Felsőbb Matematika Informatikusoknak – Sztochasztika
házi feladatok a 2. részhez – megoldások
2016 ősz

Minden héten összesen egy pontot érnek a kitűzött feladatok.

1.HF: (Beadási határidő: 2016.11.04.)

HF 1.1 A postafiókomba az emailek Poisson folyamat szerint érkeznek (éjjel-nappal egyenletesen), naponta átlagosan 24. Minden email a többitől függetlenül $\frac{1}{4}$ valószínűséggel spam; $\frac{1}{2}$ valószínűséggel nem spam, de nincs is vele teendő; a maradék $\frac{1}{4}$ valószínűséggel viszont azt eredményezi, hogy valamit csinálni kell.

a.) Mennyi annak a valószínűsége, hogy egy reggel 8-tól 10 óráig 4-nél több emailt kapok?

b.) Mennyi annak a valószínűsége, hogy 8-tól 10 óráig legalább 3 olyan email érkezik, amivel teendő lesz?

c.) Mennyi annak a valószínűsége, hogy 8-tól 10 óráig legalább 5 email érkezik, de spam egy se? (*Vigyázat: Az emailek száma és a spamek száma nem független! Tipp: a spamek száma viszont független a többi email számától. Miért is?)*

Megoldás:

a.) Legyen $X_{[8,10]}$ a reggel 8 és 10 között érkező emailek száma. Mivel az emailek Poisson folyamat szerint érkeznek, $X_{[8,10]}$ Poisson eloszlású. Várható értéke 2, mivel óránként átlag 1 email érkezik. Vagyis $X_{[8,10]} \sim Poi(2)$, vagyis $\mathbb{P}(X_{[8,10]} = k) = e^{-2}\frac{2^k}{k!}$, $k = 0, 1, 2, \ldots$-re. Ebből

$$\begin{aligned}
\mathbb{P}(X_{[8,10]} > 4) &= 1 - [\mathbb{P}(X_{[8,10]} = 0) + \mathbb{P}(X_{[8,10]} = 1) + \cdots + \mathbb{P}(X_{[8,10]} = 4)] = \\
&= 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}\right) = \\
&= 1 - e^{-2}\left(1 + 2 + \frac{4}{2} + \frac{8}{6} + \frac{16}{24}\right) = 1 - 7e^{-2} \approx 0.053 = 5.3\%.
\end{aligned}$$

b.) Legyen $Y_{[8,10]}$ a reggel 8 és 10 között érkező olyan emailek száma, amivel teendő lesz. Az ilyen emailek folyamata az összes email Poisson-folyamatának ritkítása $p = \frac{1}{4}$-szeresére, így maga is Poisson-folyamat, negyedakkora rátával. Ezért $Y_{[8,10]} \sim Poi\left(\frac{1}{2}\right)$, vagyis $\mathbb{P}(Y_{[8,10]} = k) = e^{-\frac{1}{2}}\frac{\left(\frac{1}{2}\right)^k}{k!}$, $k = 0, 1, 2, \ldots$-re. Ebből

$$\begin{aligned}
\mathbb{P}(Y_{[8,10]} \geq 3) &= 1 - [\mathbb{P}(Y_{[8,10]} = 0) + \mathbb{P}(Y_{[8,10]} = 1) + \mathbb{P}(Y_{[8,10]} = 2)] = \\
&= 1 - e^{-\frac{1}{2}}\left(\frac{\left(\frac{1}{2}\right)^0}{0!} + \frac{\left(\frac{1}{2}\right)^1}{1!} + \frac{\left(\frac{1}{2}\right)^2}{2!}\right) = \\
&= 1 - e^{-\frac{1}{2}}\left(1 + \frac{1}{2} + \frac{\frac{1}{4}}{2}\right) = 1 - \frac{13}{8}e^{-\frac{1}{2}} \approx 0.014 = 1.4\%.
\end{aligned}$$

c.) Tekintsük külön a spamek folyamatát és a *nem spam* emailek folyamatát. Mindkettő az összes email Poisson-folyamatának ritkítása, *sőt*: az egyik egy ritkítás során megmaradó eseményekből áll, a másik pedig az ugyanezen ritkítás során kidobott eseményekből. Előadásról tudjuk, hogy ezek nem csak külön-külön Poisson folyamatok, hanem *függetlenek is*.
Hogy kicsit kevesebb legyen a jelölés, mint az előző két pontban, nem írok több indexet... Legyen Z a reggel 8 és 10 között érkező *nem spam* emailek száma, V

