a χ^2 eloszláshoz tart, melynek a szabadságfoka a két modell paraméterszámának különbsége, mely esetünkben

$$(|I|^b - |I|^a)(|I| - 1)$$

lesz. Amennyiben a próbastatisztikával kiszámított érték nem esik bele az elfogadási tartományba, akkor a nullhipotézisünket elutasítjuk és az alternatív hipotézist fogadjuk el.

3.1.2. Példa

Nézzünk egy egyszerű példát a likelihood módszerhez. Legyen példa a következő két állapotú sorozatunk, valamint tegyük fel, hogy $P(X_1) = 1$:

0101011100101010

Vizsgáljuk meg, hogy vajon első-, vagy másodrendű-e ez a sorozat. Kezdjük az elsőrendű modellel, vagyis nézzük meg az átmenetvalószínűségek maximum likelihood becslését:

$$\begin{aligned} P(X_{n+1}|X_n) &= P(1|0) = \frac{6}{7} \\ P(0|0) &= \frac{1}{7} \\ P(0|1) &= \frac{6}{8} \end{aligned}$$

$$P(1|1) = \frac{2}{8}$$

Ekkor a loglikelihood értéke:

$$6 \cdot \ln(6/7) + 1 \cdot \ln(1/7) + 6 \cdot \ln(6/8) + 2 \cdot \ln(2/8) + 1 \cdot \ln(1) = \\ -7.37.$$

Most nézzük azt a modellt, amikor a sorozatunk másodrendű és ennek vizsgáljuk a maximum likelihood becslését. A becsült valószínűségek a következők lesznek:

$$\begin{aligned} P(1|00) &= 1 & P(0|00) &= 0 \\ P(1|01) &= \frac{1}{6} & P(0|01) &= \frac{5}{6} \\ P(1|10) &= \frac{4}{6} & P(0|10) &= \frac{2}{6} \\ P(0|11) &= \frac{1}{2} & P(1|11) &= \frac{1}{2} \end{aligned}$$

Így azt kapjuk, hogy:

$$\begin{aligned} &= 1 \cdot \ln(1) + 0 \cdot \ln(0) + 1 \cdot \ln(1/6) + 5 \cdot \ln(5/6) + \\ &+ 4 \cdot \ln(4/6) + 2 \cdot \ln(2/6) + 1 \cdot \ln(1/2) + 1 \cdot \ln(1/2) + 1 \cdot \ln(1) = -7.91. \end{aligned}$$

Ebből a két számolásból azt kaptuk, hogy a másodrendű modellünk jobb mint az első rendű, mivel a kapott két érték közül, a másodiknak nagyobb az értéke. Viszont ahhoz, hogy bebizonyosodjon, hogy valóban jobb a másodrendű modell, egy szignifikancia tesztet kell alkalmaznunk, hogy eldöntsük, hogy szignifikánsan jobb-e mint az elsőrendű modellünk, és nem csak azért jobb mert megnöttek a paraméterek száma.

Korábban már láttuk, hogy mi is a szignifikancia teszt próbastatisztikája. Így jelölje ezt a ${}_a\eta_b$, aminek az eloszlása a χ^2 eloszláshoz fog tartani, ahol a szabadságfok a két modell paraméterszámának különbsége.

Tehát:

$$\begin{aligned} {}_a\eta_b &= 2 \left(\mathcal{L}(P(\mathcal{D} | \vartheta_b)) - \mathcal{L}(P(\mathcal{D} | \vartheta_a)) \right) = \\ &= 2(-7.91 + 7.37) = -1.08. \end{aligned}$$

Az állapotterünk 2 elemű lesz, mivel két csúcsot használtunk (0-t és 1-et), ezért a szabadságfokot úgy kapjuk meg hogy:

$$(|I|^b - |I|^a)(|I| - 1) = (2^2 - 2^1)(2 - 1) = 2.$$

Szignifikánsan nem tér el egymástól a két modell, vagyis nem illeszkedik jobban a másodrendű modell, az elsőrendű modellhez képest.

3.2. Információelméleti módszer

Az információelméleti módszer az információelméletből származtatott entrópián alapszik. A következőkben említés szintjén lennének bemutatva a két jól ismert módszert. Az egyik az Akaike információs kritérium (*AIC*), a másik pedig a Bayes-féle információs kritérium (*BIC*).

