Felsőbb Matematika Villamosmérnököknek - Sztochasztika
2. ZH pótlása – megoldások

2018 ősz, 2018.12.06 18:00

Munkaidő: 90 perc. A nulladik feladat 0 pontos, a többi mind 9 pontot ér.

1. Legyen $n = 300$, legyenek $X_1, X_2, \dots, X_n$ független, $[-1, 1]$-en egyenletes eloszlású valószínűségi változók, és legyen $S_n = X_1 + \dots + X_n$. Ha a $\mathbb{P}(S_n > 30)$ valószínűséget a centrális határeloszlás tétel segítségével becsüljük,

   a.) Mennyit kapunk?

   b.) Legfeljebb mennyit tévedhetünk a Berry-Esseen tétel szerint? (A tételbeli konstantst vehetjük $C = 0.4748$-nak.)

   Megoldás: $m := \mathbb{E}X_i = 0$, $\sigma^2 := VarX_i = \mathbb{E}(X_i^2) = \int_{-1}^{1} \frac{1}{2}x^2 \,\mathrm{d}x = \frac{1}{3}$. (Ugyanezt kiolvashatjuk a hivatalos puskából is.)

   a.) $\mathbb{E}S_n = nm = 0$, $VarS_n = n\sigma^2 = 100$, így S_n-t a CHT szerint normális eloszlással közelítve

   $$\mathbb{P}(S_n > 30) \approx 1 - \Phi\left(\frac{30 - nm}{\sqrt{n}\sigma}\right) = 1 - \Phi\left(\frac{30}{10}\right) = 1 - \Phi(3) \tag{1}$$
   $$\approx 1 - 0.9987 = 0.0013 \approx 0.13\% \tag{2}$$

   (Itt Φ a standard normális eloszlásfüggvény.)

   b.) a Berry-Esseen tétel alkalmazásához ki kell számolnunk $\delta := \mathbb{E}\left(|X_i|^3\right)$-t. Mivel X_i sűrűségfüggvénye $[-1, 1]$-en $\frac{1}{2}$,

   $$\delta = \frac{1}{2}\int_{-1}^{1} |x|^3 \,\mathrm{d}x = \frac{1}{2}\left(\int_{-1}^{0} (-x)^3 \,\mathrm{d}x + \int_{0}^{1} x^3 \,\mathrm{d}x\right) = 2 \cdot \frac{1}{2}\int_{0}^{1} x^3 \,\mathrm{d}x = \frac{1}{4}.$$

   Így a Berry-Esseen tétel szerint a CHT közelítés hibájára

   $$|\text{hiba}| \leq \frac{C\delta}{\sqrt{n}\sigma^3} = \frac{0.4748 \cdot \frac{1}{4}}{\sqrt{300}\left(\frac{1}{\sqrt{3}}\right)^3} = \frac{3 \cdot 0.4748}{4 \cdot 10} = 0.03561 \approx 3.6\%$$

2. Egy vizsgán 120 hallgató jelenik meg, közülük 90 készült, 30 pedig nem. Aki készült, 90% valószínűséggel megy át, aki viszont nem, az csak 30%-kal. Adjuk nagy eltérés becslést annak valószínűségére, hogy a hallgatók legalább fele megbukik.

   (A p paraméterű Bernoulli eloszlás Cramér féle rátafüggvénye

   $$I(x) = x \ln\left(\frac{x}{1 - x}\frac{1 - p}{p}\right) + \ln\left(\frac{1 - x}{1 - p}\right) \quad \textit{(ha } 0 < x < 1\textit{).)}$$

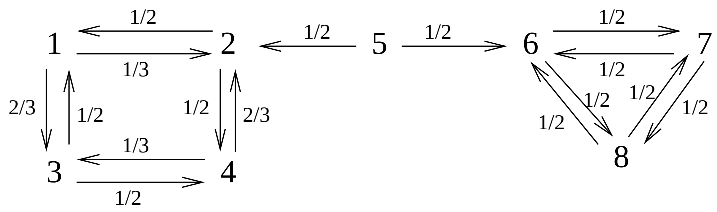
   Megoldás: Legyen $n = 120$ és $k = 1, 2, \dots, n$-re $X_k = 1$, ha a k-adik hallgató megbukik, és $X_k = 0$, ha nem. Legyen $S_n := X_1 + \dots + X_n$ a bukott hallgatók száma, így $\mathbb{P}(S_n \geq 60)$-ra keresünk nagy eltérés becslést.

