4. Markov-láncok

1. Definíció és alapvető tulajdonságok

Legyen adott egy $\mathcal{S}$ diszkrét halmaz. Leggyakrabban $\mathcal{S}$ az egész számoknak egy halmaza, például $\mathcal{S} = \{0, 1, 2, \ldots, N\}$, $\{0, 1, 2, \ldots\}$.

1. definíció. $\mathcal{S}$ értékű valószínűségi változók egy $X_0, X_1, X_2, \ldots$ végtelen sorozatát Markov-láncnak hívjuk, ha a valószínűségi változók között a következő összefüggés fennáll. Minden $n \geq 0$ egész szám, és $j, i, i_0, 1_1, \ldots, i_{n-1} \in \mathcal{S}$ állapotok esetén teljesül a

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i) \tag{1}$$

egyenlőség. A Markov-lánc időben homogén, ha

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(X_1 = j \mid X_0 = i), \quad \text{minden } n \geq 0 \text{ esetén.}$$

- Leggyakrabban az indexre úgy gondolunk, mint az idő paramétere. Így egy Markov-lánc valamilyen időben változó véletlen jelenséget ír le.
- Fontos következménye a definíciónak, pontosabban a (1) egyenlőségnek, hogy

  $$\mathbf{P}(X_{n+m} = j_m, \ldots, X_{n+2} = j_2, X_{n+1} = j_1 \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = \\ = \mathbf{P}(X_{n+m} = j_m, \ldots, X_{n+2} = j_2, X_{n+1} = j_1 = j \mid X_n = i).$$

  Ez az egyenlőség azt jelenti, hogy ha a jelen az n időindexnél van, és a jelen értékét ismerjük, $X_n = i$, akkor a Markov-lánc jövője, $(X_{n+m} = j_m, \ldots, X_{n+2} = j_2, X_{n+1} = j_1)$, független a múlttól, az $(X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0)$ eseménytől.
- Ezentúl $\mathcal{S}$-et a Markov-lánc állapottérnek hívjuk.
- Ezentúl csak időben homogén Markov-láncokkal foglalkozunk, és elhagyjuk az időben homogén jelzőt.
- Minden Markov-lánc fejlődését meghatározza a $\mathbf{P}(X_{n+1} = j \mid X_n = i)$ egylépéses átmenetvalószínűség, ha ez adott minden $j, i \in \mathcal{S}$ esetén. Ezeket az átmenet valószínűségeket egy $|\mathcal{S}|$-szer $|\mathcal{S}|$-es P mátrixba gyűjtjük, ahol a mátrix ij indexű helyén a

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

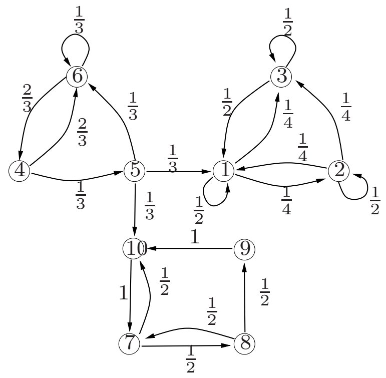
  valószínűség áll. Ha $\mathcal{S}$ végtelen, akkor P egy végtelen mátrix.

  A P mátrix neve átmenetvalószínűség mátrix, vagy egylépéses átmenetvalószínűség mátrix.
- Legyen adott állapotok egy véges hosszú sorozata: $i_1, i_2, \ldots, i_{n-1}, i_n$. Fontos tulajdonsága a Markov-láncoknak, hogy egyszerűen fel tudjuk írni annak a valószínűségét, hogy mi annak a valószínűsége, hogy a Markov-lánc egymás utáni lépéseiben pont ezeket az állapotokat veszi fel:

  $$\mathbf{P}\left(X_n = i_n, X_{n-1} = i_{n-1} \ldots, X_2 = i_2, X_1 = i_1 \mid X_1 = i_0\right) = P_{i_0 i_1} P_{i_1 i_2} \ldots P_{i_{n-1} i_n} \tag{2}$$

  Ennek az a következménye, hogy minden Markov-láncot egyértelműen meg lehet feleltetni egy irányított, súlyozott gráfon való véletlen bolyongásnak. Legyenek a gráf pontjai az $\mathcal{S}$ halmazzal indexelve. Az $i \to j$ irányított él be van húzva, ha a P_{ij} egy lépéses átmenet pozitív, az él súly a P_{ij} értéke. Ekkor a gráfon való bolyongás úgy néz ki, hogy ha n-dik lépésben a bolyongás az i állapotba ugrott, akkor a j-be ugrás valószínűsége éppen P_{ij}

  1. példa. Az 1. ábrán lévő Markov-lánc esetén annak a valószínűsége, hogy az 5-ből indulva sorrendben meglátogatja az 112 állapotokat $\frac{1}{3}\frac{1}{2}\frac{1}{4}$. Annak a valószínűsége, hogy 5-ből indulva sorrendben meglátogatja az 612 állapotokat 0.