pedig a reggel 8 és 10 között érkező spamek száma. A fentiek szerint $Z \sim Poi\left(\frac{3}{2}\right)$, $V \sim Poi\left(\frac{1}{2}\right)$ és függetlenek, vagyis

$$\begin{aligned}\mathbb{P}(Z \geq 5, V = 0) &= \mathbb{P}(Z \geq 5)\mathbb{P}(V = 0) = \\ &= \left[1 - e^{-\frac{3}{2}}\left(1 + \frac{3}{2} + \frac{1}{2}\left(\frac{3}{2}\right)^2 + \frac{1}{6}\left(\frac{3}{2}\right)^3 + \frac{1}{24}\left(\frac{3}{2}\right)^4\right)\right]e^{-\frac{1}{2}} \\ &\approx 0.011 = 1.1\%.\end{aligned}$$

2.HF: (Beadási határidő: 2016.11.14.)

HF 2.1 Az X valószínűségi változó generátorfüggvénye $g(z) = \frac{1}{(2-z)^2}$.

a.) $\mathbb{P}(X = 0) =?$
b.) $\mathbb{P}(X = 1) =?$
c.) $\mathbf{E}X =?$

Megoldás:

a.) $\mathbb{P}(X = 0) = g(0) = \frac{1}{(2-0)^2} = \frac{1}{4}$.
b.) $g'(z) = \frac{-2}{(2-z)^3}(-1) = \frac{2}{(2-z)^3}$, így $\mathbb{P}(X = 1) = g'(0) = \frac{2}{(2-0)^3} = \frac{1}{4}$.
c.) $\mathbf{E}X = g'(1) = \frac{2}{(2-1)^3} = 2$.

(Alternatív megoldás: órán láttuk, hogy a p paraméterű pesszimista geometriai eloszlás generátorfüggvénye $g_{PessGeom(p)}(z) = \frac{p}{1-qz}$, ami $p = \frac{1}{2}$-re éppen $g_{PessGeom(\frac{1}{2})}(z) = \frac{\frac{1}{2}}{1-\frac{1}{2}z} = \frac{1}{2-z}$. Tudjuk továbbá, hogy független val.változók összegének generátorfüggvénye éppen a generátorfüggvények szorzata. A feladat-beli generátorfüggvényre $g(z) = g_{PessGeom(\frac{1}{2})}(z) \cdot g_{PessGeom(\frac{1}{2})}(z)$, vagyis ha U és V független pesszimista geometriai eloszlású val.változók $p = \frac{1}{2}$ paraméterrel és $X = U + V$, akkor X generátorfüggvénye éppen g. A feladat kérdéseire ebből is azonnal adódik a válasz.)

HF 2.2 Egy atomreaktorban sok olyan atommag van, ami maghasadásra képes, ha egy neutron eltalálja. Egy, a reaktorba bekerülő neutron sorsa kétféle lehet:

i.) Kirepül a reaktorból, vagy elnyelődik (pl. egy nem hasadó atommagban) anélkül, hogy hasadást okozna - és ezzel elvész a láncreakció számára. Ennek valószínűsége legyen p.
ii.) Egy hasadásra képes magot eltalálva hasadást okoz. Ő maga elnyelődik, helyette a hasadás során véletlen számú másik neutron szabadul fel: 1, 2 vagy 3, azonos (vagyis $\frac{1-p}{3}$) valószínűséggel.

Az egyes neutronok sorsa független egymástól és az előzményektől. A p paraméter értéke a reaktor méretének és alakjának változtatásával (vagyis szabályozórudak mozgatásával) állítható.

Belövünk a reaktorba egyetlen neutront – legyen ez egymaga a „nulladik generáció". Az „első generáció" álljon a legelső neutron által okozott hasadás során létrejövő neutronokból (ha van ilyen). A „második generáció" álljon az első generáció tagjai által okozott hasadások során létrejövő neutronokból. És így tovább, az n+1-edik generáció álljon az n-edik generáció tagjai által okozott hasadások során létrejövő neutronokból, $n = 0, 1, 2, \ldots$-re. Jelölje Z_n az n-edik generáció elemeinek számát. Legyen $X = Z_1$, és legyen $N = \sum_{n=0}^{\infty} Z_n$ a láncreakcióban résztvevő neutronok össz-száma.

Válaszoljuk meg az alábbi kérdéseket

I.) ha $p = \frac{5}{8}$
II.) ha $p = \frac{1}{4}$.

a.) Mi X eloszlása?

b.) $\mathbf{E}X =?$

c.) Mi X generátorfüggvénye?

d.) $\mathbf{E}Z_{20} =?$

e.) Mi Z_2 generátorfüggvénye?

f.) $\mathbf{E}N =?$

g.) $\mathbb{P}(Z_4 = 0) =?$

h.) Mennyi a valószínűsége, hogy a láncreakció előbb-utóbb leáll, vagyis hogy valamelyik generáció már üres?

