Felsőbb matematika villamosmérnököknek - Sztochasztika
második ZH megoldás, 2021 ősz

1. Feladat

a)

$$g(z) = \left(\frac{1}{2-z}\right)^2 = (2-z)^{-2}$$
$$g'(z) = -2(2-z)^{-3} \cdot (-1) = \frac{2}{(2-z)^3}$$
$$m = g'(1) = 2$$

b)

$$\mathbb{P}(Z_2 = 0) = g_{Z_2}(0) = g\left(g(0)\right) = g\left(\frac{1}{4}\right) = \left(\frac{4}{7}\right)^2 = \frac{16}{49} \approx 0,327$$

c)

Mivel $m = 2 > 1$ szuperkritikus esetben vagyunk. A kihalás valószínűsége: z megoldása a $z = g(z)$ egyenletnek és $0 \leq z < 1$.

Numerikusan iterációs módszerrel oldhatjuk meg, ahol $z_{n+1} = g(z_n)$ rekurziót iteráljuk kellően sok lépésig valamilyen $0 \leq z_0 < 1$ kezdeti értékből, például $z_0 = 0$. Számológéppel ezt úgy tehetjük meg, hogy rögzítjük $ANS = 0$-t, majd $\left(\frac{1}{2-ANS}\right)^2$-et ismételgetjük.

Algebrai módszerrel egy harmadfokú polinom gyökét kell megkeresnünk.

$$z = g(z) = \left(\frac{1}{z}\right)^2$$
$$z(2-z)^2 = 1$$
$$z(z^2 - 4z + 4) = 1$$
$$z^3 - 4z^2 + 4z - 1 = 0$$

Innen a harmadfokú gyökét megkereshetjük numerikusan okos számológépekkel, vagy mazochisták alkalmazhatják a harmadfokú megoldóképletet.

Felhasználhatjuk viszont trükknek, hogy $z = 1$ mindig fixpont, így a baloldalt leoszthatjuk $z-1$-el polinomosztással. Ez által egy másodfokút egyenletet kapunk, amit középiskolás módszerekkel is meg tudunk már oldani.

$$z^2 - 3z + 1 = 0$$
$$z = \frac{3 \pm \sqrt{5}}{2}$$

Mivel z-nek $[0, 1[$ intervallumból kell kikerülnie a megoldás $z = \frac{3-\sqrt{5}}{2} \approx 0,382$.

2. Feladat

Megjegyzés: A feladat paraméterei Apple részvények múltbeli adatai alapján lettek megadva. (Értsd: a valóságban is lehet hasznos ilyen számításokat végezni.)

A feladat paraméterei:

$$\begin{aligned} m =& 1,4 \cdot 10^{-3} \\ \sigma =& 1,9 \cdot 10^{-2} \\ \delta =& 1,8 \cdot 10^{-5} \\ n =& 2 \cdot 365 = 730. \end{aligned}$$

Megjegyzés: Vegyük észre hogy egy adott napon bár az átlagos hozam pozitív, a szórás egy nagyságrenddel nagyobb, így gyakorlatilag teljesen véletlen, hogy több vagy kevesebb pénzzel térünk majd haza a nap végén. Viszont ne feledjük el, hogy időben a várható érték lineárisan növekszik, míg a szórás csak gyökösen, így hosszútávon a trend fog nyerni a volatilitással szemben. (Gondoljunk csak a nagy számok törvényére.)

a)

Először centralizálást alkalmazunk

$$\mathbb{P}(S_n < 0) = \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < -\frac{nm}{\sigma\sqrt{n}}\right) \approx \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < -1,99\right).$$

A centrális határeloszlás tétele alapján

$$\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < -1,99\right) \approx \Phi(-1,99) = 1 - \Phi(1,99) \approx 0.023.$$

b)

A Berri-Essén tétel alapján

$$\left|\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} < -1,99\right) - \Phi(-1,99)\right| \leq \frac{C\delta}{\sigma^3\sqrt{n}} \approx 0,046.$$

Megjegyzés: Vegyük észre, hogy a hibatag nagyobb, mint az előrejelzett valószínűség, ha bár a nagyságrendjük megegyezik. Ez a példa szemlélteti, hogy ész nélkül nem szabad normál közelítést alkalmazni, amikor az átlagtól nagyon eltérő értékre vagyunk kíváncsiak, vagy amikor túl kicsi a mintánk.

A számolásaink mindenesetre nem haszontalanok, felső becslésként alkalmazhatjuk a

$$\mathbb{P}(S_n < 0) \leq \Phi(-1,99) + \frac{C\delta}{\sigma^3\sqrt{n}} \approx 0,069$$

a veszteség valószínűségéhez.