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

Figure 1.: Markov-lánc gráf reprezentációja és átmenetvalószínűség mátrixa

- Többlépéses átmenetvalószínűségek. Legyen $P^{(n)}$ a P átmenetvalószínűség mátrixszal meghatározott Markov-lánc egylépéses átmenetvalószínűség mátrixa. Pontosabban, $P^{(n)}$ ij koordinátájú helyén a

$$P_{ij}^{(n)} = \mathbf{P}(X_n = j \mid X_0 = i), \quad i, j \in \mathcal{S}$$

valószínűség áll, azaz annak valószínűsége, hogy a Markov-lánc n lépése alatt az i állapotból eljut a j állapotba. Fontos összefüggés, hogy az n lépéses átmenetvalószínűség mátrix felírható az egylépéses átmenetvalószínűség mátrix n-dik hatványaként:

1. <u>állítás</u>. $P^{(n)} = P^n$

2. *példa.* A 1. ábrában definiált Markov-lánc esetén annak a valószínűsége, hogy 5-ból indulva 3 lépés múlva a 2-es állapotban van megegyezik a $[P^3]_{53}$ elemmel. Ez megegyezik az 5112 5122 utak valószínűségének összegével.

- Eddig mindig feltettük, hogy a Markov-lánc valamilyen i_0 állapotból indul, és úgy néztük a későbbi lépések valószínűségét. Most $\mathbf{P}(X_0 = i_0, X_1 = i_1)$, vagy $\mathbf{P}(X_n = i_n)$ kifejezésekhez hasonló valószínűségeket szeretnénk meghatározni mindenféle feltétel nélkül. Ezt úgy tudjuk megtenni, hogy feltesszük, hogy a 0 időben adott valamilyen kiindulási eloszlás, azaz adott az X_0-nak az eloszlása. Ezt az eloszlást kezdeti eloszlásnak nevezzük. A kezdeti eloszlás megadását technikailag úgy tesszük meg, hogy megadunk egy sorvektort, $\underline{a}$-t, azaz $\underline{a} = [a_i, i \in \mathcal{S}]$, úgy, hogy az i-dik elem éppen annak a valószínűsége, hogy a 0 időben a Markov-lánc az i állapotban van:

$$\mathbf{P}(X_0 = i) = a_i, \quad i \in \mathcal{S}.$$

3. *példa.* Például, ha $\mathbf{P}(X_0 = i_0) = 1$, akkor ez azt jelenti, hogy a Markov-lánc az i_0 állapotból indul. Vagy tekintsük a 1. ábrában definiált Markov-láncot. Ha feltesszük, hogy kezdetben minden állapot ugyanolyan valószínű, azaz $a_i = \frac{1}{10}$, $i = 1, 2, \ldots, 10$, akkor annak a valószínűsége, hogy az első lépés után a 3-as állapotban van az $a_3 P_{33} + a_2 P_{23} + a_1 P_{13} = \frac{1}{10}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{4}\right) = \frac{1}{10}$.

- Ekkor a (2) egyenlőséghez hasonlóan (immáron feltétel nélkül)

$$\mathbf{P}\left(X_n = i_n, X_{n-1} = i_{n-1} \ldots, X_2 = i_2, X_1 = i_1 X_1 = i_0\right) = \mathbf{P}(X_0 = i_0) P_{i_0 i_1} P_{i_1 i_2} \ldots P_{i_{n-1} i_n}$$

azaz $\mathbf{P}(X_0 = i_0) = a_{i_0}$ miatt

$$\mathbf{P}\left(X_n = i_n, X_{n-1} = i_{n-1} \ldots, X_2 = i_2, X_1 = i_1 X_1 = i_0\right) = a_{i_0} P_{i_0 i_1} P_{i_1 i_2} \ldots P_{i_{n-1} i_n}$$

4. *példa.* Tekintsük a 1. ábrában definiált Markov-láncot. Ha feltesszük, hogy kezdetben $a_i = \frac{i}{55}$, $i = 1, 2, \ldots, 10$, akkor annak a valószínűsége, hogy sorrendben az 5112 utat teszi meg az $\frac{5}{55} \frac{1}{3} \frac{1}{2} \frac{1}{4}$.