3.2.1. Akaike információs kritérium (AIC)

A módszer több modellről dönti el, hogy melyik a legjobb modell. Alapja a maximum likelihood hányados teszt. A közelítés alapja, hogy minimalizálni kell az *AIC*-ot (minimum AIC becslés - MAICB) a különböző modellek szerint. Az Akaike információs kritérium a következő lesz:

$$AIC(a) = a\eta_b - 2(|I|^b - |I|^a)(|I| - 1)$$

A teszt a Kullback-Leibler távolságon¹ és az aszimptotikus tulajdonságú likelihood hányados teszten alapszik. Az alapötlet az, hogy válasszuk *b*-t meglehetősen nagyra és teszteljük alacsonyabb rendű modelleken, amíg meg nem találjuk az optimális rendet. Vagyis egy modell jóságát az összes többi modellhez viszonyítva becsli meg. Ha túl sok paraméter kerül a modellbe, akkor nem biztos, hogy az optimális modellt tudjuk kiválasztani. Az Akaike-féle információs kritérium a két ellentétes hatást próbálja kiegyensúlyozni. Egy jól illeszkedő, nem túlillesztett modell Akaike-féle információs kritériuma kicsi. A *MAICB* (minimum AIC becslés) automatikusan kiválaszt egy modellt, ami általában jó, vagy közel áll az optimális megoldáshoz. Ha két egymásba ágyazott modellnél az *AIC* 2-nél kevesebbel tér el, akkor a két modell nem tér el egymástól szignifikánsan. Valamint attól, hogy az $a\eta_b$ és/vagy az *AIC* kicsi, még gyakran előfordul, hogy a modell rosszul illeszkedik.

¹Az információelméletben a Kullback-Leibler távolság két valószínűségi eloszlás különbözőségét méri. Az egyik tipikusan az elméleti eloszlást, míg a másik a tapasztalatit reprezentálja.

3.2.2. Bayes-féle információs kritérium (BIC)

Más néven az optimális várható érték kritériuma. A kritériumra úgy lehet tekinteni, mint a Bayes faktor közelítésére. Nagyon hasonló az előbb bemutatott *AIC*-hoz, viszont annyi különbség van benne, hogy a büntetőtag még meg van szorozva a megfigyelések számának logaritmusával. Vagyis:

$$BIC(a) = {}_a\eta_b - (|I|^b - |I|^a)(|I| - 1) \ln(n)$$

Ismét nagyra választjuk b -t, majd alacsonyabb rendű modellekkel vizsgáljuk meg. A bayesi kritérium aszimptotikusan megegyezik a maximum *a posteriori* szabállyal, vagyis azt a hipotézist választja, amelyik az adott minta mellett a legvalószínűbb. A *BIC* elég nagy n esetén maximalizálja a jó választás valószínűségét. A kritérium bizonyíthatóan konzisztens becslés.

3.2.3. Összehasonlítva a két kritérium

Gyakran mindkét kritérium ugyanazt a modellt választja. Viszont van néhány olyan eset amikor eltérnek egymástól.

Performancia szinten mind a kettőnek megvannak a maga erősségei és hátrányai. Az *AIC*-ot használva megvan rá az esély, hogy túlbecsüli a valódi rendet, függetlenül attól, hogy milyen nagy az adatunk, ez azt jelenti, hogy nem konzisztens a becslés, míg a *BIC* konzisztens becslés. Ennek ellenére kis mennyiségű adatok esetén sokkal jobban teljesít az *AIC*, mint a *BIC*.

Mindkét kritérium nagyon hasonlít a már említett likelihood hányados teszthez, viszont vannak alapvető különbségeik. Elsősorban az, hogy az információs kritériumokat lehet használni nem egymásba ágyazott modelleken is. Továbbá nem használ statisztikai próbát. Nagy előnye a *BIC*-nek, hogy könnyebben alkalmazható, ha nehéz az a priori eloszlást beállítani. Valamint akkor is érdemesebb ezt a kritériumot használni, amikor a megfigyelt adatok száma nagy.