Megoldás:

I.) a.)

k	0	1	2	3
$p_k = \mathbb{P}(X = k)$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

b.) $m := \mathbf{E}X = 0 \cdot \frac{5}{8} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = \frac{3}{4}$.

c.) X generátorfüggvénye $g(z) = \sum_{k=0}^{\infty} p_k z^k = \frac{5}{8}z^0 + \frac{1}{8}z^1 + \frac{1}{8}z^2 + \frac{1}{8}z^3 = \frac{5+z+z^2+z^3}{8}$.

d.) $\mathbf{E}Z_{20} = m_{20} = m^{20} = \left(\frac{3}{4}\right)^{20} \approx 0.00317$.

e.) Z_2 generátorfüggvénye

$$g_2(z) = g(g(z)) = \frac{5 + \frac{5+z+z^2+z^3}{8} + \left(\frac{5+z+z^2+z^3}{8}\right)^2 + \left(\frac{5+z+z^2+z^3}{8}\right)^3}{8}.$$

f.) Mivel $m < 1$ (a folyamat szubkritikus), $\mathbf{E}N = \frac{1}{1-m} = \frac{1}{1-\frac{3}{4}} = 4$.

g.) Az $r_n := \mathbb{P}(Z_n = 0)$ sorozatról tudjuk, hogy eleget tesz az $r_{n+1} = g(r_n)$ rekurziós szabálynak $r_0 = 0$ kezdeti értékkel. Így az r_n-ek egyesével számolhatók:

* $r_0 = 0$
* $r_1 = g(r_0) = g(0) = \frac{5}{8}$
* $r_2 = g(r_1) = g(\frac{5}{8}) \approx 0.78247$
* $r_3 = g(r_2) \approx 0.859226$
* $r_4 = g(r_3) \approx 0.90398$.

h.) $m < 1$, vagyis a folyamat szubkritikus (és nem elfajult, vagyis $\mathbb{P}(X = 1) \neq 1$), ezért a kihalás valószínűsége $r_\infty = 1$.

II.) a.)

k	0	1	2	3
$p_k = \mathbb{P}(X = k)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

b.) $m := \mathbf{E}X = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{3}{2}$.

c.) X generátorfüggvénye $g(z) = \sum_{k=0}^{\infty} p_k z^k = \frac{1}{4}z^0 + \frac{1}{4}z^1 + \frac{1}{4}z^2 + \frac{1}{4}z^3 = \frac{1+z+z^2+z^3}{4}$.

d.) $\mathbf{E}Z_{20} = m_{20} = m^{20} = \left(\frac{3}{2}\right)^{20} \approx 3325$.

e.) Z_2 generátorfüggvénye

$$g_2(z) = g(g(z)) = \frac{1 + \frac{1+z+z^2+z^3}{4} + \left(\frac{1+z+z^2+z^3}{4}\right)^2 + \left(\frac{1+z+z^2+z^3}{4}\right)^3}{4}.$$

f.) Mivel $m \geq 1$, $\mathbf{E}N = \infty$.

g.) Az $r_n := \mathbb{P}(Z_n = 0)$ sorozatról tudjuk, hogy eleget tesz az $r_{n+1} = g(r_n)$ rekurziós szabálynak $r_0 = 0$ kezdeti értékkel. Így az r_n-ek egyesével számolhatók:

* $r_0 = 0$
* $r_1 = g(r_0) = g(0) = \frac{1}{4}$

* $r_2 = g(r_1) = g(\frac{1}{4}) \approx 0.33203$
* $r_3 = g(r_2) \approx 0.36972$
* $r_4 = g(r_3) \approx 0.38924.$

h.) $m > 1$, vagyis a folyamat szuperkritikus, ezért számolni kell: a kihalás valószínűsége a $z = g(z)$ fixpont-egyenlet egyetlen $[0, 1)$-beli megoldása. Vagyis meg kell oldani a

$$z = \frac{1 + z + z^2 + z^3}{4}$$

egyenletet, ami nullára redukálva

$$z^3 + z^2 - 3z + 1 = 0.$$

Ez első ránézésre ijesztő harmadfokú, de szerencsére tudjuk, hogy $g(1) = 1$ mindig, így $z = 1$ mindig megoldás. Vagyis a baloldalból $z - 1$ kiemelhető. És valóban:

$$z^3 + z^2 - 3z + 1 = (z - 1)(z^2 + 2z - 1),$$

tehát az egyenletünk

$$(z - 1)(z^2 + 2z - 1) = 0.$$

Ebből $z - 1 = 0$ vagy $z^2 + 2z - 1 = 0$. Ez utóbbi már csak másodfokú, megoldásai $z = -1 \pm \sqrt{2}$.
Összefoglalva: a $z = g(z)$ fixpontegyenlet megoldásai $-1-\sqrt{2}$, $-1+\sqrt{2}$ és 1. Ezek közül pontosan egy esik a $[0, 1)$ intervallumba (ahogy annak lenni kell), és ez a kihalási valószínűség:

$$r_\infty = \mathbb{P}(\text{kihalás}) = -1 + \sqrt{2} \approx 0.4142$$

3.HF: (Beadási határidő: 2016.11.28.)

HF 3.1 Móricka 10000-szer dob egy szabályos dobókockával.

a.) Közelítsük a centrális határeloszlás tétel (CHT) segítségével annak valószínűségét, hogy a dobott számok átlaga legalább 3.7.

b.) A Berry-Esseen tétel szerint legfeljebb mennyi lehet a CHT közelítés hibája?

Megoldás: Legyen $n = 10000$ és $i = 1, 2, \dots, n$-re X_i az i-edik dobás eredménye. Legyen $S_n = X_i + \cdots + X_n$ a dobott számok összege. Feladat a $\mathbb{P}\left(\frac{S_n}{n} \geq 3.7\right) = \mathbb{P}(S_n \geq 37000)$ valószínűség közelítése. Az X_i-k függetlenek és azonos eloszlásúak.

a.) $m := \mathbf{E}X_i = \frac{1+2+3+4+5+6}{6} = 3.5$, $\mathbf{E}X_i^2 = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$, vagyis $\sigma^2 := Var(X_i) = \mathbf{E}X_i^2 - m^2 = \frac{35}{12}$. Így S_n váratő értéke $\mathbf{E}S_n = nm = 35000$, szórása pedig $\mathbf{D}S_n = \sqrt{n}\sigma \approx 170.8$. Vagyis ha a várható értéket 2000-rel túl akarjuk lépni, az a szórás körülbelül 11.7-szeresét jelenti. A CHT szerint

$$\begin{aligned}\mathbb{P}(S_n \geq 37000) &= \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \geq \frac{37000 - 35000}{100\sqrt{\frac{35}{12}}}\right) \approx \\ &\approx \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \geq 11.0\right) \approx 1 - \Phi(11.7) = \Phi(-11.7),\end{aligned}$$

ahol Φ a standard normális eloszlásfüggvény. Ezt kevés táblázatból lehet 11.7-ben kiolvasni, de az mindegyikből látszik, hogy $\Phi(11.7) > 0.9999$, vagyis $1-\Phi(11.7) < 10^{-4}$. A táblázatkezelőm szerint viszont $\Phi(-11.7) \approx 6.4 \cdot 10^{-32}$.

b.) Ehhez kell még az X_i-k eloszlásának $\delta := \mathbf{E}(|X_i - m|^3)$ jellemzője. Esetünkben

$$\delta = \frac{|1-3.5|^3 + |2-3.5|^3 + |3-3.5|^3 + |4-3.5|^3 + |5-3.5|^3 + |6-3.5|^3}{6} = \frac{51}{8}.$$

Vagyis a Berry-Esseen tétel szerint a CHT közelítés hibájára $C = 0.4748$-cal

$$\text{hiba} \leq \frac{C\delta}{\sqrt{n}\sigma^3} = \frac{0.4748 \cdot \frac{51}{8}}{100\sqrt{\frac{35}{12}}^3} \approx 0.006.$$

Vagyis a becslés hibája sok nagyságrenddel nagyobb (lehet), mint a becslés maga.

HF 3.2 Egy véletlen algoritmus egy eldöntendő kérdésre $p = 0.55$ valószínűséggel ad helyes választ. Lefuttatjuk az algoritmust $n = 1000$-szer egymástól függetlenül, és megnézzük, hogy melyik válasz jön ki többször. Adjunk nagy eltérés becslést annak valószínűségére, hogy a végeredmény rossz – vagyis hogy a hibás válasz jön ki többször, vagy az eredmény döntetlen,

a.) a Hoeffding egyenlőtlenség segítségével,

b.) a Cramér tétel segítségével.

c.) Es akkor mi jön ki, ha $n = 10000$?