- Most már, hogy adott egy $\underline{a}$ kezdeti eloszlás, képesek vagyunk meghatározni X_n eloszlását. Tehát meg kell adnunk $\mathbf{P}(X_n = j)$-t minden $j \in \mathcal{S}$ esetén. A teljes valószínűség tételét használva:

$$\mathbf{P}(X_n = j) = \sum_{i \in \mathcal{S}} \mathbf{P}(X_n = j \mid X_0 = i) \mathbf{P}(X_0 = i) = \sum_{i \in \mathcal{S}} a_i P_{ij}^{(n)}.$$

Mivel ennek az összegnek ugyanolyan a struktúrája, mint a mátrix szorzásnak ezért ezért

$$\mathbf{P}(X_n = j) = \sum_{i \in \mathcal{S}} a_i P_{ij}^{(n)} = \left[\underline{a} P^n\right]_j ,$$

azaz amit kapunk az a $\underline{a}P^n$ sorvektor j-dik eleme.

- Tehát X_n eloszlását egy sorvektorként képzelhetjük el, ahol a sorvektor j-dik koordinátája a $\mathbf{P}(X_n = j)$ valószínűség. Ez függ a kezdeti eloszlástól. Ha jelölni szeretnénk a kezdeti eloszlástól való függést, akkor a

$$\mathbf{P}_{\underline{a}}(X_n = j)$$

kifejezést írjuk.

2. Markov-lánc állapoteres felbontása

A Markov-lánc állapotterét szeretnénk felbontani olyan osztályokra, amelyek *ha nagyon sokáig nézzük a lánc fejlődését*, akkor lényegesen máshogy viselkednek. Ehhez először definiáljunk két relációt.

2. definíció. j elérhető az i állapotból, jelben $i \to j$, ha létezik pozitíva valószínűségű út ib-ől j-be, azaz létezik egy n, hogy $\mathbf{P}(X_n = j \mid X_0 = i) > 0$.

3. definíció. i érintkezik a j állapottal, jelben $i \leftrightarrow j$, ha i elérhető jből, és j is elérhető az i-ből

2. <u>állítás</u>. Az érintkezési reláció ekvivalencia reláció, azaz, (1) reflexív $i \leftrightarrow i$, (2) szimmetrikus $i \leftrightarrow j \Rightarrow j \leftrightarrow i$, és (3) tranzitív $i \leftrightarrow j$ és $j \leftrightarrow k \Rightarrow i \leftrightarrow k$.

Ennek az a következménye, hogy az egymással érintkező állapotok egy osztályba tartoznak.
A 1. ábrán definiált Markov-lánc esetén 3 osztály van: $C_1 = \{1, 2, 3\}$, $C_2 = \{4, 5, 6\}$, $C_3 = \{7, 8, 9, 10\}$.

4. definíció. Ha egy osztályból nem megy ki pozitív valószínűségű út, akkor zártnak nevezzük.

A példában a C_1 és a C_3 osztály zárt.
Minden osztályra megmondható az, hogy mi a periódusa, és az, hogy visszatérő-e. Ezeket definiáljuk most.

5. definíció. Az i állapot *periódusa* $d(i)$ a $\{n : P(X_n = i \mid X_0 = i) > 0\}$ halmaz legnagyobb közös osztója. Ha a periódus 1, azaz $d(i) = 1$, akkor azt mondjuk, hogy az i állapot *aperiodikus*.

Bebizonyítható, hogy ha egy i állapot periódusa $d(i) = d$, akkor elegendően sok idő után minden n számra teljesül, hogy $P_{ii}^{(nd)} > 0$, azaz, d lépésenként mindig visszatér pozitív valószínűséggel, ha az i állapotból indult a Markov-lánc. Természetesen ha $d(i) = 1$, akkor ez azt jelenti, hogy egy idő után mindig pozitív valószínűséggel (lehet, hogy egyre csökkenő) tartózkodik az i állapotban, ha az i állapotból indult a Markov-lánc.

6. definíció. Egy i állapot *visszatérő*, ha i-ből indulva 1 valószínűséggel visszatér, azaz

$$\mathbf{P}(\exists n X_n = i \mid X_0 = i) = 1.$$

Egy állapot *átmeneti*, ha nem visszatérő.

3. <u>állítás</u>. A periódus és a visszatérőség osztálytulajdonság. Azaz, az egy osztályban lévő állapotoknak ugyanaz a periódusa. Továbbá, az egy osztályban lévő állapotok vagy mind visszatérők, vagy mind átmenetiek.

