Felsőbb matematika villamosmérnököknek - Sztochasztika

2. ZH megoldások, 2022 ősz

Minden megoldást részletesen indokolni kell. Azon belül minden alkalmazott jelölést be kell vezetni.

Munkaidő: 90 perc

1. Egy járvány során egy fertőzött személy véletlen számú ismerősét fertőzi meg, mielőtt meggyógyul. Röst von Piripócs számításai szerint a megfertőzött ismerősök számának eloszlását az alábbi generátorfüggvény írja le:

$$g(z) = \frac{1}{\sqrt{5-4z}}.$$

a.) Mi az esélye, hogy egy fertőzött személy egyik ismerősének sem adja tovább a betegséget? (2 pont)

b.) Mi az esélye, hogy valaki pontosan egy ismerősét fertőzi meg? (2 pont)

c.) Átlagosan hány embert fertőz meg valaki? (2 pont)

d.) A fertőzött ismerősök továbbadhatják a fertőzést másoknak: mindenki ugyanolyan eloszlás szerint, az előzményektől függetlenül. Mi a valószínűsége, hogy a járvány globális járvánnyá fog kinőni? (Vagyis, mekkora valószínűséggel él tovább a fertőzés „örökké”?) *(Tipp: ha egy harmadfokú egyenletnek egy gyökét tudjuk előre, az segít a többi gyök megtalálásában.)* (4 pont)

Megoldás: Legyen X a megfertőzött ismerősök száma, ennek generátorfüggvénye g.

a.)

$$\mathbb{P}(X=0) = g(0) = \frac{1}{\sqrt{5}}$$

b.)

$$g'(z) = 2\,(5-4z)^{-\frac{3}{2}}$$
$$\mathbb{P}(X=1) = g'(0) = 2 \cdot 5^{-\frac{3}{2}}$$

c.)

$$m := \mathbb{E}X = g'(1) = 2$$

d.) A fertőzöttek között nevezzük *nulladik generációnak* a kezdeti egyetlen fertőzöttet, és $n \geq 0$-ra nevezzük $n+1$-*edik generációnak* azokat, akiket az n-edik generáció tagjai megfertőznek. Legyen Z_n az n-edik generáció tagjainak száma. Így Z_n Galton-Watson elágazó folyamat X egylépéses utódszámmal. Mivel $m = \mathbb{E}X > 1$, a folyamat pozitív valószínűséggel nem hal ki, a kihalási valószínűség pedig a $g(z) = z$ fixpont-egyenlet egyetlen $0 \leq z < 1$ megoldása. Az iterációval kapott numerikus megoldásokat elfogadjuk. Papíron az alábbi módon számolható:

$$z = \frac{1}{\sqrt{5-4z}}$$

Mindkét oldal nemnegatív, így négyzetre emelhetjük.

$$\begin{aligned} z^2 =& \frac{1}{5-4z} \\ z^2(5-4z) =& 1 \\ 4z^3 - 5z^2 + 1 =& 0 \end{aligned}$$

Vegyük észre, hogy $z = 1$ mindig megoldás, így a baloldalon polinom osztást hajthatunk végre $z-1$-gyel.

$$\begin{aligned} 4z^2 - z - 1 =& 0 \\ z =& \frac{1+\sqrt{17}}{8} \approx 0,6404 \end{aligned}$$

a kihalás valószínűsége. Így tehát annak a valószínűsége, hogy nem hal ki, durván 36%.

2. Egy matematika weboldalra átlagosan 8 óránként érkezik egy meme. A világ bármely pontjáról küldhetőek be memek, így az esti órákban sem csökken a gyakoriságuk. A memek 20%-a szóvicc, 30%-a fizikus vicc ($\pi^2 = g = 10$), 40% jelölésekből eredő vicc, 10% pedig egyéb.

   a.) Pistike csak a fizikus vicceket szereti. Mi a valószínűsége, hogy 2 nap alatt legfeljebb egy ilyen viccet talál? (5 pont)

   b.) Egy napon 5 meme érkezett be. Mi az esélye, hogy ebből 3 este 8 és éjfél között jelent meg? (5 pont)

