Sztochasztika 2 vizsga Felsőbb matematika tárgy.

2014. január 21. 12:00. Munkaidő: 60 perc.

1. Egy kis telefonközpontba érkező, egymást követő hívások között eltelt idő mindig exponenciális eloszlású 1 perc várható értékkel, és független az előzményektől. Adjunk nagy eltérés becslést annak valószínűségére, hogy reggel 8 órától számítva a 400-adik hívásra kevesebb, mint 5 órát kell várni.

   (Segítség: a λ paraméterű exponenciális eloszlás Cramer féle rátafüggvénye

   $$I(x) = \lambda x - \ln(\lambda x) - 1 \quad (\text{ha } x > 0).$$

   A λ paraméterű Poisson eloszlás Cramer féle rátafüggvénye

   $$I(x) = x \ln(x/\lambda) - x + \lambda \quad (\text{ha } x > 0).)$$

1. Megoldás: Legyen $n = 400$ és $X_1, X_2, \ldots, X_n$ független azonos 1 paraméterű exponenciális eloszlású valószínűségi változók: azt jelentik, hogy az egyes hívások között mennyi idő telik el (percben). Így $S_n := X_1 + \cdots + X_n$ a 400-adik hívás ideje, és a kérdés $\mathbb{P}(S_n \leq 300)$. Erre a Hoeffding-egyenlőtlenség *nem alkalmazható*, mert az X_k-k nem korlátosak. Marad a Cramer tétel. Ehhez a kérdéses valószínűséget $\mathbb{P}(S_n \leq 300) = \mathbb{P}(\frac{S_n}{n} \in (0, \frac{3}{4}]) = \mathbb{P}(\frac{S_n}{n} \in (a, b])$ alakba írjuk. Mivel $\mathbf{E}X_k = m$-re $b < m$, a Cramer tétel szerint (az exponenciális eloszlás rátafüggvényét használva $\lambda = 1$-gyel)

$$\mathbb{P}(\frac{S_n}{n} \in (a, b]) \lessapprox e^{-nI(b)} = e^{-400I(\frac{3}{4})} \approx e^{-15.07} \approx 2.84 \cdot 10^{-7}.$$

2. Megoldás: Vegyük észre, hogy a hívások Poisson folyamat szerint érkeznek, ezért az 1 perc alatt érkező hívások száma Poisson eloszlású $\lambda = 1$ várható értékkel, és az egyes percek függetlenek. Így ha $n = 300$ és $S_n = X_1 + \cdots + X_n$ az 5 óra alatt befutott hívások száma, ahol $X_k \sim Poi(1)$, akkor a kérdés $\mathbb{P}(S_n \geq 400)$. Mivel $\frac{4}{3} > m = \mathbf{E}X_k = 1$, a Cramer tétel szerint (a Poisson eloszlás rátafüggvényét használva $\lambda = 1$-gyel)

$$\mathbb{P}(S_n \geq 400) = \mathbb{P}(\frac{S_n}{n} \in [\frac{4}{3}, \infty)) \lessapprox e^{-300 \cdot I(\frac{4}{3})} \approx e^{-15.07} \approx 2.84 \cdot 10^{-7}.$$

3. Megoldás: Pontosan ugyanezt kapjuk akkor is, ha egybevesszük az 5 óra alatt érkező összes hívást: a 300 perc alatt érkező hívások száma Poisson eloszlású $\lambda = 300$ várható értékkel. Így alkalmazhatjuk a Cramer tételt az $S_n = X_1$ egytagú összegre ($n = 1$), ahol $X_1 \sim Poi(300)$, és a kérdés $\mathbb{P}(S_n \geq 400)$. Mivel $400 > m = \mathbf{E}X_1 = 300$, a Cramer tétel szerint (a Poisson eloszlás rátafüggvényét használva $\lambda = 300$-zal)

$$\mathbb{P}(S_n \geq 400) = \mathbb{P}(\frac{S_n}{n} \in [400, \infty)) \leq e^{-1 \cdot I(400)} \approx e^{-15.07} \approx 2.84 \cdot 10^{-7}.$$

2. 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 résznél 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 $\lambda_2 1 = 2 \cdot \frac{1}{3} = \frac{2}{3}$ 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 $\lambda_3 4 = 1$. Ha 2 sérült, akkor mindkettő lábadozik, így *valamelyikük* már $\lambda_2 3 = 2 \cdot 1 = 2$ rátával épül fel. Ugyanígy, ha 3 sérült, akkor mindhárom lábadozik, ezért $\lambda_1 2 = 3$, és ha mind a 4 sérült, akkor mind a négy lábadozik, így $\lambda_0 1 = 4$.

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 $\pi f = \pi_0 \approx 0.002$, 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 $1 - F_{\lambda = \frac{2}{3}}(\frac{1}{3}) = e^{-\frac{2}{3} \cdot \frac{1}{3}} = e^{-\frac{2}{9}} \approx 0.80$.

3. Mintát vettünk egy X normális eloszlású valószínűségi változóból, melynek várható értéke *ismert*: $m = 1000$, de szórása ismeretlen. Azt kaptuk, hogy 997, 1002, 998, 1003, 996, 1001, 998, 1004, 1005. Adjunk maximum likelihood becslést az eloszlás szórására.

Megoldás: $n = 9$, $m = 1000$ ismert és a likelihood-függvény

$$L(\sigma) = \prod_{i=1}^{n} f_{m,\sigma}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - m)^2}{2\sigma^2}},$$

amiből a log-likelihood függvény

$$l(\sigma) = \ln L(\sigma) = \sum_{i=1}^{n} \left[-\ln \sqrt{2\pi} - \ln \sigma - \frac{(x_i - m)^2}{2\sigma^2} \right] = c - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2.$$

Ennek maximumát keressük, ehhez megoldjuk a $l'(\sigma) = 0$ egyenletet:

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - m)^2,$$

amiből

$$\sigma_{ML} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2},$$

éppen a (korrigálatlan) tapasztalati szórás. Esetünkben

$$\sum_{i=1}^{n}(x_i - m)^2 = (-3)^2 + 2^2 + (-2)^2 + 3^3 + (-4)^2 + 1^2 + (-2)^2 + 4^2 + 5^2 = 88,$$

amiből

$$\sigma_{ML} = \sqrt{\frac{88}{9}} \approx 3.13\,.$$

4. 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.*