A 1. ábrán definiált Markov-lánc esetén a C_1 és a C_2 osztály periódusa 1, míg a C_3 osztály periódusa 2. Továbbá, a C_1 és a C_3 osztály visszatérő, míg a C_2 osztály átmeneti. Ennek a következő az oka.

4. <u>állítás</u>. Ha egy véges sok elemből álló osztály zárt, akkor visszatérő. Ha megy ki belőle pozitív valószínűségű út, akkor visszatérő.

Ez utóbbi állítás azért igaz szemléletesen, mert egy ilyen pozitív valószínűségű út "kipumpálja az ott tartózkodás valószínűségét az osztályból". Az is bebizonyítható, hogy egy véges elemből álló átmeneti osztályban tartózkodás valószínűsége exponenciálisan csökken. Pontosabban, ha C egy véges átmeneti osztály és $i \in C$ akkor,

$$\mathbf{P}(X_n \in C \mid X_0 = i) \leq K\delta^n,$$

ahol $K > 0$ és $0 < \delta < 1$.

Általában az $\mathcal{S}$ állapottér felbontható

$$\mathcal{S} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \dots \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots$$

osztályok uniójára, ahol $\mathcal{T}_i$ osztályok jelölik az átmeneti osztályokat, és $\mathcal{C}_i$ jelölik a visszatérő osztályokat.

7. definíció. Ha a Markov-lánc egy osztályból álla, akkor azt mondjuk, hogy irreducibilis (felbonthatatlan).

A visszatérő osztályok zártak, nem megy ki belőlük él.
Ha a Markov-lánc egy zárt osztályból indul, akkor végig abban az osztályban marad.
Ha a Markov-lánc egy átmeneti osztályból indul, akkor véges sok lépés alatt 1 valószínűséggel elhagyja.

3. Stacionárius eloszlás

8. definíció. Egy $X_0, X_1, \dots$ Markov-láncot (erősen) stacionárius folyamatnak nevezünk, ha minden $n \geq 1$, $m \geq 1$, és $i_0, i_1, \dots, i_n \in \mathcal{S}$ állapotsorozat esetén

$$\mathbf{P}\left(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\right) = \mathbf{P}\left(X_m = i_0, X_{m+1} = i_1, \dots, X_{m+n} = i_n\right).$$

Ez azt jelenti, hogy egy véges trajektória $(i_0, i_1, \dots, i_n)$ valószínűsége minden időben ugyanaz

9. definíció. Ha $\underline{\pi}$ sorvektor az $\mathcal{S}$ állapottéren adott egy eloszlás ($\underline{\pi} = [\pi_i, i \in \mathcal{S}]$), akkor stacionárius eloszlásnak hívjuk, ha

$$\underline{\pi} P = \underline{\pi}. \tag{3}$$

Figyeljük meg, hogy ez azt jelenti, hogy X_0 eloszlása ($\mathbf{P}(X_n = j) = \pi_j$, miden $j \in \mathcal{S}$) megegyezik X_1 eloszlásával ($\mathbf{P}(X_n = j) = [\pi P]_j$, miden $j \in \mathcal{S}$)

Ennek az a következménye, hogy minden n-re X_n-nek ugyanaz az eloszlása:

$$\mathbf{P}(X_n = j) = \pi_j, \quad \forall n \geq 0, \forall j \in \mathcal{S},$$

ugyanis, ahogy már korábban felírtuk X_n eloszlása felírható $\underline{\pi} P^n$ sorvektor alakban. Továbbá, a (3)-ból az következik, hogy nem csak egy hatványra, hanem tetszőleges n-re igaz, hogy

$$\underline{\pi} P^n = \underline{\pi}.$$

Bizonyítható a

5. <u>állítás</u>. Ha $\underline{\pi}$ stacionárius eloszlás, és ez a Markov-lánc kezdeti eloszlása ($\underline{a} = \underline{\pi}$), akkor az $X_0, X_1, \dots$ Markov-lánc stacionárius.

Érdekes, hogy ha

1. <u>tétel</u>. Ha egy Markov-lánc irreducibilis és aperiodikus, akkor az

$$\underline{y} P = \underline{y}$$

sajátérték egyenletnek létezik egyetlen megoldása.

Két egymást kizáró eset lehetséges:

(1) Ha az $\underline{y}$ sorvektor elemeinek az összege véges, azaz $\sum_{i \in \mathcal{S}} y_i < \infty$, akkor létezik egyetlen stacionárius eloszlás $\underline{\pi}$, ahol

$$\pi_j = \frac{y_j}{\sum_{i \in \mathcal{S}} y_i}$$

pozitív szám.

