Felsőbb Matematika Villamosmérnököknek – Sztochasztika
házi feladatok, 2019 ősz

Minden héten összesen 2 pontot érnek a kitűzött feladatok.

1.HF: (Beadási határidő: 2019.10.08.)

HF 1.1 Pistike egyszer dob egy szabályos dobókockával. Ezután Móricka addig dobál egy szabályos dobókockával, amíg nem sikerül legalább akkora számot dobnia, mint Pistikének. Legyen X Móricka dobásainak a száma.

a.) Mennyi a $\mathbb{P}(X = 1)$ valószínűség?

b.) Mennyi a $\mathbb{P}(X = 2)$ valószínűség?

c.) Mennyi az $\mathbb{E}X$ várható érték?

Megoldás:

a.) Ez még könnyű. $X = 1$ akkor és csak akkor, ha Móricka első dobása legalább akkora, mint Pistikéé. Ezen két dobás eredménye összesen $6 \cdot 6 = 36$ féle lehet. Ebből jó az a 6, amikor ugyanannyit dobnak, meg a maradék 30-nak az a fele, amikor Móricka dobása nagyobb. Vagyis

$$\mathbb{P}(X = 1) = \frac{6 + 15}{36} = \frac{21}{36} = \frac{7}{12} \approx 0.58 = 58\%.$$

b.) Legyen $l = 1, 2, \ldots, 6$-ra A_l az az esemény, hogy Pistike dobása l. Persze $(A_1, \ldots, A_6)$ teljes eseményrendszer és $\mathbb{P}(A_l) = \frac{1}{6}$. Azon feltétel mellett, hogy A_l bekövetkezik, Móricka dobásainak száma geometriai eloszlású $p_l = \frac{7-l}{6}$ paraméterrel, hiszen Móricka kockájának 6 oldalából $7 - l$ nyerő. Vagyis

$$\mathbb{P}(X = k \mid A_l) = (1 - p_l)^{k-1} p_l \quad k = 1, 2, \ldots \text{ -re.}$$

Konkrétan $\mathbb{P}(X = 2 \mid A_l) = (1 - p_l)p_l = \frac{l-1}{6}\frac{7-l}{6} = \frac{(l-1)(7-l)}{36}$. Így a teljes valószínűség tétel szerint

$$\mathbb{P}(X = 2) = \sum_{l=1}^{6} \mathbb{P}(A_l)\mathbb{P}(X = 2 \mid A_l) = \sum_{l=1}^{6} \frac{1}{6} \frac{(l-1)(7-l)}{36}$$
$$= \frac{0 \cdot 6 + 1 \cdot 5 + 2 \cdot 4 + 3 \cdot 3 + 4 \cdot 2 + 5 \cdot 1}{216} = \frac{35}{216} \approx 0.16 = 16\%.$$

c.) Az előző pont jelöléseivel: A_l feltétel mellett X geometriai eloszlású $p_l = \frac{7-l}{6}$ paraméterrel, vagyis $\mathbb{E}(X \mid A_l) = \frac{1}{p_l} = \frac{6}{7-l}$. Így a teljes várható érték tétel szerint

$$\mathbb{E}X = \sum_{l=1}^{6} \mathbb{P}(A_l)\mathbb{E}(X \mid A_l) = \sum_{l=1}^{6} \frac{1}{6} \frac{6}{7-l} = \sum_{l=1}^{6} \frac{1}{7-l} =$$
$$= \frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} = 2.45$$

HF 1.2 A radioaktív ^{14}C atommag élettartama (vagyis a létrejöttétől a bomlásáig eltelt idő) exponenciális eloszlású. A felezési idő 5730 év, ami azt jelenti, hogy egy sok atommagból álló mintának ennyi idő alatt bomlik el a fele.

a.) Mennyi az élettartam eloszlásának λ paramétere (rátája)

i.) ha az időt években mérjük?

ii.) ha az időt másodpercben mérjük?

b.) Veszünk egyetlenegy ^{14}C magot. Mennyi a valószínűsége, hogy egy másodpercen belül elbomlik? És hogy két másodpercen belül? És hogy 3 másodpercen belül? *(Figyelem: nem csak képleteket kérek, hanem konkrét számokat!)*

c.) Veszünk egy ezermilliárd (vagyis 10^{12}) magból álló mintát, és X-szel jelöljük a 3 másodperc alatt *bekövetkező* bomlások számát. Mennyi a $\mathbb{P}(X = 12)$ valószínűség? *(Figyelem: nem csak képletet kérek, hanem egy konkrét számot!)*

d.) A 10^{12} magból álló minta bomlásait egy detektorral figyeljük, ami csak a bomlások egy részét észleli: minden bomlást a többitől függetlenül $\frac{1}{4}$ valószínűséggel. Legyen Y a 3 másodperc alatt *észlelt* bomlások száma. Mennyi a $\mathbb{P}(Y = 3)$ valószínűség? *(Figyelem: nem csak képletet kérek, hanem egy konkrét számot!)*

e.) Legyen T a detektor által észlelt első bomlás időpontja (másodpercben). Mi X eloszlása?