(Tipp: Legyen $i = 1, 2, \ldots, n$-re $X_i = 1$, ha az i-edik válasz helyes, és $X_i = 0$, ha hibás.)

Segítség: a p paraméterű Bernoulli eloszlás Cramér féle rátafüggvénye

$$I(x) = -x \ln \frac{p}{x} - (1-x) \ln \frac{1-p}{1-x}.$$

Megoldás: Legyen $i = 1, 2, \ldots, n$-re $X_i = 1$, ha az i-edik válasz helyes, és $X_i = 0$, ha hibás. Legyen $S_n = X_1 + \cdots + X_n$ a helyes válaszok száma. Kérdés a

$$\mathbb{P}\left(S_n \leq \frac{n}{2}\right) = \mathbb{P}\left(\frac{S_n}{n} \leq \frac{1}{2}\right)$$

valószínűség. Az X_i-k függetlenek és Bernoulli eloszlásúak $p = 0.55$ paraméterrel, vagyis $m := \mathbf{E}X_i = p = 0.55$.

a.) Az X_i-k korlátosak $0 = a_i \leq X_i \leq b_i = 1$ korlátokkal. A Hoeffding egyenlőtlenséghez kelleni fog, hogy

$$\sum_{i=1}^{n} (b_i - a_i)^2 = n(1-0)^2 = n = 1000.$$

$\mathbf{E}S_n = nm = 550$, így $t = 50$ választással a Hoeffding egyenlőtlenség szerint

$$\mathbb{P}(S_n \leq 500) = \mathbb{P}(S_n \leq \mathbf{E}S_n - t) \leq \exp\left\{-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\} = e^{-5} \approx 0.0067.$$

b.) A kérdés $\mathbb{P}\left(\frac{S_n}{n} \leq \frac{1}{2}\right) = \mathbb{P}\left(\frac{S_n}{n} \in (a, b]\right)$, ahol $a = -\infty$ és $b = \frac{1}{2}$. mivel $b = \frac{1}{2} < 0.55 = m$, a Cramér tétel szerint

$$\mathbb{P}\left(\frac{S_n}{n} \leq \frac{1}{2}\right) = \mathbb{P}\left(\frac{S_n}{n} \in (a.b]\right) \lessapprox e^{-nI(b)} = e^{-1000 I(\frac{1}{2})},$$

ahol I a $p = 0.55$ paraméterű Bernoulli eloszlás rátafüggvénye, vagyis $q = 1 - p$ jelöléssel

$$I\left(\frac{1}{2}\right) = -\frac{1}{2}\ln\frac{p}{\frac{1}{2}} - \left(1 - \frac{1}{2}\right)\ln\frac{q}{1-\frac{1}{2}} = -\frac{1}{2}\ln(4pq).$$

esetünkben $I(\frac{1}{2}) = -\frac{1}{2}\ln(0.99) \approx 0.005025$, így

$$\mathbb{P}(S_n \le 500) \lessapprox e^{-1000\cdot 0.005025} = e^{-5.025} \approx 0.0066.$$

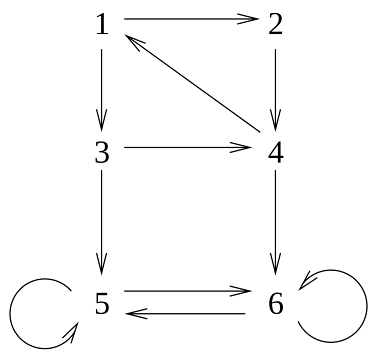
c.) $n = 10000$ esetén a Hoeffding egyenlőtlenségből

$$\mathbb{P}(S_{10000} \le 5000) \le e^{-50} \approx 1.93 \cdot 10^{-22}$$

jön ki, a Cramér tételből pedig

$$\mathbb{P}(S_{10000} \le 5000) \lessapprox e^{-50.25} \approx 1.50 \cdot 10^{-22}.$$

4.HF: (Beadási határidő: 2016.12.09.)



1. ábra. Markov lánc gráf-reprezentációja (valószínűségek nélkül)

HF 4.1 Az 1. ábrán látható gráf egy diszkrét idejű, időben homogén Markov lánc pozitív valószínűségű egylépéses átmeneteit mutatja. Osztályozzuk az állapotokat aszerint, hogy melyik melyikkel érintkezik! Minden osztályról állapítsuk meg, hogy

- zárt-e vagy nyílt,
- lényeges-e vagy lényegtelen,
- visszatérő-e vagy átmeneti,
- mennyi a periódusa.