(2) Ha az $\underline{y}$ sorvektor elemeinek az összege véges, azaz $\sum_{i\in\mathcal{S}} y_i = \infty$, akkor nem létezik stacionárius eloszlás.

1. *megjegyzés.* Fontos megjegyzés, hogy ha az állapottér véges, akkor az $\sum_{i\in\mathcal{S}} y_i$ összeg automatikusan véges, így minden irreducibilis és aperiodikus Markov-láncnak létezik stacionárius eloszlása.

5. *példa.* Tekintsük a 1. ábrán definiált Markov-láncot megszorítva a C_1 osztályra. Ez egy irreducibilis, aperiodikus Markov-lánc. Határozzuk meg a stacionárius eloszlását. Ehhez meg kell oldani az

$$[\pi_1\ \pi_2\ \pi_3] \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \end{bmatrix} = [\pi_1\ \pi_2\ \pi_3]$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

egyenletrendszert. Azt kapjuk, hogy $\pi_1 = \frac{4}{9}$, $\pi_2 = \frac{2}{9}$, $\pi_3 = \frac{3}{9}$.

6. *példa.* Legyen $\mathcal{S} = \{0, 1, 2, \dots\}$. Legyen adott egy $p_1, p_2, \dots$ sorozat, ahol $0 < p_i < 1$ teljesül minden $i = 1, 2, \dots$ számra.

Tekintsük a következő Markov-láncot az $\mathcal{S}$ állapottéren.

$$P_{i,i+1} = p_i, \quad P_{i,i-1} = 1 - p_i =: q_i, \quad \text{ha } i = 1, 2, \dots \text{ és}$$

$$P_{01} = 1.$$

Határozzuk meg, hogy mikor létezik a Markov-láncnak stacionárius eloszlása. Írjuk fel a stacionárius eloszlást.

Megoldás: Az 1. tétel alapján az $\underline{y}P = \underline{y}$ egyenletet kell megoldani első lépésben. Azt kapjuk, hogy

$$\begin{aligned} y_0 &= y_1 q_1 \\ y_j &= y_{j+1} q_{j+1} + y_{j-1} p_{j-1}, \quad j \geq 1. \end{aligned}$$

Ebből egy rekurziót lehet kapni y_j-kre. $p_j + q_j = 1$ miatt azt kapjuk, hogy

$$\begin{aligned} y_0 p_0 &= y_1 q_1 \\ y_{j+1} q_{j+1} - y_j p_j = y_j q_j - y_{j-1} p_{j-1}, \quad j \geq 1. \end{aligned}$$

A szomszédos sorokat összeadva, majd egyszerűsítve azt kapjuk, hogy

$$y_{j+1} q_{j+1} = y_j p_j, \quad j \geq 0,$$

amiből a rekurzió az

$$y_{j+1} = \frac{p_0 p_1 \dots p_j}{q_1 q_2 \dots q_{j+1}} y_0, \quad j \geq 0$$

alakot ölti. Ekkor a 1. tétel szerint pontosan akkor van stacionárius eloszlás, ha y_j-k összege véges, azaz

$$\sum_{j=0}^{\infty} y_j / y_0 = \sum_{j=0}^{\infty} \frac{p_0 p_1 \dots p_j}{q_1 q_2 \dots q_{j+1}} < \infty.$$

Abban a speciális esetben, mikor $p_i = p$ rögzített szám ($q_i = q := 1 - p$), akkor ez a feltétel az

$$p \sum_{j=0}^{\infty} \left(\frac{p}{q}\right)^j = p \frac{1}{1 - \frac{p}{q}} < \infty.$$

alakot ölti. Ez akkor teljesül, ha $p < \frac{1}{2}$. Ekkor a stacionárius eloszlás

$$\pi_j = \rho^j \pi_0,$$

ahol $\rho = \frac{p}{q}$.

2. <u>tétel</u>. Ha egy Markov-lánc irreducibilis és aperiodikus és létezik stacionárius eloszlása, akkor X_n eloszlása n növekedésével konvergál a stacionárius eloszláshoz függetlenül attól, hogy mi a kezdeti eloszlás. Pontosabban,

$$\mathbf{P}_{\underline{a}}(X_n = j) \to \pi_j, \quad n \to \infty,$$

vagy mátrixosan megfogalmazva

$$\underline{a}P^n \to \underline{\pi}, \quad n \to \infty,$$

ahol a sorvektorok konvergenciája azt jelenti, hogy minden koordinátában konvergál.