Megoldás:

a.) Jelöljük ξ-vel az (egyetlen) atommag élettartamát, F-fel ennek eloszlásfüggvényét, $T_{1/2}$-vel a felezési időt, vagyis $T_{1/2} = 5730\text{év} \approx 1.8 \cdot 10^{11}s$. Az exponenciális eloszlás definíciója szerint $x \geq 0$-ra

$$F(x) = 1 - e^{-\lambda x}$$

a felezési idő definíciója szerint pedig

$$\frac{1}{2} = \mathbb{P}(\xi < T_{1/2}) = F(T_{1/2}) = 1 - e^{-\lambda T_{1/2}},$$

amiből

$$\lambda = \frac{\ln \frac{1}{2}}{-T_{1/2}} = \frac{\ln 2}{T_{\frac{1}{2}}} \approx \frac{0.693}{5730\text{év}} \approx \frac{0.693}{1.8 \cdot 10^{11}s}$$

i.) $\lambda = \frac{0.693}{5730\text{év}} \approx 1.21 \cdot 10^{-4} \frac{1}{\text{év}}$, vagyis ha az időt években mérjük, akkor $\lambda \approx 1.21 \cdot 10^{-4}$.

ii.) $\lambda \approx \frac{0.693}{1.8 \cdot 10^{11}s} \approx 3.83 \cdot 10^{-12} \frac{1}{s}$, vagyis ha az időt másodpercben mérjük, akkor $\lambda \approx 3.83 \cdot 10^{-12}$.

b.) Mérjük az időt másodpercben!

i.) 1 másodpercen belül: Az eltelt idő $t = 1$.

$$\mathbb{P}(\xi < t) = F(t) = 1 - e^{-\lambda t} \approx 3.83 \cdot 10^{-12}$$

Hát persze: Mivel λt nagyon kicsi, $e^{-\lambda t}$-t nagyon jól közelíti az elsőfokú Taylor polinomja: $e^{-\lambda t} \approx 1 - \lambda t$, vagyis $F(t) \approx \lambda t$. Éppen ezért hívják λ-t az exponenciális eloszlás *rátájának.*

(Alternatív érvelés: mivel λt kicsi, a ξ élettartam $f(x) = \lambda e^{-\lambda x}$ (ha $x > 0$) sűrűségfüggvénye jó közelítéssel konstans a $[0, t]$ intervallumon, így $\mathbb{P}(\xi < t) = \int_0^t f(x)\,\mathrm{d}x \approx f(0)(t - 0) = \lambda t$.)

ii.) Hasonlóan 2 és 3 másodpercre:

$$\mathbb{P}(\xi < 2) \approx 2\lambda \approx 7.66 \cdot 10^{-12}$$
$$\mathbb{P}(\xi < 3) \approx 3\lambda \approx 1.15 \cdot 10^{-11}$$

c.) Szigorúan véve X binomiális eloszlású $n = 10^{12}$ és $p = \mathbb{P}(\xi < 3) \approx 3\lambda \approx 1.15 \cdot 10^{-11}$ paraméterekkel, így

$$\begin{aligned}\mathbb{P}(X = 12) &= \binom{n}{12} p^{12} (1 - p)^{n-12} \\ &\approx \binom{1000000000000}{12} \left(1.15 \cdot 10^{-11}\right)^{12} \left(1 - 1.15 \cdot 10^{-11}\right)^{999999999988}.\end{aligned}$$

Ezt az én számológépem nem is bírja kiszámolni, ezért Poisson eloszlással közelítünk: mivel n nagy és p kicsi, jó közelítéssel X Poisson eloszlású $np \approx 11.5$ paraméterrel. Vagyis

$$\mathbb{P}(X = 12) \approx e^{-11.5}\frac{11.5^{12}}{12!} \approx 0.113 = 11.3\%$$

d.) Y a fenti, Poisson eloszlással modellezhető X ritkítása, így maga is (jó közelítéssel) Poisson eloszlású $\frac{np}{4} \approx 2.875$ paraméterrel. Így

$$\mathbb{P}(Y = 3) \approx e^{-2.875}\frac{2.875^{3}}{3!} \approx 0.223 = 22.3\%$$

e.) A kérdés persze sajtóhibás, bocs. Persze a T eloszlása érdekel. Mivel az anyag másodperces időskálán nézve nagyon lassan fogy el, rövid ideig nyugodtan úgy tekinthetjük, hogy az észlelések Poisson folyamat szerint történnek. (Hosszú időre ez persze nem igaz, 5730 év alatt például a felére csökken az aktivitás (vagyis a ráta).) Mivel a bomlások rátája $n\lambda$, az észlelések rátája $\frac{n\lambda}{4} \approx 0.96$. Így az első észlelésig eltelt T idő (nagyon jó közelítéssel) exponenciális eloszlású 0.96 paraméterrel.

2.HF: (Beadási határidő: 2019.10.29.)

HF 2.1 Feldobunk egy 10 és egy 20 forintos érmét egyszerre, és ezt addig ismételgetjük, amíg a 20-ason fej nem jön ki. Jelölje X, hogy ez alatt *a 10-essel* hány fejet dobtunk. Számoljuk ki X várható értékét és generátorfüggvényét!