Megoldás:

a.) Válasszunk egy napot egység időnek. Ekkor a bejövő memek Poisson folyamat szerint érkeznek $\lambda = 3$ paraméterrel. A fizikus memek ennek a 30%-os ritkítása, így egy 0.9 paraméterű Poisson folyamattal van dolgunk. Tehát a két nap alatt beérkező fizikus memek száma $X \sim \text{Poi}(1.8)$. Ebből

$$\mathbb{P}(X \leq 1) = e^{-1.8} + 1.8 \cdot e^{-1.8} = 2.8 \cdot e^{-1.8} \approx 0.4628$$

b.) Fontos, hogy *feltételes* valószínűséget keresünk. Ha *tudjuk* a nap folyamán beérkezett memek számát, akkor ezen belül az egyes memek érkezési időpontjai egyenletes eloszlásúak és függetlenek, így mindegyik a többitől függetlenül $p := \frac{4}{24} = \frac{1}{6}$ eséllyel érkezik $20:00 - 0:00$ között. Így tehát az ezen időszakban érkező memek száma $Y \sim \text{Bin}(5, \frac{1}{6})$. Ebből

$$\mathbb{P}(Y = 3) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0.032$$

Alternatív megoldás: Az időt mérjük most órában, és legyen $X_{[a,b]}$ az $[a,b]$ időintervallumban érkező memek száma, így $X_{[a,b]} \sim \text{Poi}(\frac{b-a}{8})$. Fontos, hogy $X_{[0,24]}$ és $X_{[20,24]}$ NEM függetlenek, viszont $X_{[0,20]}$ és $X_{[20,24]}$ már igen, így

$$\begin{aligned} \mathbb{P}(X_{[20,24]} = 3 \,|\, X_{[0,24]} = 5) =& \frac{\mathbb{P}(X_{[20,24]} = 3, X_{[0,24]} = 5)}{\mathbb{P}(X_{[0,24]} = 5)} = \frac{\mathbb{P}(X_{[0,20]} = 2, X_{[20,24]} = 3)}{\mathbb{P}(X_{[0,24]} = 5)} = \\ =& \frac{\mathbb{P}(X_{[0,20]} = 2)\mathbb{P}(X_{[20,24]} = 3)}{\mathbb{P}(X_{[0,24]} = 5)} = \frac{e^{-\frac{20}{8}}\frac{\left(\frac{20}{8}\right)^2}{2!}e^{-\frac{4}{8}}\frac{\left(\frac{4}{8}\right)^3}{3!}}{e^{-\frac{24}{8}}\frac{\left(\frac{24}{8}\right)^5}{5!}} = \\ =& \frac{5!}{2!3!}\left(\frac{20}{24}\right)^2\left(\frac{4}{24}\right)^3 = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0.032 = 3.2\% \end{aligned}$$

3. A piripócsi panel-lakótelepen az utcák szabályos négyzetrácsot alkotnak: 4 észak-déli és 4 kelet-nyugati utca van, 16 utcasarokkal. Egy macska esténként véletlenszerűen sétál a panelek között, minden sarkon egyenlő valószínűséggel választva a kivezető utak közül (beleértve azt is, amelyiken jött), az előzményektől függetlenül.

   a.) Feltéve, hogy a lakótelep dél-nyugati sarkából indult, mi a valószínűsége, hogy két lépés után visszatért? (3 pont)

   b.) Feltéve, hogy a lakótelep dél-nyugati sarkából indult, körülbelül mennyi a valószínűsége, hogy 200 lépés után az észak-keleti sarokban lesz? (3 pont)

   c.) Jellemezzük a macska helyzetét (x, y) koordinátákkal, ahol $x, y = 0, 1, 2, 3$! Cicánk egy (nagyon) hosszú téli estén N-szer ér el valamelyik utcasarokra. Ezeknek hányad részében jár a $(0, 0)$, a $(0, 2)$ illetve a $(2, 1)$ koordinátáknál? (4 pont)

Megoldás:

a.) Bárhová is lép először, onnan mindenképpen $\frac{1}{3}$ a visszalépés valószínűsége, így

$$\mathbb{P}(\text{2 lépésben visszatér} \,|\, \text{sarokból indul}) = \frac{1}{3}.$$

b.) Legyen X_n a macska által az n-edik lépésben felkeresett utcasarok. Így X_n időben stacionárius Markov lánc. Az S állapottér véges (16 elemű) és irreducibilis, a 200 lépés pedig hosszú idő, ezért a *Markov láncok alaptétele* szerint a keresett valószínűség jó közelítéssel az egyetlen π stacionárius eloszlás által az észak-keleti sarokhoz rendelt súly LENNE, ha a Markov lánc aperiodikus LENNE. Ám ez a Markov lánc *periodikus* 2 periódussal, mert visszatérni mindenhova csak páros lépésben lehet. Így azokra a csúcsokra – köztük az észak-keletire, amik elérhetők 200 lépésben, dupla valószínűség jut:

$$\mathbb{P}(X_{200} = \text{ÉK} \,|\, X_0 = \text{DNy}) \approx 2\pi_{\text{ÉK}}.$$