Nagyon fontos tulajdonsága *véges* állapotterű Markov-láncoknak, hogy a fenti konvergencia exponenciálisan gyors, azaz

3. <u>tétel</u>. Ha egy véges állapotterű Markov-lánc irreducibilis és aperiodikus (ekkor létezik stacionárius eloszlása), akkor

$$\left|\mathbf{P}_{\underline{a}}(X_n = j) - \pi_j\right| \le C\delta^n,$$

ahol $C > 0$ és $0 < \delta < 1$. Vagy mátrixosan megfogalmazva, minden $j \in \mathcal{S}$ állapotra

$$|[\underline{a}P^n]_j - \pi_j| \le C\delta^n.$$

2. megjegyzés. A fenti exponenciális konvergencia nem csak véges állapotterű Markov-láncokra igaz. Bizonyos feltételek mellett végtelen állapotterű Markov-láncokra is teljesül.

4. Ergod-tétel Markov-láncokra

Legyen $X_0, X_1, \ldots$ egy Markov-lánc $\mathcal{S}$ állapottérrel. Továbbá, adott egy valós értékű f költségfüggvény az állapotokon értelmezve:

$$f : \mathcal{S} \to \mathbb{R}$$

Ha a Markov-láncnak adott egy $i_0 i_1 \ldots i_n$ trajektóriája, akkor ennek a trajektóriának a költsége

$$f(i_0) + f(i_1) + \cdots + f(i_n).$$

Ha nem határozzuk meg a trajektóriát, akkor az első $n+1$ állapot költsége egy valószínűségei változó

$$f(X_0) + f(X_1) + \cdots + f(X_n).$$

Ekkor egy lépésre jutó átlagos költség

$$\frac{f(X_0) + f(X_1) + \cdots + f(X_n)}{n+1}.$$

Arra vagyunk kíváncsiak, hogyha a Markov-láncot végtelen sokáig futtatjuk, akkor mi lesz az egy lépésre jutó költség, vagy másként a *hosszútávú átlagos* költség:

$$\lim_{n\to\infty} \frac{f(X_0) + f(X_1) + \cdots + f(X_n)}{n+1}$$

Erre a választ a következő tétel adja meg.

4. <u>tétel</u>. Legyen $X_0, X_1, \ldots$ egy irreducibilis, aperiodikus Markov-lánc, amelynek létezik stacionárius eloszlása, $\underline{\pi}$. Továbbá legyen adott egy $f : \mathcal{S} \to \mathbb{R}$ költségfüggvény úgy, hogy a stacionárius eloszlás szerinti átlaga véges,

$$\sum_{i\in\mathcal{S}} f(i)\pi_i < \infty.$$

Ekkor a hosszútávú átlagos költség egyszerűen számolható,

$$\lim_{n\to\infty} \frac{f(X_0) + f(X_1) + \cdots + f(X_n)}{n+1} = \sum_{i\in\mathcal{S}} f(i)\pi_i.$$

7. példa. Tekintsük a 1. ábrán definiált Markov-láncot megszorítva a C_1 osztályra. Vezessük be a következő költségfüggvényt:

$$f(1) = 10, \quad f(2) = 20, \quad f(3) = 40.$$

Határozzuk meg a Markov-lánc lépéseinek hosszútávú átlagos költségét. Az 4. tételből következik, hogy ehhez ki kell számolni a stacionárius eloszlást. Az 5. példában ezt kiszámoltuk. Tehát a 4. tétel alapján a válasz

$$10\frac{4}{9} + 20\frac{2}{9} + 40\frac{3}{9}.$$

5. Reverzibilis Markov-láncok

A 3. tételbeli gyors konvergenciát felhasználhatjuk arra, hogy egy tetszőleges, véges $\mathcal{S}$ halmazon adott $\underline{\pi}$ eloszlásból generáljunk mintát. A Később majd ejtünk róla szót, hogy miért fontos, hogy meg tudjunk oldani ilyen típusú feladatokat.

Azt fogjuk tenni, hogy megadunk egy Markov-láncot, amelynek a stacionárius eloszlása éppen $\underline{\pi}$. Ehhez először reverzibilis/megfordítható Markov-láncokat tanulmányozzuk.

10. definíció. Egy $X_0, X_1, \ldots, X_T$ Markov-lánc megfordítása az $X_0^*, X_1^*, \ldots, X_T^*$ folyamat, ahol $X_t^* = X_{T-t}$

6. <u>állítás</u>. Az $X_0^*, X_1^*, \ldots, X_T^*$ folyamat Markov-lánc.