Megoldás:

- $\{1, 2, 3, 4\}$ nyílt, lényegtelen, átmeneti, periódusa 3.
- $\{5, 6\}$ zárt, lényeges, visszatérő, periódusa 1 (vagyis aperiodikus).

HF 4.2 Egy fagyisnál a sorban álló gyerekek száma 0 és 4 között változhat (beleértve az éppen kiszolgálás alatt állót is): ha már 4-en vannak, és egy újabb gyerek be akarna állni, az apukája elrángatja. A fagyis bácsi nagyon igyekszik, de mindig csak $\frac{3}{4}$ valószínűséggel sikerül egy gyereket kiszolgálnia azelőtt, hogy egy újabb érkezne – az előzményektől függetlenül. Kivétel ez alól, ha 4-en vannak, mert akkor persze biztosan sikerül (új gyerek nem tud jönni), illetve ha a sor üres, mert akkor nincs is kit kiszolgálni.

Tekintsük a sorban állók számát *diszkrét időben*: a fagyis bácsi csettint egyet, valahányszor egy gyerek *érkezik vagy elmegy*, vagyis valahányszor a sor hossza változik. A sor hossza mindig pontosan 1-gyel változik (egyszerre csak 1 gyerek tud érkezni és elmenni is), és a fentiek szerint $\frac{3}{4}$ valószínűséggel csökken, a maradék $\frac{1}{4}$ valószínűséggel pedig nő, az előzményektől függetlenül (kivéve ha 4 vagy 0).

Legyen X_n a sor hossza az n-edik csettintés után (vagyis az n-edik sorhossz-változás után).

a.) Kezdetben a sor üres. Mennyi a valószínűsége, hogy 4 lépés után ismét üres?

b.) Kezdetben a sor üres. Mennyi a valószínűsége, hogy 5 lépés után ismét üres?

c.) Adjuk meg az X_n Markov lánc állapotterét és rajzoljuk fel a gráf-reprezentációját!

d.) Adjuk meg az X_n Markov lánc átmenetmátrixát!

e.) Keressük meg a Markov lánc stacionárius eloszlásait!

f.) Kezdetben a sor üres. Körülbelül mekkora a valószínűsége, hogy 100 lépés után ismét üres? *(Vigyázat: a feladat cseles, és az erre való tétel csak óvatosan alkalmazható. Egy hibásan alkalmazott tételnél jobb, ha precíz indoklás nélkül megsejtjük a helyes eredményt.)*

g.) **Bónusz kérdés:** Mi a válasz az előző kérdésre, ha a sor hosszára nincs felső korlát?

Megoldás: Kezdjük a c.) és d.) kérdéssel, utána jön az a.) és b.).

c.) Az állapottér $S = \{0, 1, 2, 3, 4\}$, a gráf-reprezentáció

$$\begin{tikzcd} 0 \arrow[r, shift left, "1"] & 1 \arrow[l, shift left, "3/4"] \arrow[r, shift left, "1/4"] & 2 \arrow[l, shift left, "3/4"] \arrow[r, shift left, "1/4"] & 3 \arrow[l, shift left, "3/4"] \arrow[r, shift left, "1/4"] & 4 \arrow[l, shift left, "1"] \end{tikzcd}$$

d.) Az átmenetmátrix

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

a.) A 0 állapotból a 0-ba 4 lépésben visszatérni csak kétféleképpen lehet: $0 \to 1 \to 0 \to 1 \to 0$ vagy $0 \to 1 \to 2 \to 1 \to 0$. Ezek (feltételes) valószínűsége (feltéve, hogy 0-ból indulunk $P_{01} \cdot P_{10} \cdot P_{01} \cdot P_{10} = 1 \cdot \frac{3}{4} \cdot 1 \cdot \frac{3}{4} = \frac{9}{16}$ illetve $P_{01} \cdot P_{12} \cdot P_{21} \cdot P_{10} = 1 \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{64}$, így

$$\mathbb{P}(X_4 = 0 | X_0 = 0) = \frac{9}{16} + \frac{9}{64} = \frac{45}{64} \approx 0.703.$$

b.) A 0 állapotból a 0-ba 5 lépésben visszatérni sehogy sem lehet, azért

$$\mathbb{P}(X_5 = 0 | X_0 = 0) = 0.$$

e.) A $(P^T - I)\pi^T$ lineáris egyenletrendszert kell megoldani, ahol I az egységmátrix, $\pi = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix}$ a keresett stacionárius eloszlás (sorvektor):

$$\begin{pmatrix} -1 & 3/4 & 0 & 0 & 0 \\ 1 & -1 & 3/4 & 0 & 0 \\ 0 & 1/4 & -1 & 3/4 & 0 \\ 0 & 0 & 1/4 & -1 & 1 \\ 0 & 0 & 0 & 1/4 & -1 \end{pmatrix} \begin{pmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Az egyenletrendszer a szokásos bővített mátrix jelöléssel

$$\left(\begin{array}{ccccc|c} -1 & 3/4 & 0 & 0 & 0 & 0 \\ 1 & -1 & 3/4 & 0 & 0 & 0 \\ 0 & 1/4 & -1 & 3/4 & 0 & 0 \\ 0 & 0 & 1/4 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1/4 & -1 & 0 \end{array}\right).$$