Megoldás: Jelölje Y, hogy hányszor kell dobnunk a 20 forintos érmével az első fejhez. Ekkor Y geometriai eloszlású 0.5 paraméterrel.

$$\mathbb{E}Y = 2, \quad g_Y(z) = \frac{\frac{1}{2}z}{1 - \frac{1}{2}z} \tag{1}$$

Legyen X_i egy $\frac{1}{2}$ paraméterű Bernoulli változó minden $i \in \mathbb{N}$ indexre, ami az 1 értéket veszi fel, ha az i-edik dobásunk a 10 forintos érmével fej.

$$\mathbb{E}X_1 = \frac{1}{2}, \quad g_{X_1}(z) = \frac{1}{2}z + \frac{1}{2}$$

Ezekkel a jelölésekkel X a következő véletlen tagszámó összegként írható fel:

$$X = \sum_{i=1}^{Y} X_i$$

Mivel a 2 pénzérme feldobásának kimenetelei függetlenek, az órán tanultak alapján X várható értéke és generátor függvénye a következő:

$$\mathbb{E}X = \mathbb{E}X_1 \cdot \mathbb{E}X_i = 1, \quad g_X(z) = g_Y(g_{X_1}(z)) = \frac{1}{\frac{4}{z+1} - 1} = \frac{1+z}{3-z}.$$

HF 2.2 Egy boltba a vevők egyesével érkeznek, percenként pontosan egy, és beállnak az egyetlen sorba. Amikor sorra kerülnek, a kiszolgálásuk az előzményektől független véletlen ideig tart: $\frac{1}{3}$ valószínűséggel 0 percig (a vevő csak szétnéz és elmegy, kár is volt sorba állnia), p valószínűséggel 1 percig, és a maradék $\frac{2}{3} - p$ valószínűséggel 2 percig.

A boltba 08:00-kor érkezik az első vevő: nevezzük őt egymagát *nulladik generációnak*. Azokat a vevőket, akik az ő kiszolgálása alatt érkeznek, nevezzük *első generációnak* (lehet, hogy ilyen egy sincs). Azokat a vevőket, akik az első generáció kiszolgálása alatt érkeznek, nevezzük *második generációnak*, stb: Azok a vevők, akik an n-edik első generáció tagjainak kiszolgálása alatt érkeznek, alkotják az $n+1$-edik generációt. Jelölje Z_n az n-edik generáció tagjainak számát.

Az alábbi kérdéseket válaszoljuk meg

I.) ha $p = \frac{1}{2}$,

II.) ha $p = \frac{1}{6}$.

a.) Mennyi $\mathbb{E}Z_0$?

b.) Mennyi $\mathbb{E}Z_{10}$?

c.) Mi Z_0 generátorfüggvénye?

d.) Mi Z_2 generátorfüggvénye?

e.) Mennyi annak a valószínűsége, hogy $Z_3 = 0$?

f.) Mennyi annak a valószínűsége, hogy a sor előbb-utóbb elfogy?

g.) Legyen N a sor első elfogyásáig kiszolgált vevők teljes száma (a legelső vevőt is beleértve). Mennyi N várható értéke?

h.) Mi N generátorfüggvénye?

Megoldás: I.) $p = \frac{1}{2}$, így ha X jelöli az utódeloszlást

$$g_X(z) = \frac{1}{3} + \frac{1}{2}z + \frac{1}{6}z^2 \quad , \quad \mathbb{E}X = \frac{1}{3}\cdot 0 + \frac{1}{2}\cdot 1 + \frac{1}{6}\cdot 2 = \frac{5}{6}.$$

Az első vevő alkotja a nulladik generációt, ezért $g_{Z_0}(z) = z$ valamint $\mathbb{E}Z_0 = 1$. Galton-Watson folyamatokról ismert, hogy az n-edik generáció generátora az utódeloszlás generátorának n-szeres kompozíciója, azaz $g_{Z_n}(z) = g_X \circ .. \circ g_X(z)$, így $\mathbb{E}Z_{10} = (\mathbb{E}X)^{10} = \left(\frac{5}{6}\right)^{10}$.

A generátor függvény segítségével könnyen számolhatóak a kihalási valószínűségek is.

$$\mathbb{P}\left(Z_3 = 0\right) = g_X(g_X(g_X(0))) = g_X(g_X(\frac{1}{3})) = g_X(\frac{14}{27}) = \frac{1394}{2187}$$

Mivel $\mathbb{E}X < 1$, a folyamat biztosan kihal. Ekkor N egy jól definiált valószínűségi változó. Az órán tanultak alapján $\mathbb{E}N = \frac{1}{1-\mathbb{E}X} = 6$.