Bizonyítás. A megfordítás definíciója szerint:

$$\mathbf{P}(X_{t+1}^* = j \mid X_t^* = i, X_{t-1}^* = i_{t-1}, \ldots, X_0^* = i_0) = \\ = \mathbf{P}(X_{T-(t+1)} = j \mid X_{T-t} = i, X_{T-(t-1)} = i_{t-1}, \ldots, X_T = i_0).$$

Mivel adott jelen, $X_{T-t} = i$, mellett a Markov lánc jövője $(X_{T-(t-1)} = i_{t-1}, \ldots, X_T = i_0)$ és a múltja $(X_{T-(t+1)} = j)$ függetlenek, ezért a jobboldal egyenlő $\mathbf{P}(X_{T-(t+1)} = j \mid X_{T-t} = i)$ valószínűséggel, ami pedig éppen $\mathbf{P}(X_{(t+1)}^* = j \mid X_t^* = i)$. Azaz a megfordított folyamat is Markov-lánc. □

7. <u>állítás</u>. Ha az $X_0, X_1, \ldots, X_T$ Markov-lánc stacionárius, akkor a megfordított Markov-lánc is stacionárius,

$$P_{ij}^* = \frac{P_{ji}\pi_j}{\pi_i} \tag{4}$$

átmenetvalószínűséggel, ahol $\underline{\pi}$ az eredeti Markov-lánc stacionárius eloszlása.

Proof. A Bayes-tételt használva, majd azt, hogy a Markov-lánc a stacionárius eloszlásból indul, így minden n-re $\mathbf{P}(X_n = j) = \pi_j$.

$$P_{ij}^* = \mathbf{P}(X_{t+1}^* = j \mid X_t^* = i) = \mathbf{P}(X_{T-t-1} = j \mid X_{T-t} = i) = \frac{\mathbf{P}(X_{T-t-1} = j X_{T-t} = i)}{\mathbf{P}(X_{T-t} = i)} = \\ = \frac{\mathbf{P}(X_{T-t} = i \mid X_{T-t-1} = j)\mathbf{P}(X_{T-t-1} = j)}{\mathbf{P}(X_{T-t} = i)} = \frac{P_{ji}\pi_j}{\pi_i}.$$

□

Ezek után definiáljuk ezen alfejezet legfontosabb fogalmát

11. definíció. Egy $X_0, X_1, \ldots, X_T$ Markov-láncot megfordíthatónak nevezünk, ha eloszlása megegyezik a megfordításának eloszlásával, azaz

$$P_{ij}^* = P_{ij}.$$

A (4) egyenletből következik, hogy a megfordítható Markov-lánc átmenetvalószínűségére fenáll

$$P_{ij}^* = \frac{P_{ji}\pi_j}{\pi_i} = P_{ij},$$

másképpen

$$\pi_j P_{ji} = \pi_i P_{ij}, \quad \text{minden } i, j \in \mathcal{S}. \tag{5}$$

Ha ez az egyenlőség fennáll egy Markov-láncra, akkor a Markov-lánc megfordítható.

• A megfordítható Markov-láncok reprezentálhatók egy *irányítatlan* súlyozott gráfon való bolyongással a következőképpen. Legyen adott egy gráf, ahol a csúcspontok az $\mathcal{S}$ halmazzal vannak indexelve. Ha az $i - j$ él be van húzva, akkor az él súlya legyen $w_{ij} = w_{ji} > 0$. Ekkor a bolyongás a következőképpen fejlődik. Ha az n-dik lépésben az i állapotban van, akkor

$$P_{ij} = \frac{w_{ij}}{w_i}$$

valószínűséggel lép a j állapotba a következő lépésben, ahol

$$w_i = \sum_{i-j \text{ él be van húzva}} w_{ij}$$

Így egy Markov-láncot definiáltunk, amelynek a stacionárius eloszlása

$$\pi_j = \frac{w_j}{w},$$

ahol

$$w = \sum_{i \in \mathcal{S}} w_i$$

a súlyok összege. Azt, hogy ez a stacionárius eloszlás, látszik a következő állítás bizonyításából.

8. <u>állítás</u>. A fent definiált Markov-lánc reverzibilis.

Proof. Elég ellenőrizni a (5) egyenlőség meglétét. Ez viszont teljesül:

$$\pi_j P_{ji} = \frac{w_j}{w}\frac{w_{ji}}{w_j} = \frac{w_{ji}}{w} = \frac{w_{ij}}{w} = \frac{w_i}{w}\frac{w_{ij}}{w_i} = \pi_i P_{ij}.$$

□

8. *példa.* Mutassuk meg, hogy az 6. példában definiált Markov-lánc reverzibilis.