Eliminációval nagyon könnyen kijön, hogy

$$\left(\begin{array}{ccccc|c} -1 & 3/4 & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & 0 & 0 \\ 0 & 0 & 0 & -1 & 4 & 0 \end{array}\right).$$

Az utolsó egyenlet kiesett, ahogy kell. Az jött ki, hogy

$$\begin{aligned} \pi_0 &= \frac{3}{4}\pi_1 \\ \pi_1 &= 3\pi_2 \\ \pi_2 &= 3\pi_3 \\ \pi_3 &= 3\pi_4. \end{aligned} \tag{1}$$

(Megjegyzés: ezt erre a nagyon speciális gráfra a a gráf-reprezentációból is könnyen ki lehet találni.)

$\tilde{\pi}_4 := 1$ választással kijön egy megoldás: $\tilde{\pi} = \begin{pmatrix} 27 & 36 & 12 & 4 & 1 \end{pmatrix}$. Az összes többi megoldás ennek konstans-szorosa, tehát az egyetlen stacionárius eloszlás ennek lenormáltja (hogy a sorösszeg 1 legyen):

$$\pi = \begin{pmatrix} \frac{27}{80} & \frac{36}{80} & \frac{12}{80} & \frac{4}{80} & \frac{1}{80} \end{pmatrix}.$$

Azt, hogy pontosan egy stacionárius eloszlás van, persze előre tudtuk abból, hogy a Markov lánc véges állapotterű és irreducibilis.

f.) A Markov lánc véges állapotterű és irreducibilis. Ha aperiodikus is *lenne*, akkor a Markov láncok alaptétele szerint $\mathbb{P}(X_{100} = 0) \approx \pi_0 = \frac{27}{80} = 0.3375$ *lenne*, a kezdeti állapottól függetlenül. Igen ám, de ez a Markov lánc *periodikus*, éspedig a periódusa 2. Mivel a feltevés szerint $X_0 = 0$, $n = 100$ lépésben pont lehetséges eljutni a 0-ba – a keresett val.ség nem nulla.

Intuitív érvelés 1: 0-ból indulva a 0-ban csak minden páros lépésben járhatunk. Mégis, az ergodtétel szerint hosszú távon az időnek $\pi_0 = 33.75\%$-át töltjük ott. Ezért a *páros* időpontok $2 \cdot 33.75\% = 67.5\%$-át töltjük 0-ban, ezért

$$\mathbb{P}(X_{100} = 0 | X_0 = 0) \approx 2\pi_0 = 0.675 = 67.5\%.$$

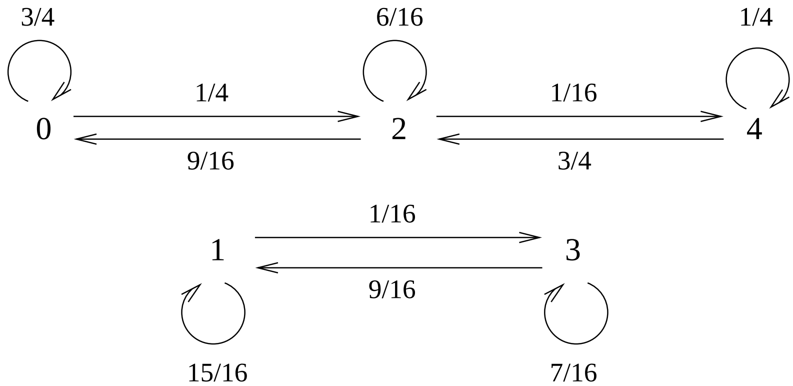
Intuitív érvelés 2: 0-ból indulva a $n = 100$ ugrás után csak a páros állapotokban lehetünk. Ezeken belül viszont – mivel $n = 100$ sok idő – jó közelítéssel a stacionárius eloszlás adja meg a tartózkodási valószínűségeket: persze **abban az értelemben**, hogy páros i-re az i-ben való tartózkodás valószínűsége π_i-vel arányos. Vagyis mivel a páros i-khez tartozó π_i-k összege nem 1, újra kell őket normálni 1-re. Ennek megfelelően

$$\mathbb{P}(X_{100} = 0 | X_0 = 0) \approx \frac{\pi_0}{\pi_0 + \pi_2 + \pi_4} = \frac{\frac{27}{80}}{\frac{27}{80} + \frac{12}{80} + \frac{1}{80}} = \frac{27}{40} = 0.675 = 67.5\%.$$