N generátor függvénye egy olyan $y(z)$ függvény, melyre

$$y(z) = z \cdot g_X(y(z)) = z\left(\frac{1}{3} + \frac{1}{2}y(z) + \frac{1}{6}(y(z))^2\right)$$

teljesül minden z esetén. Ez egy másodfokú egyenlet megoldására vezet, amelyben z paraméterként szerepel.

$$y_{1,2}(z) = \frac{1 - \frac{1}{2}z \pm \sqrt{\frac{z^2}{4} - z + 1 - \frac{4z^2}{18}}}{\frac{1}{3}z}$$

Minden generátor függvénynek fixpontja van 1-ben, így azt a megoldást kell választanunk, ami ezt teljesíti.

$$y_1(1) = \frac{\frac{1}{2} + \frac{1}{6}}{\frac{1}{3}} = 2, \quad y_2(1) = \frac{\frac{1}{2} - \frac{1}{6}}{\frac{1}{3}} = 1$$

Tehát N generátor függvéne $y_2(z)$.

II.) $p = \frac{1}{6}$, így ha X jelöli az utódeloszlást

$$g_X(z) = \frac{1}{3} + \frac{1}{6}z + \frac{1}{2}z^2, \mathbb{E}X = \frac{1}{3} \cdot 0 + \frac{1}{6} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{7}{6}.$$

Az I.) rész megoldásánál használt formulákat és érvelést alkalmazva

$$\mathbb{E}Z_0 = 1, \quad \mathbb{E}Z_{10} = \left(\frac{7}{6}\right)^{10}, \mathbb{P}\left(Z_3 = 0\right) = \frac{41}{81}$$

.

A kihalás valószínűsége megegyezik $g_X(z)$ legkisebb fixpontjával:

$$z = \frac{1}{3} + \frac{1}{6}z + \frac{1}{2}z^2$$
$$z_{1,2} = \frac{5 \pm \sqrt{25-24}}{6} = \frac{5 \pm 1}{6}$$

A kisebb gyök $\frac{2}{3}$, így ez lesz a megoldás. A folyamat csupán 1-nél kisebb valószínűséggel hal ki, emiatt N várható értéke ∞ és a generátora nem definiált.

3.HF: (Beadási határidő: 2019.11.12.)

HF 3.1 Pistike 6999-szer dob egy szabályos dobókockával.

a.) Közelítsük a centrális határeloszlás tétel segítségével annak a valószínűségét, hogy 1000-nél kevesebb hatost sikerül dobnia!

b.) Legfeljebb mennyi lehet a fenti közelítés hibája a Berry-Esseen tétel szerint?

Megoldás: Legyen $n = 6999$ és $k = 1, 2, \ldots, n$-re $X_k = 1$, ha az k-edik dobás 6-os, és $X_k = 0$, ha nem. $S_n = X_1 + \cdots + X_n$ a dobott hatosok száma. Feladat a $\mathbb{P}(S_n < 1000)$ valószínűség közelítése. Az X_k-k függetlenek és azonos *Bernoulli* $\left(\frac{1}{6}\right)$ eloszlásúak.

a.) $m := \mathbb{E}X_k = \frac{1}{6}$, $\sigma^2 := Var(X_k) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$. Így S_n váható értéke $\mathbb{E}S_n = nm = 1166.5$, szórása pedig $\mathbf{D}S_n = \sqrt{n}\sigma \approx 31.18$. Vagyis ha a várható értéktől 166.5-del el akarunk maradni, az a szórás körülbelül 5.34-szeresét jelenti. A CHT szerint

$$\mathbb{P}(S_n < 1000) = \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} < \frac{1000 - 1166.5}{\sqrt{6999}\sqrt{\frac{5}{36}}}\right) \approx$$
$$\approx \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} < -5.34\right) \approx \Phi(-5.34) = 1 - \Phi(5.34),$$

ahol Φ a standard normális eloszlásfüggvény. Ezt kevés táblázatból lehet 5.34-ben kiolvasni, de az mindegyikből látszik, hogy $\Phi(5.34) > 0.9999$, vagyis $1 - \Phi(5.34) < 10^{-4}$. A táblázatkezelőm szerint viszont $\Phi(-5.34) \approx 4.65 \cdot 10^{-8}$.

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

$$\delta = \frac{5}{6}\left|0 - \frac{1}{6}\right|^3 + \frac{1}{6}\left|1 - \frac{1}{6}\right|^3 = \frac{5}{6}\left(\frac{1}{6}\right)^3 + \frac{1}{6}\left(\frac{5}{6}\right)^3 \approx 0.100.$$

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

$$\text{hiba} \leq \frac{C\delta}{\sqrt{n}\sigma^3} = \frac{0.4748 \cdot 0.1}{\sqrt{6999}\sqrt{\frac{5}{36}}^3} \approx 0.011.$$

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

HF 3.2 Mórickának 1000 szabályos dobókockája van. Mindegyikkel addig dobál, amíg ki nem jön rajta a 6-os. Adjunk nagy eltérés becslést annak a valószínűségére, hogy összesen legalább 7000 dobásra lesz szüksége!

(Segítség: a p paraméterű Bernoulli eloszlás Cramér féle rátafüggvénye

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

(ahol $q = 1-p$ és $0 < x < 1$). A p paraméterű (optimista) geometriai eloszlás Cramér féle rátafüggvénye

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

(ahol $q = 1-p$ és $x > 1$).)