Végezetül még azt lehet elmondani a két kritériumról, hogy az *AIC*-ot előrejelzésre, míg a *BIC*-et magyarázatra érdemes használni.

3.3. Cross-validation

Egy újabb módszer arra, hogyan is lehet meghatározni a Markov lánc rendjét.

A cross-validation vagy más néven keresztvalidáció egy prediktív módszer. Mivel olyan jelenségekre vagyunk kíváncsiak, amik még nem zajlottak le, ezért jön az az ötlet, hogy a már meglévő adathalmazunknak egy részét tekintsük úgy, mintha jövőbeli megfigyelések lennének, majd illesszük erre a halmazra azokat a modelleket amiket a maradék megfigyeléseinkből építettünk fel. Így tesztelve a modellek jóságát, teljesítményét.

Majd a modelleket összehasonlítja úgy, hogy valamilyen módszer szerint felosztja az adathalmazunkat egy tanító halmazra (training set), valamint egy validációs halmazra (validation set). A tanuló adathalmazon felépíti a modellt, valamilyen módszer szerint, aztán a validációs halmazon pedig előrejelzéseket ad, méri a modell illeszkedését.

Egy keresztvalidációs módszer a K -fold Cross-Validation, amely annyit tesz, hogy felosztja az adathalmazunkat k darab, közel egyenlő méretű diszjunkt részhalmazra. Majd $k - 1$ részt használ a modell felépítésére és a maradék 1 halmazt pedig validálásra használja. Ezt véghez viszi k -szor, majd a létrejövő k darab modell átlagát tekinti a modell minőségének.

3.3.1. A módszer

Ebben a módszerben az átmenetvalószínűségekre fókuszálunk. A tanító/tanuló halmazból megbecsüljük az átmenetmátrixot maximum likelihood módszerrel, aztán meghatározzuk minden i -edik csúcsból j -edik csúcsba menő él rangját. Ezt úgy csináljuk, hogy az i -edik csúcsból kimenő éleket az átmenetvalószínűségük alapján sorba rendezzük, majd a legnagyobb valószínűség megkapja az 1-es rangot, míg a második legnagyobb valószínűség a 2-es rangot, és így tovább. Majd vesszük ezek átlagát a validáló halmazon.

Formálisan határozzuk meg az átlag rangot minden D_f validáló halmazra:

$$\overline{r(D_f)} = \frac{\sum_i \sum_j n_{ij} r_{ij}}{\sum_i \sum_j n_{ij}}$$

ahol n_{ij} az átmenetek száma i csúcsból j csúcsba a D_f sorozatban, r_{ij} pedig a be-

csült átmenetvalószínűség mátrix i -edik sorában lévő j -edik elem rangja.

Ha egy sorban megegyezik két, vagy akár több valószínűség, akkor azok rangja megegyezik a közülük legkisebb valószínűséggel rendelkező ranggal.

Nézzük meg a második fejezetben lévő **2.2.2-es példát**, ahol az alábbi számsorozatokból próbáltunk a valószínűségekre becsülni:

132431231322314144123

Felhasználva az alábbi becsült átmenet-valószínűség mátrixot:

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/5 & 3/5 & 1/5 \\ 3/5 & 2/5 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/4 \end{pmatrix},$$

megkapjuk a rangok mátrixát:

$$R = \begin{pmatrix} 4 & 3 & 3 & 3 \\ 4 & 3 & 1 & 3 \\ 1 & 2 & 4 & 4 \\ 1 & 4 & 3 & 3 \end{pmatrix}$$

Legyen ez a számsorozat a tanító számsorozatunk, valamint legyen a lánc elsőrendű. Ekkor a **2.2.2-es példában** is vizsgált 12322314 számsorozat legyen a validáló számsorozat, amelyre alkalmazzuk a megadott képletet.

$$\overline{r(D_f)} = \frac{\sum_i \sum_j n_{ij} r_{ij}}{\sum_i \sum_j n_{ij}} = \frac{3 + 1 + 2 + 3 + 1 + 1 + 3}{7} = \frac{14}{7} = 2$$

A módszer ezt elvégzi minden megadott k -ad rendű tanító halmazra, majd az $\overline{r(D_f)}$ rangokat kiátlagolja. Ha több modell közül kell választani, akkor azt a modellt választjuk, amelyeknek legkisebb a rangja.