• Markov chain Monte Carlo (MCMC). Ebben a pontba írjuk le az alfejezet kiindulási problémájának megoldását. Tehát, ha adott egy $\mathcal{S}$ állapottér és egy $\underline{\pi}$ eloszlás $\mathcal{S}$-en, akkor készítünk egy reverzibilis Markov-láncot, amelynek a stacionárius eloszlása éppen $\underline{\pi}$. Mivel $\mathcal{S}$ véges, a lépések (n) számának növelésével a Markov-lánc n időben vett eloszlása exponenciálisan gyorsan konvergál a stacionárius eloszlásához, $\underline{\pi}$-hez. Tehát, hogyan konstruálunk ilyen reverzibilis Markov-láncot? Meg kell határozni P-t, az átment valószínűség mátrixot.

Induljunk ki (5) karakterizációból, azaz az ismeretlen P_{ij}-nek ki kell elégíteni a

$$\frac{P_{ij}}{P_{ji}} = \frac{\pi_j}{\pi_i}$$

egyenlőséget. Vegyük a logaritmusát:

$$\log P_{ij} - \log P_{ji} = \log \pi_j - \log \pi_i.$$

Átrendezve azt kapjuk, hogy

$$\log P_{ij} = \log P_{ji} + \log \pi_j - \log \pi_i,$$

továbbá

$$\begin{aligned}\log P_{ij} - \frac{1}{2}(\log \pi_j - \log \pi_i) &= \log P_{ji} + \frac{1}{2}(\log \pi_j - \log \pi_i) \\ \log P_{ij} - \frac{1}{2}(\log \pi_j - \log \pi_i) &= \log P_{ji} - \frac{1}{2}(\log \pi_i - \log \pi_j). \qquad (6)\end{aligned}$$

Legyen S' egy $|\mathcal{S}| \times |\mathcal{S}|$ mátrix, amelynek az ij-dik eleme legyen

$$S'_{ij} := \log P_{ij} - \frac{1}{2}(\log \pi_j - \log \pi_i), \quad \text{ha } i \neq j.$$

Ekkor az (6) egyenlőség alapján S' szimmetrikus mátrix (i és j felcserélésével ugyanazt a számot kapjuk). Ekkor az ismeretlen P_{ij} átmenetvalószínűségekre teljesül a

$$\log P_{ij} = S'_{ij} + \frac{1}{2}(\log \pi_j - \log \pi_i) = S'_{ij} - \frac{1}{2}(\log \pi_i - \log \pi_j), \quad \text{ha } i \neq j.$$

egyenlőség. Ekkor bevezetve az S szimmetrikus mátrixot, amelyet az $S_{ij} = e^{S'_{ij}}$ egyenlőség definiál minden ij párra, azt kapjuk, hogy

$$P_{ij} = S_{ij} e^{-\frac{1}{2}} \frac{\pi_i}{\pi_j}, \quad \text{ha } i \neq j.$$

Most meg kell választanunk P_{ii} átmenet valószínűségeket, $i \in \mathcal{S}$. Arra kell vigyáznunk, hogy ne legyen $\sum_{j:j\neq i} P_{ij} > 1$. Mivel S_{ij}-k helyett tetszőleges $\delta > 0$ számszorosa is jó szimmetrikus mátrixnak, ezért legyenek S_{ij}-k $(i, j \in \mathcal{S})$ akkorák, hogy

$$\sum_{j:j\neq i} P_{ij} \leq 1, \quad \text{azaz} \quad \sum_{j:j\neq i} S_{ij} e^{-\frac{1}{2}} \frac{\pi_i}{\pi_j} \leq 1.$$

Ekkor már definiálhatjuk P_{ii}-t, legyen

$$P_{ii} := 1 - \sum_{j:j\neq i} P_{ij}.$$

Szokásos választás, ha az S mátrixnak az incidencia mátrix C, valamely δ-szorosát vesszük $(\delta > 0)$. Emlékeztetőül, a incidencia mátrix egy négyzetes mátrix, amely a csúcsok közötti kapcsolatot írja le úgy, hogy

$$C_{ij} = \begin{cases} 1 & \text{ha az } i - j \text{ él be van húzva,} \\ 0 & \text{különben.} \end{cases}$$

Ekkor

$$P_{ij} = \delta C_{ij} \frac{\pi_i}{\pi_j}.$$