1. a,b megoldás: Annak persze semmi jelentősége, hogy Móricka 1000 különböző kockét dobál. Mondhattuk volna azt is, hogy egyetlen kockét dobál addig, amíg ki nem jön 1000 darab 6-os. A kérdés tehát átfogalmazható a következőképpen: Ha Móricka 6999-szer dob a kockával, mennyi a valószínűsége, hogy ebből csak 1000-nél kevesebblesz 6-os?

Ezért legyen $n = 6999$ és $k = 1, 2, \ldots, n$-re legyen $X_k = 1$ ha a k-adik dobás 6-os, és legyen $X_k = 0$ ha nem.Így az X_k-k függetlenek és $p = \frac{1}{6}$ paraméterű Bernoulli eloszlásúak. $S_n := X_1 + \cdots + X_n$ a dobott 6-osok száma, a kérdés pedig $\mathbb{P}(S_n < 1000)$.

a.) A legkönnyebb megoldást a Hoeffding egyenlőtlenség adja. $\mathbb{E}S_n = np = \frac{6999}{6} = 1166.5$, így legyen $t = \mathbb{E}S_n - 1000 = \frac{999}{6} = 166.5$, $a_k = 0$ és $b_k = 1$ minden $k = 1, 2, \ldots, n$-re. A Hoeffding egyenlőtlenség szerint

$$\mathbb{P}(S_n < \mathbb{E}S_n - t) \leq \exp\left\{ -\frac{2t^2}{\sum_{k=1}^{n} (b_k - a_k)^2} \right\} = \exp\left\{ -\frac{2 \cdot \left(\frac{999}{6}\right)^2}{6999} \right\} \approx e^{-7.92} \approx 0.00036.$$

b.) Pontosabb becslést kaphatunk a Cramér tétel segítségével. Ez is könnyű – ha a segítség alapján tudjuk a rátafüggvényt. Legyen $m = \mathbb{E}X_k = p = \frac{1}{6}$, így a kérdés $\mathbb{P}(S_n < 1000) = \mathbb{P}\left(\frac{S_n}{n} < \frac{1000}{6999}\right) = \mathbb{P}\left(\frac{S_n}{n} \in (a, b)\right)$, ahol $a = -\infty$ és $b = \frac{1000}{6999}$. Mivel $b < m$, a Cramér tétel azt állítja, hogy

$$\mathbb{P}\left(\frac{S_n}{n} \in (a, b)\right) \lessapprox e^{-n \cdot I(b)} = e^{-6999 \cdot I\left(\frac{1000}{6999}\right)} \approx e^{-14.85} \approx 0.00000036.$$

(A Bernoulli eloszlás rátafüggvényét számoltuk ki a segítség alapján $p = \frac{1}{6}$, $x = \frac{1000}{6999}$-ben.)

2. megoldás: A kérdést közvetlenül is nézhetjük: legyen $n = 1000$ és $k = 1, 2, \ldots, n$-re jelentse Y_k azt, hogy hányszor kell dobni a k-adik kockán a hatosig. Így az Y_k-k függetlenek és (optimista) geometriai eloszlásúak $p = \frac{1}{6}$ parameterrel, $S_n := Y_1 + \cdots + Y_n$ pedig az 1000 darab 6-oshoz szükséges dobások teljes száma. A kérdés most $\mathbb{P}(S_n \geq 7000)$.

Vegyük észre, hogy az Y_k-k *nem korlátosak*, így a Hoeffding egyenlőtlenség nem alkalmazható. A Cramér tétel viszont most is működik. Alkalmazásához átírjuk a kérdést $\mathbb{P}(S_n \geq 7000) = \mathbb{P}(\frac{S_n}{n} \geq 7) = \mathbb{P}(\frac{S_n}{n} \in [a, b))$ alakba, ahol $a = 7$ és $b = \infty$. Legyen $m = \mathbb{E}Y_k = \frac{1}{p} = 6$. Mivel $a > m$, a Cramér tétel azt állítja, hogy

$$\mathbb{P}\left(\frac{S_n}{n} \in [a, b)\right) \lessapprox e^{-n \cdot I(a)} = e^{-1000 \cdot I(7)} \approx e^{-14.87} \approx 0.0.00000035.$$

(A geometriai eloszlás rátafüggvényét számoltuk ki a segítség alapján $p = \frac{1}{6}$, $x = 7$-ben.)

Megjegyzés: A kérdéses valószínűségnek a centrális határeloszlás tétellel történő becslésére tett minden kísérlet hibás. a CHT nem *nagy eltérés tétel, és egyáltalán nem alkalmas az átlag ilyen szélsőséges értékei ilyen kicsi valószínűségeinek becslésére - lásd az előző feladatot.*

4.HF: (Beadási határidő: 2019.12.03.)