A cross-validation módszer egyik hátránya, hogy nem tudjuk biztosra, hogyan is jó felosztani az adatunkat.

3.4. Bayes módszer

A bayesi modellezés egy külön ágazata a statisztikának. A megfigyelésekből próbál az adott jelenséget kiváltó okra következtetni, valamint valószínűsíteni ennek helyességét. A Bayes statisztika abban tér el a klasszikus statisztikától, hogy figyelembe veszi a becslés előtt esetlegesen már meglévő információkat a paraméter hozzávetőleges értékéről. Ezért a modellbeli paraméterekre úgy tekint, mint egy valószínűségi változóra.

Jelölje D az adatokat, ϑ_k a paramétert, valamint adott modellt az M . Esetünkben k a Markov lánc rendjét jelöli. A Bayes-tétel szerint azt kapjuk, hogy:

$$P(\vartheta_k|D, M_k) = \frac{P(D|\vartheta_k, M_k)P(\vartheta_k|M_k)}{P(D|M_k)},$$

ahol a $P(D|\vartheta_k, M_k)$ a *likelihood* függvény, a $P(\vartheta_k|M_k)$ eloszlást nevezik a *priori*-nak, a $P(D|M_k)$ a normáló tényező, valamint a $P(\vartheta_k|D, M_k)$ valószínűséget nevezik a *poszteriori* eloszlásnak, ami a ϑ_k paraméter mintavétel utáni eloszlása. Vizsgáljuk meg külön-külön a valószínűségeket.

A **likelihood** függvényt már korábban definiáltuk, vagyis:

$$P(D|\vartheta_k, M_k) = p(x_1) \prod_i \prod_j p_{ij}^{n_{ij}}$$

3.4.1. A priori eloszlás

Az a priori a megelőző tudást, a tapasztalatot jelenti. Megmondja, milyen valószínűséggel gondoljuk az adatok ismerete előtt, hogy a modell mennyire helyes. Az a priori eloszlás a valószínűségi változónak tekintett ϑ_k paraméter eloszlása.

Mivel az a poszteriori eloszlás a likelihood és az a priori eloszlás szorzatával arányos, ezért érdemes olyan a priori választani, aminek függvényformája megegyezik a likelihoodéval. Az azonos függvényforma biztosítja, hogy a szorzás elvégzése után az a priori és az a poszteriori formája megegyezzen, azaz ugyanabban az eloszláscsaládban maradnak. Ha egy eloszláscsaládból származik az a priori eloszlás és az a poszteriori eloszlás, akkor könnyebben végezhetőek el a számítások. Az így alkalmazott a priorikat hívjuk konjugált priornak.

Esetünkben minden i -re az i után következő állapotok kategorikus eloszlásúak. A kategorikus eloszlás konjugált a priori eloszlása a Dirichlet eloszlás. A Dirichlet eloszlás pedig nem más mint a többváltozós általánosítása a béta eloszlásnak.

Jelölje az eloszlást $Dir(\alpha)$. Ennek sűrűségfüggvénye:

$$Dir(\alpha) = \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j x_j^{\alpha_j - 1},$$

ahol a Γ a gamma függvény, a hiperparaméter $\alpha_j > 0$ minden j -re, valamint $\sum_j x_j = 1$. Az átmenet-valószínűség mátrixban minden i -re a következő lesz az a priori valószínűség:

$$Dir(\alpha_i) = \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \prod_j p_{ij}^{\alpha_{ij} - 1}$$

Tudjuk, hogy minden i -re:

$$\sum_j p_{ij} = 1$$

Így az egész átmenetmátrixra a következő lesz az a priori valószínűség:

$$P(\vartheta_k | M_k) = \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \prod_j p_{ij}^{\alpha_{ij} - 1}$$