3. Feladat

Hoeffding-egyenlőtlenséget kell alkalmaznunk, elvégre feltehetjük, hogy az embereket egymástól függetlenül éri baleset, viszont a balesetből adódó károk eloszlását nem ismerjük, s feltehetően nem is azonos eloszlású az emberek között.

Indexeljük az embereket úgy, hogy $1 \leq i \leq 75000$ indexű kliensek normál, $75001 \leq i \leq 100000$ indexűek pedig prémium szerződést kötöttek.

A költségek nem negatívak és mind a normál, mind a prémium szerződéseknél felülről korlátos az értékük. Ez alapján

$$\begin{aligned} a_i &= 0 \\ b_i &= \begin{cases} 5000 \text{ ha } 1 \leq i \leq 75000 \\ 20000 \text{ ha } 75001 \leq i \leq 100000. \end{cases} \end{aligned}$$

Így a Hoeffding-egyenlőtlenségben használt négyzetösszeg:

$$\sum_{i=1}^{100000} (b_i - a_i)^2 = 7500 \cdot 5000^2 + 2500 \cdot 20000^2 \approx 1,19 \cdot 10^{13}$$

A várható költség

$$\mathbb{E}\left(S_n\right) = \sum_{i=1}^{75000} \mathbb{E}(X_i) + \sum_{i=75001}^{100000} \mathbb{E}(X_i) = 75000 \cdot 50 + 25000 \cdot 100 = 6,25 \cdot 10^6.$$

Bontsuk szét a K korlátot $K = t + \mathbb{E}(S_n)$-re! Hoeffding szerint

$$\mathbb{P}(S_n \geq K) = \mathbb{P}(S_n \geq t + \mathbb{E}(S_n)) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Úgy kell megválasztanunk t-t, hogy a fizetésképtelenség esélye legfeljebb $0,1\%$ legyen. (S természetesen a lehetőségekhez képest a legkisebb t értéknek örülünk.)

$$\begin{aligned} \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) &= 0,001 \\ t &= \sqrt{-\frac{1}{2} - \log(0,001) \sum_{i=1}^{n}(b_i - a_i)^2} \approx 6,4 \cdot 10^6 \end{aligned}$$

A korlátunk tehát $K \approx 1,265 \cdot 10^7$.

Megjegyzés: Ez a korlát egy nagyságrenddel kisebb, mint a konzervatív $8,75 \cdot 10^8$ becslés, melyet úgy kapunk, hogy feltesszük minden kliensnél a maximális kárt kell megfizetnünk.

4. Feladat

Megjegyzés: A 700-as szám realisztikus a KSH adatbázisa szerint, csak nem Pest Megyében, hanem Budapesten. (Elnéztem.) Az autó és motorbalesetek számáról nem találtam megfelelő adatbázist, a százalékokat nem kell komolyan venni.

Az első negyedévben 90 nap van, így napban mérve a ráta az összbalesetekre $\lambda = \frac{700}{90} \approx 7,78$. Hasonlóan az autó és motor balesetek rátája rendre:

$$\begin{aligned}\lambda_a =& 0,8 \cdot \lambda \approx 6,22\\ \lambda_m =& 0,05 \cdot \lambda \approx 0,39.\end{aligned}$$

Mondanom sem kell, Poisson folyamattal modellezzük a jelenséget.

a)

Jelölje N a balesetek számát az adott napon.

$$\mathbb{P}(N = 0) = e^{-\lambda} \approx 4,19 \cdot 10^{-4}$$

b)

Legyenek N_a, N_m rendre az autó és motorbalesetek száma az adott napon, melyek független változók.

$$\begin{aligned}\mathbb{P}(N_a \geq 3, N_m \geq 1) =& \mathbb{P}(N_a \geq 3)\mathbb{P}(N_m \geq 1)\\ =& \left(1 - e^{-\lambda_a} - \lambda_a e^{-\lambda_a} - \frac{\lambda_a^2}{2} e^{-\lambda_a}\right)\left(1 - e^{-\lambda_m}\right) \approx 0,305\end{aligned}$$

5. Feladat

A feladat nem mondja ki pontosan, de egy diszkrét idejű Markov-láncról van szó, ahol az átmenetvalószínűségek skálájából adódva jól közelíthetünk folytonos idejű Markov-folyamattal. Mindkét esetben a stacionárius eloszlást kell kiszámolnunk, mivel a 10. hét egy távoli időpont.

Megoldás diszkrét időben

Van két állapotunk: egészséges (E), és beteg (B), s a hozzá tartozó átmenetmátrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

ahol $p = 1,16 \cdot 10^{-10}$ és $q = 1,16 \cdot 10^{-9}$. Megjegyezzük, hogy $p = 10q$, ami a későbbi számításoknál hasznos lesz.