HF 4.1 Egy diszkrét idejű, időben homogén X_n Markov lánc állapottere $S = \{1, 2, 3\}$. A Markov lánc az 1-es állapotból $50 - 50\%$ valószínűséggel ugrik a 2-es és 3-as állapotba. Ha a 2-es állapotban van, akkor 50% valószínűséggel ott is marad, 50% valószínűséggel pedig a 3-as állapotba ugrik. A 3-as állapotból mindig az 1-esbe ugrik. A Markov lánc X_0 kezdeti állapotát kockadobással sorsoljuk, egyenlő esélyt adva mindhárom állapotnak.

a.) Rajzoljuk fel a Markov lánc gráf-reprezentációját.

b.) Írjuk fel a Markov lánc átmenetmátrixát.

c.) Írjuk fel a Markov lánc kezdeti eloszlás vektorát.

d.) Mennyi a valószínűsége, hogy a folyamat kezdetén a 131223 állapot-sorozatot figyeljük meg (a 0-dik (kezdő) állapotot is beleértve)?

e.) Mennyi a $\mathbb{P}(X_4 = 1 \,|\, X_0 = 1)$ átmenetvalószínűség?

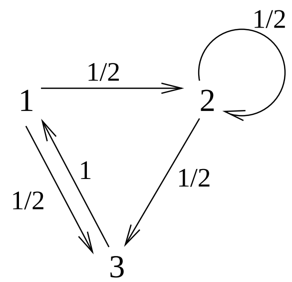
f.) Keressük meg a Markov lánc stacionárius eloszlásait.

g.) Közelítőleg mennyi a valószínűsége, hogy 100 lépés után a Markov lánc a 2-es állapotban lesz?

h.) Legyen az $f : S \to \mathbb{R}$ függvény olyan, hogy $f(1) = 0$, $f(2) = 1$ és $f(3) = 5$. Mennyi lesz az $f(X_n)$ sorozat (idő)átlaga hosszú távon?

Megoldás:

a.) A gráf-reprezentáció



b.) Az átmenetmártix $P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$

c.) A kezdeti eloszlás vektor $\pi(0) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, mert mindhárom állapotnak azonos esélyt adtunk.

d.)

$$\begin{aligned}\mathbb{P}(131223) &= \mathbb{P}(X_0 = 1, X_1 = 3, X_2 = 1, X_3 = 2, X_4 = 2, X_5 = 3) = \\ &= \pi_1(0) P_{13} P_{31} P_{12} P_{22} P_{23} = \frac{1}{3} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{48}.\end{aligned}$$

e.) Az 1-es állapotból az 1-esbe négy lépésben visszajutni csak kétféleképpen lehet: $1 \to 2 \to 2 \to 3 \to 1$, vagy $1 \to 3 \to 1 \to 3 \to 1$. Ezek feltételes val.sége (feltéve, hogy $X_0 = 1$) $P_{12} P_{22} P_{23} P_{31} = \frac{1}{8}$, illetve $P_{13} P_{31} P_{13} P_{31} = \frac{1}{4}$, így

$$\mathbb{P}(X_4 = 1 \,|\, X_0 = 1) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}.$$

f.) Mivel a Markov lánc irreducibilis és véges állapotterű, pontosan egy π stacionárius eloszlás (sorvektor) van, éspedig a

$$(P^T - I)\pi^T = 0$$

lineáris egyenletrendszer egyetlen olyan megoldása, ahol a sorösszeg 1. (Itt I az egységmátrix.) Jelen esetben

$$P^T - I = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix},$$

vagyis a lineáris egyenletrendszer a szokásos mátrix-reprezentációval

$$\left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & 0 \end{array}\right).$$

Ezt megoldjuk pl. Gauss eliminációval:

$$\left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \end{array}\right) \sim$$

$$\sim \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right),$$

amiből $\pi_1 = \pi_3$ és $\pi_1 = \pi_2$. Így pl. a $\pi_3 := 1$ önkényes választással megkapjuk a lineáris egyenletrendszer egy megoldását: $\tilde{\pi} = (1\,1\,1)$. Ez még nem az, amit keresünk, mert az elemek összege nem 1, hanem $1 + 1 + 1 = 3$, ezért a keresett megoldást úgy kapjuk, hogy ezt a $\tilde{\pi}$-t lenormáljuk, vagyis leosztjuk 3-mal:

$$\pi = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

g.) A Markov lánc irreducibilis, aperiodikus és véges állapotterű, $n = 100$ lépés pedig hosszú idő, ezért a Markov láncok alaptétele értelmében a kezdeti eloszlástól függetlenül

$$\mathbb{P}(X_{100} = 2) \approx \pi_2 = \frac{1}{3}.$$

h.) Az f függvényt célszerű oszlopvektor formájába írni:

$$f = \begin{pmatrix} 0 \\ 1 \\ 5 \end{pmatrix}.$$

A Markov lánc irreducibilis és véges állapotterű, ezért az ergodtétel értelmében az $f(X_n)$ sorozat (idő)átlaga hosszú távon 1 valószínűséggel a stacionárius eloszlás szerinti várható értékhez tart:

$$\lim_{n\to\infty} \frac{f(X_0) + f(X_1) + \cdots + f(X_{n-1})}{n} = \mathbb{E}_\pi f = \sum_{i\in S} \pi_i f(i) = \pi f =$$

$$= \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 5 \end{pmatrix} = \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 5 = 2.$$