   Mivel az X_k-k nem azonos eloszlásúak (a készületlenek nagyobb valószínűséggel buknak), a Cramér tétel (közvetlenül) nem alkalmazható, így a Hoeffding-egyenlőtlenséget fogjuk alkalmazni. Ehhez minden k-ra $a_k = 0$, $b_k = 1$-gyel $a_k \leq X_k \leq b_k$, valamint $\mathbb{E}S_n =$

$90 \cdot 0.1 + 30 \cdot 0.7 = 30$, vagyis $t = 30$ választással a Hoeffding-egyenlőtlenség azt adja, hogy

$$\mathbb{P}(S_n \geq 60) = \mathbb{P}(S_n \geq \mathbb{E}S_n + t) \leq \exp\left\{ -\frac{2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2)} \right\} = e^{-\frac{2 \cdot 30^2}{120 \cdot 1^2}} = e^{-15} \approx 3.06 \cdot 10^{-7}.$$

3. Legyen az X_n diszkrét idejű Markov lánc gráf-reprezentációja a következő:



Adjuk meg közelítőleg az alábbi valószínűségeket. A válaszokat indokoljuk.

a.) $\mathbb{P}(X_{1000} = 7 \mid X_0 = 6) \approx ?$

b.) $\mathbb{P}(X_{1000} = 2 \mid X_0 = 1) \approx ?$

c.) $\mathbb{P}(X_{1000} = 2 \mid X_0 = 6) \approx ?$

d.) $\mathbb{P}(X_{1000} = 7 \mid X_0 = 5) \approx ?$

Megoldás: A Markov lánc NEM irreducibilis. Három osztálya közül kettő zárt: a $C_1 := \{1, 2, 3, 4\}$ osztály periódusa 2, a $C_2 := \{6, 7, 8\}$ osztály pedig aperiodikus, mert pl. 6-ból 6-ba vissza lehet jutni 2 és 3 lépésben is. Így

a.) $\mathbb{P}(X_{1000} = 7 \mid X_0 = 6) \approx \frac{1}{3}$, mert 6-ból indulva örökre bent maradunk a C_2 osztályban. A Markov lánc ide megszorítva irreducibilis és aperiodikus, így a Markov láncok alaptétele szerint hosszú idő elteltével az eloszlás a stacionáriussal közelíthető. A C_2 irreducibilis komponensen a(z egyetlen) π stacionárius eloszlás szimmetria okból az egyenletes, így $\mathbb{P}(X_{1000} = 7 \mid X_0 = 6) \approx \pi_7 = \frac{1}{3}$.

b.) $\mathbb{P}(X_{1000} = 2 \mid X_0 = 1) = 0$, mert 1-ből indulva örökre bent maradunk a C_2 osztályban, ez viszont periodikus 2 periódussal, így páros sok lépésben csak 1-be és 3-ba juthatunk el.

c.) $\mathbb{P}(X_{1000} = 2 \mid X_0 = 6) = 0$, mert 6-ból indulva örökre bent maradunk a C_2 osztályban, vagyis 2-be nem lehet eljutni.

d.) $\mathbb{P}(X_{1000} = 7 \mid X_0 = 5) \approx \frac{1}{6}$, mert 5-ből indulva $\frac{1}{2}$ valószínűséggel az első lépésben a C_1 osztályba lépünk és ott is ragadunk, $\frac{1}{2}$ valószínűséggel viszont a C_2-be, és innen kezdve az a.) pont szerinti $\frac{1}{3}$ az esélyünk hosszú idő alatt 7-be érkezni.

4. A Faláb FC focicsapatának 4 csatára van összesen. A csatárok közül esetleg néhány sérült. A csapat mindig 2 egészséges csatárral játszik (ha ennél kevesebb csatáruk egészséges, akkor az összes egészséges csatár játszik). Ha egy csatár játszik, akkor átlagosan 3 havonta sérül le. Egy sérülés átlagosan 1 hónapig tart. Ha egy csatár nem játszik, nem sérül meg.

Jelölje az egészséges csatárok számát a t időpontban X_t. Az időt mérjük hónapokban.

a.) Modellezzük X_t-t folytonos idejű Markov-lánccal! Írjuk fel a generátort. *Legyünk észnél a rátákkal!*

b.) Számítsuk ki a stacionárius eloszlást.

c.) Az idő mekkora részében kénytelen a csapat csatár nélkül játszani? Miért?

d.) Átlagosan hány csatárral játszanak? Miért?

e.) Tegyük fel, hogy éppen minden csatár egészséges. Becsüljük meg annak a valószínűségét, hogy a következő 10 napban ez végig így marad (a 10 napot tekinthetjük 1/3 hónapnak).

Megoldás: Az állapottér $S = \{0, 1, 2, 3, 4\}$. Ha 0 csatár egészséges, persze nincs sérülés. Ha 1 egészséges (vagyis $X_t = 1$), akkor az az egy $\frac{1}{3}$ rátával sérül meg, vagyis $\lambda_{10} = \frac{1}{3}$. Ha 2 csatár egészséges, akkor mindkettő játszik is, így *valamelyikük* már $\lambda_{21} = 2 \cdot \frac{1}{3} = \frac{2}{3}$ rátával sérül meg. Ha 3 vagy 4 csatár egészséges, akkor is csak kettő játszik, így a sérülés rátája ugyanennyi: $\lambda_{32} = \lambda_{43} = \frac{2}{3}$.