Vegyük észre, hogy X_n *egyszerű szimmetrikus bolyongás* az utcasarkok által alkotott 16 csúcsú gráfon. **Gyakorlatról tudjuk, hogy ennek stacionárius eloszlása arányos a csúcsok fokszámával**: minden x csúcsra

$$\pi_x = \frac{d(x)}{\sum_{y \in S} d(y)}.$$

A fokszámok összege az élek számának kétszerese, esetünkben $2 \cdot 4 \cdot 3 = 48$, vagyis $\pi_x = \frac{d(x)}{48}$. Az észak-keleti sarok fokszáma 2, így

$$\mathbb{P}(X_{200} = \text{ÉK} \,|\, X_0 = \text{DNy}) \approx 2\pi_{\text{ÉK}} = 2\frac{2}{48} = \frac{1}{12}.$$

Megjegyzés: A feladat sajtóhibás. Eredetileg dél-keleti sarkot akartam kérdezni, ahova 200 *lépésben pont nem lehet eljutni, így a valószínűség nulla. Bocs, bocs. Cserébe aki rájött, hogy a Markov lánc periodikus, az megkapott a 3 pontból 2-t.*

c.) Mivel az X_n Markov lánc véges állapotterű és irreducibilis, az N lépés pedig sok, az *ergodtétel* szerint hosszú távon minden x állapotban a (diszkrét) idő π_x hányadát tölti, ahol π az egyetlen stacionárius eloszlás. Ezt szerencsére már az előző pontban kiszámoltuk: hosszú távon az N diszkrét idő

$$\begin{aligned}\pi_{(0,0)} =& \frac{2}{48}\\ \pi_{(0,3)} =& \frac{3}{48}\\ \pi_{(2,1)} =& \frac{4}{48}\end{aligned}$$

hányadát tölti az említett utcasarkokon.

4. Egy tantárgyra 200 hallgató jár. Korábbi évek statisztikái alapján minden hallgató 60% eséllyel ér el legalább elégségest pót-ZH nélkül, a többiektől függetlenül.

a.) Normális eloszlás segítségével adjunk közelítést annak az esélyére, hogy több mint 136 hallgató lesz, akinek nincs szüksége pót-ZH-ra! (5 pont)

b.) Adjunk felső becslést a fenti közelítés hibájára! (5 pont)

Megoldás:

a.) Legyen $n = 200$ és $i = 1, 2, \ldots n$-re legyen $X_i = 1$, ha az i-edik hallgató elkerüli a pótZH-t, és $X_i = 0$, ha nem. Így az X_i valószínűségi változók *i.i.d.* Ber(0.6) eloszlásúak és $S_n := X_1 + \cdots + X_n$ a sikeres hallgatók száma. Így

$$\begin{aligned}m :=& \mathbb{E}(X_1) = 0.6\\ \sigma^2 :=& \mathbb{D}^2(X_1) = 0.6 \cdot 0.4 = 0.24\end{aligned}$$

és a centrális határeloszlás tétel (CHT) szerint $\frac{S_n - nm}{\sqrt{n}\sigma}$ jó közelítéssel standard normális eloszlású, ezért

$$\begin{aligned}\mathbb{P}(S_n \geq 136) =& \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \geq \frac{150 - nm}{\sqrt{n}\sigma}\right) = \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \geq \frac{16}{\sqrt{200 \cdot 0.24}}\right) =\\ & \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \geq \frac{4\sqrt{3}}{3}\right) \approx 1 - \Phi\left(\frac{4\sqrt{3}}{3}\right) \approx 0.01\end{aligned}$$

b.) A CHT közelítés hibája a Berry-Esseen tétellel becsülhető. Ehhez szükségünk van a harmadik centrális momentumra.

$$\begin{aligned}\delta =& \mathbb{E}\left(|X_1 - 0.6|^3\right) = \mathbb{P}(X_1 = 0)|0 - m|^3 + \mathbb{P}(X_1 = 1)|1 - m|^3 =\\ =& 0.4 \cdot 0.6^3 + 0.6 \cdot 0.4^3 = 0.1248\end{aligned}$$

Innen $C = 0.4748$ konstanssal

$$\text{hiba} = \left|\mathbb{P}(S_n \geq 136) - \left(1 - \Phi\left(\frac{4\sqrt{3}}{3}\right)\right)\right| \leq \frac{C\delta}{\sigma^3\sqrt{n}} = \frac{0.4748 \cdot 0.1248}{0.24^{\frac{3}{2}} \cdot \sqrt{200}} \approx 0.036 = 3.6\%$$