HF 4.2 Egy fagyis bácsi előtt gyerekek állnak sorba. Ő minden sorra kerülő gyereket exponenciális eloszlású véletlen idő alatt szolgál ki, fél perc várható értékkel, az előzményektől és a sorban állók számától függetlenül. A gyerekek Poisson folyamat szerint érkeznek, percenként átlagosan 1, a múlttól és a sor hosszától függetlenül. Kivétel, ha a sorban már 3 gyerek áll, mert akkor több nem állhat be (az apukája elvonszolja). Jelölje $X(t)$ a sorban állók számát t idő elteltével. Az időt mérjük percben. Modellezzük a sor hosszát folytonos idejű Markov lánccal.

a.) Mi a Markov lánc állapottere?

b.) Rajzoljuk fel a Markov lánc gráf-reprezentációját – vagyis a lehetséges átmenetek irányított gráfját az egyes átmenetek rátáival.

c.) Írjuk fel a Markov lánc ráta-mátrixát.

d.) Írjuk fel a Markov lánc tartózkodási idő paraméter vektorát.

e.) Ha a sor hossza éppen 2, várhatóan mennyi idő múlva fog megváltozni?

f.) Ha a sor hossza éppen 2, mennyi a valószínűsége, hogy a következő állapot a 3 lesz?

g.) Írjuk fel a Markov lánc infinitezimális generátorát.

h.) Keressük meg a Markov lánc stacionárius eloszlásait.

i.) Ha a sor $t = 0$-kor üres, közelítőleg mennyi a valószínűsége, hogy 2 óra elteltével 3 lesz a hossza?

j.) Az idő hány százalékát tölti a fagyis bácsi tétlenül (mert üres a sor) hosszú távon?

Megoldás:

a.) A sorban állók száma 0,1,2 vagy 3 lehet, így az állapottér $S = \{0, 1, 2, 3\}$.

b.) Egyszerre csak egy gyerek mehet el, és csak egy érkezhet, így ugrani minden állapotból csak szomszédos állapotokba lehet. A felfelé ugrás rátája mindig a gyerekek érkezési folyamatának rátája, vagyis 1 (mert percenként átlag 1 gyerek érkezik). A lefelé ugrás rátája pedig a kiszolgálás rátája, vagyis 2, mert a bácsi percenként átlag 2 gyereket szolgál ki. *FONTOS, hogy a ráta, ami* 2*, nem egyenlő a kiszolgálási idő várható értékével, ami* $\frac{1}{2}$*, hanem annak a reciproka.* Így a gráf-reprezentáció

$$0 \underset{2}{\overset{1}{\rightleftarrows}} 1 \underset{2}{\overset{1}{\rightleftarrows}} 2 \underset{2}{\overset{1}{\rightleftarrows}} 3$$

c.) Ez alapján a ráta-mátrix

$$\underline{\underline{\lambda}} = \begin{pmatrix} * & 1 & 0 & 0 \\ 2 & * & 1 & 0 \\ 0 & 2 & * & 1 \\ 0 & 0 & 2 & * \end{pmatrix}.$$

d.) A tartózkodási idő paraméter vektor a ráta-mátrix sorösszegeiből áll, vagy más szóval az egyes állapotokból való elugrás rátáiból:

$$\underline{\lambda} = \begin{pmatrix} 1 & 3 & 3 & 2 \end{pmatrix}.$$

e.) A 2 állapotból való elugrás rátája $\lambda_2 = 3$ vagyis a tartózkodási idő várható értéke $\frac{1}{\lambda_2} = \frac{1}{3}$.

f.) Az egyes irányokba való ugrások valószínűségeit a beépített diszkrét idejű Markov lánc átmenetmátrixa adja meg: $Q_{ij} = \frac{\lambda_{ij}}{\lambda_1}$ ha $i \neq j$, és $Q_{ij} = 0$, ha $i = j$. Most a kérdés csak a 2-ből 3-ba ugrás val.ségére vonatkozott, vagyis a válasz

$Q_{23} = \frac{\lambda_{23}}{\lambda_2} = \frac{1}{3}$. De ha már itt tartunk, felírom az egész mátrixot:

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

g.) Az infinitezimális generátort úgy kapjuk a ráta-mátrixból, hogy a főátlóba beírjuk a tartózkodási idő paraméter vektor eleminek ellentettjét (vagyis annyit, hogy a sorösszegek nullák legyenek):

$$G = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 2 & -3 & 1 & 0 \\ 0 & 2 & -3 & 1 \\ 0 & 0 & 2 & -2 \end{pmatrix}.$$

h.) Mivel a Markov lánc irreducibilis és véges állapotterű, pontosan egy π stacionárius eloszlás (sorvektor) van, éspedig a $G^T \pi^T = 0$ lineáris egyenletrendszer egyetlen olyan megoldása, ahol a sorösszeg 1. Jelen esetben az egyenletrendszer a szokásos mátrix-reprezentációval

$$\left(\begin{array}{cccc|c} -1 & 2 & 0 & 0 & 0 \\ 1 & -3 & 2 & 0 & 0 \\ 0 & 1 & -3 & 2 & 0 \\ 0 & 0 & 1 & -2 & 0 \end{array}\right),$$

aminek a normált megoldása

$$\pi = \begin{pmatrix} \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \end{pmatrix}.$$