Precíz érvelés: A kétlépéses átmenetmátrix

$$P^2 = \begin{pmatrix} 3/4 & 0 & 1/4 & 0 & 0 \\ 0 & 15/16 & 0 & 1/16 & 0 \\ 9/16 & 0 & 6/16 & 0 & 1/16 \\ 0 & 9/16 & 0 & 7/16 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix}.$$

Nézzünk a Markov láncra csak minden páros időpontban. Így a $Y_k := X_{2k}$ Markov lánv átmenetmátrixa a fenti P^2, gráf-reprezentációja



Látható, hogy az Y_k Markov lánc *reducibilis*, kommunikáló osztályai a $\{0,2,4\}$ és az $\{1,3\}$, mindkettő zárt. Ezért az $\tilde{S} := \{0,2,4\}$ rész-állapottér önálló életet él, és az Y_k Markov láncot nézhetjük csak ezen. Így az álapottér már irreducibilis, az átmenetmátrix pedig a fenti P^2 páros indexű elemeiből áll:

$$\tilde{P} = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 9/16 & 6/16 & 1/16 \\ 0 & 3/4 & 1/4 \end{pmatrix}.$$

Vagyis kaptunk egy véges állapotterű, irreducibilis és immár *aperiodikus* Markov láncot. Ennek a szokásos módon kiszámohatjuk a stacionárius eloszlását, és az jön ki, hogy

$$\bar{\pi} = \begin{pmatrix} \bar{\pi}_0 & \bar{\pi}_2 & \bar{\pi}_4 \end{pmatrix} = \begin{pmatrix} \frac{27}{40} & \frac{12}{40} & \frac{1}{40} \end{pmatrix}.$$

Így a Markov láncok alaptétele szerint

$$\mathbb{P}(X_{100} = 0 | X_0 = 0) = \mathbb{P}(Y_{50} = 0 | Y_0 = 0) \approx \bar{\pi}_0 = \frac{27}{40} = 0.765.$$

g.) Ha a sorhosszra nincs korlát, a (1) egyenletrendszer megfelelőjére azt kapjuk, hogy

$$\begin{aligned} \pi_0 &= \frac{3}{4}\pi_1 \\ \pi_1 &= 3\pi_2 \\ \pi_2 &= 3\pi_3 \\ \pi_3 &= 3\pi_4 \\ \pi_4 &= 3\pi_5 \\ &\vdots \\ \pi_i &= 3\pi_{i+1} \\ &\vdots \end{aligned}$$

Ebből $\pi_0 := c$ választással kapható egy megoldás:

$$\tilde{\pi} = \begin{pmatrix} c & \frac{4}{3}c & \frac{4}{9}c & \frac{4}{27}c & \frac{4}{81}c & \dots & \frac{4}{3^i}c & \dots \end{pmatrix}$$

ezek összege $c(1 + \sum_{i=1}^{\infty} \frac{4}{3^i}) = c(1 + \frac{4/3}{1-1/3}) = 3c$, vagyis pontosan akkor 1, ha $c = \frac{1}{3}$. Vagyis $c := \frac{1}{3}$ választással kapunk egy stacionárius eloszlást:

$$\pi = \begin{pmatrix} \frac{1}{3} & \frac{4}{9} & \frac{4}{27} & \frac{4}{81} & \dots & \frac{4}{3^{i+1}} & \dots \end{pmatrix}.$$

Avagy:

$$\pi_i = \begin{cases} \frac{1}{3}, & \text{ha } i = 0 \\ \frac{4}{3^{i+1}}, & \text{ha } i \geq 1 \end{cases}.$$

Ellenőrizhető, hogy ez tényleg stacionárius eloszlás.

Nem tanultuk, de ebben a speciális végtelen állapotterű esetben is érvényes a Markov láncok alaptétele és az ergodtétel, mint a véges állapottér esetén (csak a konvergencia lassabb). Így az f.) pont-beli érvelésből az jön ki, hogy

$$\mathbb{P}(X_{100} = 0 | X_0 = 0) \approx 2\pi_0 = \frac{2}{3}.$$