3.4.2. A normáló tényező

Ahhoz, hogy kiszámítsuk a nevezőt, szükségünk van a ϑ_k paraméter lehetséges értékeinek a súlyozott átlagára. Ezért ki kell integrálnunk ϑ_k szerint:

$$P(D | M_k) = \int P(D | \vartheta_k, M_k) P(\vartheta_k | M_k) d\vartheta_k$$

Behelyettesítve a korábbi képleteket:

$$= \int \left(p(x_1) \prod_i \prod_j p_{ij}^{n_{ij}} \right) \left(\prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \prod_j p_{ij}^{\alpha_{ij} - 1} \right) d\vartheta_k$$

Az integrálból kiemeljük a konstansokat:

$$= p(x_1) \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \int \prod_i \prod_j p_{ij}^{n_{ij}} p_{ij}^{\alpha_{ij}-1} d\vartheta_k$$

Amiből azt kapjuk, hogy:

$$= p(x_1) \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \int \prod_{i,j} p_{ij}^{n_{ij}+\alpha_{ij}-1} d\vartheta_k$$

Tudjuk, hogy a sűrűségfüggvény integrálja megegyezik 1-gyel.

Így:

$$\int \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j x_j^{\alpha_j-1} dx = 1$$

$$\frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \int \prod_j x_j^{\alpha_j-1} dx = 1$$

$$\int \prod_j x_j^{\alpha_j-1} dx = \frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)}$$

Amelyből esetünkben az következik, hogy:

$$\int \prod_j p_{ij}^{n_{ij}+\alpha_{ij}-1} d\vartheta_k = \frac{\prod_j \Gamma(n_{ij} + \alpha_{ij})}{\Gamma(\sum_j (n_{ij} + \alpha_{ij}))}$$

Végül:

$$P(D|M_k) = p(x_1) \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \frac{\prod_j \Gamma(n_{ij} + \alpha_{ij})}{\Gamma(\sum_j (n_{ij} + \alpha_{ij}))}$$

3.4.3. A posteriori

Az a posteriori eloszlás a ϑ_k paraméter mintavétel utáni eloszlása, vagyis azt mutatja meg, hogy mennyit tudunk a nem megfigyelhető paraméterekről a megfigyelhető adatok elemzése után. Összegzi az adatokból kinyert információkat, valamint az eddigi ismereteinket.

Az eddigi levezetések alapján fel tudjuk írni az a poszteriori sűrűségét, ami nem más

mint:

$$\begin{aligned} P(\vartheta_k|D, M_k) &= \left(\prod_i \prod_j p_{ij}^{n_{ij}} \right) \prod_i \frac{\Gamma(\sum_k (n_{ik} + \alpha_{ik}))}{\prod_k \Gamma(n_{ik} + \alpha_{ik})} \prod_j p_{ij}^{\alpha_{ij}-1} \\ &= \prod_i \frac{\Gamma(\sum_k (n_{ik} + \alpha_{ik}))}{\prod_k \Gamma(n_{ik} + \alpha_{ik})} \prod_j p_{ij}^{n_{ij} + \alpha_{ij} - 1} \end{aligned}$$

Ez az egyenlőség megegyezik a Dirichlet eloszlással, ahol a paraméterek $n_j + \alpha_j$ minden i sorra, vagyis

$$P(\vartheta_k|D, M_k) = \prod_i Dir(n_i + \alpha_i)$$

Az a poszteriori eloszlás az a priori tudás és a megfigyelt adat kombinációja. Bayes becslésnek nevezzük az a poszteriori eloszlás várható értékét. A várható érték és a szórásnégyzet a következő lesz:

$$E[p_{ij}] = \frac{n_{ij} + \alpha_{ij}}{\sum_k (n_{ik} + \alpha_{ik})}$$

$$D^2[p_{ij}] = \frac{(n_{ij} + \alpha_{ij})(\sum_k (n_{ik} + \alpha_{ik}) - (n_{ij} + \alpha_{ij}))}{(\sum_k (n_{ik} + \alpha_{ik}))^2 (\sum_k (n_{ik} + \alpha_{ik}) + 1)}$$

A várható értéket tovább alakíthatjuk, mégpedig:

$$E[p_{ij}] = \frac{1}{\sum_k (n_{ik} + \alpha_{ik})} \left(\sum_k n_{ik} \frac{n_{ij}}{\sum_l n_{il}} + \sum_k \alpha_{ik} \frac{\alpha_{ij}}{\sum_l \alpha_{il}} \right)$$