Ha minden csatár egészséges, akkor persze egy se tud meggyógyulni. A pontosan 1 sérült (vagyis $X_t = 3$), akkor azaz egy 1 rátával épül fel, vagyis $\lambda_{34} = 1$. Ha 2 sérült, akkor mindkettő lábadozik, így *valamelyikük* már $\lambda_{23} = 2 \cdot 1 = 2$ rátával épül fel. Ugyanígy, ha 3 sérült, akkor mindhárom lábadozik, ezért $\lambda_{12} = 3$, és ha mind a 4 sérült, akkor mind a négy lábadozik, így $\lambda_{01} = 4$.

Mivel egy valószínűséggel egyszerre csak egy csatár tud megsérülni vagy felépülni (a folytonos Markov modell szerint), az összes többi (nem szomszédos állapotok közötti) ugrási ráta 0.

a.) Így az infinitezimális generátor

$$G = \begin{pmatrix} -4 & 4 & 0 & 0 & 0 \\ 1/3 & -10/3 & 3 & 0 & 0 \\ 0 & 2/3 & -8/3 & 2 & 0 \\ 0 & 0 & 2/3 & -5/3 & 1 \\ 0 & 0 & 0 & 2/3 & -2/3 \end{pmatrix}.$$

(A főátlón kívülre az ugrási rátákat írjuk, a főátlóba meg annyit, hogy minden sorösszeg nulla legyen.)

b.) A $\pi = (\pi_0, \ldots, \pi_4)$ stacionárius eloszlás kiszámításához vagy megoldjuk a $G^T \pi^T = 0$ homogén lineáris egyenletrendszert azzal a kiegészítő feltétellel, hogy $\pi_0 + \cdots + \pi_4 = 1$, vagy kihasználjuk, hogy X_t születési-halálozási folyamat, amiből *szomszédos* i, j állapotokra $\frac{\pi_i}{\pi_j} = \frac{\lambda_{ji}}{\lambda_{ij}}$. Mindkettőből az jön ki, hogy

$$\pi = c \cdot \begin{pmatrix} 1 & 12 & 54 & 162 & 243 \end{pmatrix} = \begin{pmatrix} \frac{1}{472} & \frac{12}{472} & \frac{54}{472} & \frac{162}{472} & \frac{243}{472} \end{pmatrix}$$
$$\approx \begin{pmatrix} 0.002 & 0.025 & 0.114 & 0.343 & 0.515 \end{pmatrix}.$$

c.) A 0 állapotban eltöltött időnek a teljes időhöz mért aránya nem egyéb, mint $f(X_t)$ időátlaga, ahol $f : S \to \mathbb{R}$ a 0 állapot indikátora: oszlopvektor formájában $f = (1\,0\,0\,0\,0)^T$. Mivel a Markov lánc folytonos idejű, véges állapotterű és irreducibilis, az ergodtétel szerint ez az időátlag tart $\pi f = \pi_0 \approx 0.002$-hez, vagyis a Faláb FC hosszú távon az idő kb. 2 ezrelékében játszik csatár nélkül.

d.) A játszó csatárok számát az állapot függvényében a $g = (0\,1\,2\,2\,2)^T$ függvény adja meg. Ennek időátlaga hosszú távon – ismét az ergodtétel miatt $\pi g = \pi_1 + 2 \cdot (\pi_2 + \pi_3 + \pi_4) \approx 1.97$.

e.) A 4 állapotból való elugrás rátája $\frac{2}{3}$, így annak valószínűsége, hogy $t = \frac{1}{3}$ ideig nem történik ugrás, pontosan annak valószínűsége, hogy egy $\lambda = \frac{2}{3}$ paraméterű exponenciális eloszlás $t = \frac{1}{3}$-nál nagyobb értéket vesz fel, vagyis $1 - F_{\lambda = \frac{2}{3}}(\frac{1}{3}) = e^{-\frac{2}{3} \cdot \frac{1}{3}} = e^{-\frac{2}{9}} \approx 0.80$.

5. Egy hegy tengerszint feletti magasságára vagyunk kíváncsiak, de csak hibával terhelten tudjuk megmérni: a mérési eredmény normális eloszlású valószínűségi változó, melynek

várható értéke az (általunk nem ismert) tényleges magasság, szórása pedig 20 méter. Végrehajtottunk 10 egymástól független mérést, és a következő számokat kaptuk (méterben): 7009, 7023, 6999, 6994, 6978, 7014, 6989, 6997, 7009, 6993.

Döntsünk 95%-os szinten arról a hipotézisről, hogy a hegy legalább 7000 méter magas.

Megoldás: Egymintás egyoldali u-próbát végzünk $X \sim \mathcal{N}(m, 20^2)$ eloszlású mintával, ahol $\mu = 7000$ és a nullhipotézis $m \geq \mu$. Ehhez először kiszámoljuk a $\bar{x}$ mintaátlagot, ami $\bar{x} = 7000.5$-nek adódik. Ebből $\bar{x} > \mu$ miatt rögtön látszik, hogy a teszt-statisztika *pozitív lesz*. A hipotézis-beli egyenlőtlenség iránya olyan, hogy ez éppen hogy megerősíti a hipotézist (a t-t valami negatív küszöbszámmal kellene összehasonlítani), így további számolás nélkül biztos, hogy *a hipotézis elfogadjuk*.