(Az egyenletrendszer megoldásának és a megoldás lenormálásának menetét illetően lásd az előző feladat megoldását, vagy bármelyik lineáris algebra könyvet.)

i.) A Markov lánc folytonos idejű, irreducibilis és véges állapotterű, $t = 120$ perc pedig hosszú idő, ezért a Markov láncok alaptétele értelmében a kezdeti eloszlástól függetlenül

$$\mathbb{P}(X(120) = 3) \approx \pi_3 = \frac{1}{15}.$$

j.) Legyen

$$f = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

függvény az állapotéren. Ez éppen a 0 állapot indikátora, így $f(X(t))$ időátlaga éppen azt mondja meg, hogy a sor az idő mekkora hányadában üres. A Markov lánc irreducibilis és véges állapotterű, ezért az ergodtétel értelmében az $f(X(t))$ függvény (idő)átlaga hosszú távon 1 valószínűséggel a stacionárius eloszlás szerinti várható értékhez tart:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X(t))\,\mathrm{d}t = \mathbb{E}_\pi f = \sum_{i\in S} \pi_i f(i) = \pi f =$$

$$= \begin{pmatrix} \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{8}{15}.$$

5.HF: (Beadási határidő: 2019.12.18.)

HF 5.1 Móricka és Pistike is büszke arra, hogy dombon lakik. Szeretnék eldönteni, hogy melyikük lakik magasabban. Ehhez egy GPS vevőről olvasták le a tengerszint feletti magasságot, minkettőjük toronyszobájának ablakában. A GPS vevő által jelzett magasság hibával terhelt: a mutatott érték normális eloszlású, aminek várható értéke a tényleges magasság, szórása viszont ismeretlen (de minden mérésnél ugyanaz). Móricka és Pistike mindkét magasságot többször is megmérte különböző időpontokban, így a mérések függetlennek tekinthetők. A következő adatokat mérték (méterben):

Móricka	265	259	266	274	268	268
Pistike	275	267	271	268	261	265

a.) Döntsünk 95%-os szinten arról a hipotézisről, hogy Pistike ablaka legalább olyan magasan van, mint Mórickáé!

b.) Döntsünk 95%-os szinten arról a hipotézisről, hogy Móricka ablaka legalább olyan magasan van, mint Pistikéé!

Megoldás: Legyen m_1 Móricka ablakának magassága, m_2 pedig Pistike ablakának magassága, méterben. A null-hipotézis az a.) kérdésben $m_1 \leq m_2$, míg a b.) kérdésben $m_1 \geq m_2$. A feladat szövege szerint $\mathcal{N}(m_1, \sigma)$ és $\mathcal{N}(m_2, \sigma)$ eloszlásokból vettünk mintát, ahol m_1, m_2 és σ is ismeretlen. Ezért kétmintás, egyoldali t-próbát végzünk:

* Kétmintásat, mert a nullhipotézis két ismeretlen mennyiség összehasonlításáról szól, és két minta tartozik hozzá.
* Egyoldalit, mert a nullhipotézis egyenlőtlenség (és nem pedig egyenlőség).
* t-próbát, mert σ ismeretlen.

A képletgyűjtemény szerint a teszt-statisztika

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1-1)\sigma_1^{*2} + (n_2-1)\sigma_2^{*2}}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

ezt kell összehasonlítani a t_ε küszöbértékkel, ami az $n_1 + n_2 - 2$ szabadsági fokú t eloszlás $1 - \varepsilon$ kvantilise.

a.) Az adatokból $\bar{x} \approx 266.67$ és $\bar{y} \approx 267.83$. Így nem is kell többet számolni: mivel $\bar{x} < \bar{y}$, biztosan $t < 0$, ami a null-hipotézist *megerősíti.* (Azt csak akkor lehetne okunk elvetni, ha $\bar{x} > \bar{y}$ lenne.) Ezért a nullhipotézist **elfogadjuk.**

b.) Ehhez számolni kell: $\bar{x} < \bar{y}$ amiből $t < 0$, és a kérdés az, hogy ez szignifikáns-e 95%-os szinten. Ebből az is látszik, hogy a döntéshez t-t nem t_ε-nal, hanem $-t_\varepsilon$-nal kell összehasonlítani.
Az adatokból: $n_1 = 6$, $\bar{x} \approx 266.67$, $\sigma_1^* \approx 4.885$, $n_2 = 6$, $\bar{y} \approx 267.83$, $\sigma_2^* \approx 4.834$, $\varepsilon = 0.05$. A korrigált tapasztalati szórásokat zsebszámológéppel számoltam, az $\varepsilon = 0.05$ annak felel meg, hogy 95%-os konfidenciaszinten kell döntenünk. A szabadsági fokok száma $df = n_1 + n_2 - 2 = 10$. Ebből

$$t \approx -0.416 \quad , \quad t_\varepsilon \approx 1.812$$

Döntés: $t \geq -t_\varepsilon$, ezért a nullhipotézist **elfogadjuk**.