Legyen $c = \frac{\sum_k n_{ik}}{\sum_k (n_{ik} + \alpha_{ik})}$, ekkor át tudjuk írni az a posteriori eloszlás várható értékét, vagyis:

$$E[p_{ij}] = c \frac{n_{ij}}{\sum_l n_{il}} + (1 - c) \frac{\alpha_{ij}}{\sum_l \alpha_{il}}$$

Így a várható érték felírható a maximum likelihood becslés és az a priori várható érték konvex lineáris kombinációjaként. Ha a megfigyelések száma nagyobb ($n_{ij} \gg$

α_{ij}), akkor c 1-hez közelít és a várható érték megegyezik a maximum likelihood becsléssel.

Eddig azt figyeltük meg, hogy mik is a valószínűségek egy bizonyos adott k -ra a modellnek. Viszont arra vagyunk kíváncsiak, hogy hogyan is dönthetünk két modell közül. Ehhez szükségünk van az adat (D) és a modell (M) együttes-valószínűség eloszlására, azaz a $P(D, M_k)$ -ra. Ezt úgy tudjuk kiszámolni, ha alkalmazzuk a bayes szabályt a modellre, valamint az adatra. Azaz:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{P(D)},$$

ahol $P(D)$ a modellek súlyozott átlaga:

$$P(D) = \sum_k P(D|M_k)P(M_k).$$

A $P(D|M_k)$ mennyiséget már korábban meghatároztuk, ami a súlyozott átlaga minden adott M_k modell ϑ_k paraméterének.

A modellek összehasonlításában két gyakori prior szokott lenni: (i) minden modell azonos valószínűségű, valamint (ii) a bonyolultabb modellhez kisebb a priori valószínűséget rendelünk, hogy elkerüljük a túlillesztést.

(i) esetben

$$P(M_k) = \frac{1}{|M|},$$

ahol $|M|$ a modellek száma.

(ii) esetben gyakori a következő:

$$P(M_k) = \frac{e^{-|I_k|}}{\sum_k e^{-|I_k|}},$$

ahol $|I_k|$ a paraméterek száma az M_k modellben, melyet az alábbi módon lehet kiszámolni:

$$|I_k| = |I|^k(|I| - 1),$$

ahol $|I|$ az állapotok száma.

Végül megkapjuk az a posteriori eloszlást, amivel ki tudjuk választani a legjobb modellt ((*ii*) esetben például):

$$P(M_k|D) = \frac{P(D|M_k)e^{-|I_k|}}{\sum_k P(D|M_k)e^{-|I_k|}}$$

A számítások jobb végrehajtásához logaritmizálhatjuk a nevezőt, hogy elkerüljük az elszállást a nagyon kicsi számok esetén. Először számoljuk ki a $P(D|M_k)$ logaritmusát majd számoljuk ki a normalizáló konstans logaritmusát, azaz:

$$\log(C) = \log \left(\sum_k e^{\log(P(D|M_k))} \right).$$

A közvetlen kiszámítása a $e^{\log(P(D|M_k))}$ -nak még elszálláshoz vezet, ezért vonjuk ki belőle a következő tagot: $E_{max} = \max(\log(P(D|M_k)))$. Végül azt kapjuk, hogy:

$$\log(C) = E_{max} + \log \left(\sum_k e^{\log(P(D|M_k)) - E_{max}} \right)$$

Egyik nagy hátránya a Bayes-módszernek, hogy gyakran nehezen számolható ki a Bayes-faktor. Pontosabban a leggyakrabban az integrálok kiszámítása a legnehezebb, ezért ennek elkerülése végett különböző alternatívákra van szükség. A likelihood hányados teszthez viszonyítva a Bayes-módszer nem teszi szükségessé, hogy a modellek egymásba legyenek ágyazva. Fő előnye, hogy egy büntetőtaggal is számol, így segíti a túlillesztés elkerülését.

4. fejezet

Módszerek használata elemzéshez

Az elemzésekhez a [6, 7, 8] irodalmat használtam.

A Markov elemzés az események sorozatát vizsgálja, és elemzi egy esemény tendenciáját, amelyet egy másik esemény követ. A Markov folyamat hasznos lehet az egymástól függő véletlen események elemzésében, vagyis olyan eseményekben, ahol a jelenlegi állapot nagy valószínűséggel függ az előző állapottól. Például nem egy jó példa az érmedobás modellezése, hiszen minden egyes alkalommal az előző esettől függetlenül kapunk eredményt. Nincs semmilyen memória arra vonatkozóan, hogy mi volt az előző dobás eredménye. A fejek és írások sorozata nem kapcsolódnak egymáshoz, mivel ők független események.

Viszont nagyon sok eseményt érint az, hogy valójában mi is történt előtte, mivel ezek nem független események. Ilyen például a tegnapi időjárás, ami nagyban befolyásolja, hogy milyen időjárásra számíthatunk a mai napon. De például a Markov elemzést könyvelésben is lehet használni arra, hogy megbecsüljék a rossz adósságokat vagy a behajthatatlan követeléseket. De a marketingben arra is használják, hogy megbecsüljék a fogyasztóknak jövőbeni márkahűségüket a vásárlási szokásuk szerint.

A Markov modell képes vizsgálni az esős, és napos idők sorozatát és elemzi annak valószínűségét, hogy egy bizonyos időjárású napot milyen másik időjárású nap követhet. A továbbiakban egy mindennapi jelenség lesz bemutatva az eddig elhangzott módszereken keresztül. Ez a jelenség Párizs időjárása 2016-os évben január 1-től egész december 31-ig. Ezekon a napokon megkülönböztetünk napos, esős vagy zivataros, valamint ködös napot. Azokat a napokat ahol sütött a nap, és nem volt a másik két jelenség, 0-val jelöljük, az esős vagy zivataros napokat 1-gyel,

míg a ködös napokat 2-vel. Az elemzés során az R és a Python programnyelvet vesszük segítségül, valamint a Philipp Singer által megírt programkódokat, amik a <https://github.com/psinger/PathTools> oldalon találhatóak meg.

4.1. Az adatok

Az adatok a <https://www.wunderground.com/> oldalról lettek letöltve. Fő szempontként az lett előtérbe helyezve, hogy ne legyen nagy az állapottér elemszáma. Három esetet különböztetünk meg, az első a napos idő, második az esős, harmadik pedig a ködös. Ezeket rendre 0,1,2 jelöli. 2016-ban az év 365 napjából 164 esetben napos idő volt, 169 esős és 32 ködös nap fordult elő.

4.2. Likelihood módszer

Korábban már be lett mutatva, hogy hogyan is működik a likelihood módszer. Először nézzük meg, hogyan is néz ki az átmenetek gyakoriság-mátrixa, valamint az átmenet-valószínűség mátrix.

Az átmenet mátrix, ami megmutatja, hogy i csúcsból hányszor mentünk j csúcsba:

$$N = \begin{pmatrix} 111 & 43 & 10 \\ 45 & 114 & 10 \\ 8 & 11 & 12 \end{pmatrix}$$

Ez alapján meg tudjuk határozni a becsült átmenet-valószínűség mátrixot:

$$P = \begin{pmatrix} 0.68 & 0.26 & 0.06 \\ 0.27 & 0.67 & 0.06 \\ 0.26 & 0.36 & 0.39 \end{pmatrix}$$

Hasonlóan eljárva a magasabb rendű modelleknél, meg tudjuk határozni a k -ad rendű modellek maximum likelihoodjait ($k = 0, 1, 2, 3$):

k	log likelihood
0	-339.23
1	-295.19
2	-287.27
3	-256.13

Most hasonlítsuk össze a log likelihood értékeket a likelihood hányados teszttel:

próbatasztika	χ^2	szabadságfok
$0\eta_1$	88.09	4
$0\eta_2$	104.02	16
$0\eta_3$	166.20	52
$1\eta_2$	15.94	12
$1\eta_3$	78.12	48
$2\eta_3$	62.18	36

A táblázatból egyértelműen látszik, hogy nem nulladrendű a folyamat, vagyis nem független az egyes napoknak az időjárása. A másodrendű modell nem illeszkedik szignifikánsan jobban mint az elsőrendű, viszont a harmadrendű modell szignifikánsan jobb mint az első kettő modell.

4.3. Információs kritériumok

Az információ elméletben két főbb módszert határoztunk meg, az AIC-et és a BIC-et. Nézzük meg mindkettőre a likelihood ismeretében a kritériumok értékeit:

k	AIC	BIC
0	62.20	-140.60
1	-17.89	-204.95
2	-9.82	-150.02

Ennél a módszernél a legkisebb AIC és BIC értéket preferáljuk, így az AIC-nél és a BIC-nél is az elsőrendű modell lesz preferálva a többihez képest.

4.4. Bayes módszer

A Bayes módszernél az a poszteriori valószínűségekre vagyunk kíváncsiak. Vizsgáljuk meg, hogy adott k -ra milyen valószínűségek jönnek ki az (i) priort használva:

k	valószínűség
0	$9.03 \cdot 10^{-15}$
1	0.99
2	$3.16 \cdot 10^{-7}$
3	$3.12 \cdot 10^{-14}$

A táblázat alapján azt lehet megállapítani, hogy a Bayes-módszer az elsőrendű modellt preferálja jobban.

5. fejezet

Összegzés

Megismerkedhettünk a diszkrét Markov láncokkal, azok általánosításával, valamint magasabb rendű Markov láncokkal. Láthattunk példát átmenet-valószínűségekre, illetve arra, hogy hogyan lehet egy mátrixban elhelyezni azokat, vagy akár egy irányított gráfként szemléltetni. Példákon keresztül láthattuk, hogy minden magasabb rendű Markov lánc visszavezethető elsőrendű Markov láncá.

Ezután a rend becslése fejezetben megismerkedtünk a likelihood módszerrel, ami a maximum likelihoodot, valamint a likelihood hányados tesztet veszi alapul. Ezután az információs elméleti kritériumokkal foglalkoztunk, amik a likelihood módszert vetették alapul, de egy egészen más megközelítéssel. Továbbá a Bayes-módszer és a Cross validation módszer is bemutatásra került, amelyek szintén egy más megközelítéssel próbálják megbecsülni a Markov lánc rendjét. Végül némely megismert módszert arra alkalmaztunk, hogy megbecsüljük Párizs 2016-os évi időjárásának rendjét.

Irodalomjegyzék

- [1] Csiszár Villó, *Diszkrét és folytonos paraméterű Markov láncok* elektronikus jegyzet
- [2] Arató Miklós, Prokaj Vilmos, Zempléni András, *Bevezetés a valószínűségszámításba és alkalmazásaiba: példákkal, szimulációkkal*, 2013 elektronikus jegyzet
- [3] Michelberger Pál, Szeidl László, Várlaki Péter *Alkalmazott folyamatstatisztika és idősor-analízis*, Typotex, 2001
- [4] Székely Balázs *Markov láncok* elektronikus tananyag
http://math.bme.hu/~szbalazs/vill/het_4_Markov.pdf
- [5] <https://hu.wikipedia.org/wiki/Markov-lánc>
- [6] Philipp Singer, Thomas Niebler, Markus Strohmaier and Andreas Hotho, *Computing Semantic Relatedness from Human Navigational Paths: A Case Study on Wikipedia*, *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol 9(4), 41-70, 2013
- [7] Philipp Singer, Denis Helic, Behnam Taraghi and Markus Strohmaier, *Detecting Memory and Structure in Human Navigation Patterns Using Markov Chain Models of Varying Order*, *PLoS ONE*, vol 9(7), 2014
- [8] Philipp Singer, Denis Helic, Andreas Hotho and Markus Strohmaier, *HypTrails: A Bayesian Approach for Comparing Hypotheses About Human Trails on the Web*, 24th International World Wide Web Conference, Florence, Italy, 2015 (Best Paper Award)
- [9] Lang Zsolt *Többváltozós statisztika*, elektronikus tananyag
http://www2.univet.hu/users/zslang/tobbvalt/Likelihood_deviancia_AIC.pdf

- [10] Paksy László *Bayes megközelítés: induktív logikára alapozott statisztika*
<http://www.staff.u-szeged.hu/rajko/Kemometria99/Abstracts/PaksyL.html>