A stacionárius eloszláshoz meg kell oldanunk a $\pi = \pi P$ sajátvektor egyenletet, vagy komponensenként kiírva

$$\begin{aligned} (1-p)\pi_E + q\pi_B =& \pi_E \\ p\pi_E + (1-q)\pi_B =& \pi_E. \end{aligned}$$

Az egyenlet azonosságra fog vezetni, ami annak tudható be, hogy a megoldások száma végtelen (elvégre $(c\pi) = (c\pi)P$ ugyanúgy megoldás). Ezt úgy orvosolhatjuk, hogy a stacionárius eloszlásnak teljesítenie kell még, hogy $\pi_E + \pi_B = 1$. Ezzel az egyenlettel helyettesíthetjük valamelyik sort - például a másodikat.

$$\begin{aligned} (1-p)\pi_E + q\pi_B =& \pi_E \\ \pi_E + \pi_B =& 1. \end{aligned}$$

Az egyenletet megoldva kapjuk, hogy a stacionárius eloszlás $\pi = [\frac{q}{p+q}, \frac{p}{p+q}] = [\frac{10}{11}, \frac{1}{11}]$, vagyis a gyakorlatvezető körülbelül $\frac{10}{11}$ eséllyel lesz egészséges.

Megoldás folytonos időben (stacionárius eloszlással)

Vegyük észre hogy, mind az átmenetvalószínűségek, mind a lépésközök (ezredmásodperc) kis mennyiségek. Nagyobb időegységet választva, például napot már sokkal barátságosabb számot kapunk, s ilyen skálán a Markov-lánc is jól közelíthető folytonos idejű Markov-folyamattal. (Analóg módon ahhoz, ahogy a kis lépésközű összegek közelíthetőek integrálokkal.)

Válasszuk az egység időt egy napnak, s így a lépésköz: $\Delta t \approx 1,158 \cdot 10^{-8}$. Legyen a megbetegedés átmenetrátája λ míg a gyógyulásé μ. Jó közelítéssel $p \approx \lambda\Delta t$, $q \approx \mu\Delta t$, vagyis (kerekítést elhagyva) $\lambda = 0,01$ és $\mu = 0,1$.

A folyamathoz tartozó rátamátrix

$$G = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

A stacionárius megoldást ezúttal a $\pi G = 0$ egyenlet oldja meg, melyet a diszkrét esettel analóg módon oldhatunk meg. $\pi = [\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu}] = [\frac{10}{11}, \frac{1}{11}]$, a megoldás tehát ismét $\frac{10}{11}$

Megjegyzés : Azzal, hogy napokban mérjük az időt így $t = 6 \cdot 7 + 5 = 47$ időpont eloszlását helyettesítjük a stacionáriussal. Látszólag t sokkal nagyobb, ha az időt mégiscsak ezred másodpercekben mérjük, s így jóval kisebb eltérést gondolhatnánk naivan olyankor. Ez természetesen nem igaz, az idő átskálázása nem befolyásolja az eredményt. Viszont sokkal jobban megítélhetjük "mennyire nagy a t" ha olyan mértékegységet választunk, ahol G mátrix értékei "értelmes számok", melyeket nem kell normál alakban leírnunk, hogy kezelhetőek legyenek.

Megoldás folytonos időben (differenciálegyenlettel)

Természetesen így senki nem oldotta meg a feladatot, de a teljesség kedvéért ez is szerepel itt.

Jelölje $pi(t) = [pi_E(t), pi_B(t)]$ a gyakorlatvezető egészségügyi állapotának eloszlását t időpontban. Kezdetben $pi(0) = [1, 0]$, mivel egészségesként indul. $pi(t)$ időbeli alakulását a Kolmogorov-egyenlet írja le:

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi(t) = \pi(t)G.$$

Az E koponenset kiírva:

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi_E(t) = -\lambda\pi_E(t) + \mu\pi_B(t)$$

Mivel $pi(t)$ egy eloszlás vektor, teljesülnie kell $\pi_B(t) = 1 - \pi_E(t)$-nek, így némi behelyettesítés és átrendezés után kapjuk, hogy

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi_E(t) + (\lambda + \mu)\pi_E(t) = \mu$$

Így egy elsőrendű inhomogén lineáris ODE-hez jutunk, melynek megoldása

$$\begin{aligned}\pi_E(t) =& e^{-(\lambda+\mu)t}\left(\pi_E(0) + \int_0^t \mu e^{(\lambda+\mu)s}\right)\\ =& e^{-(\lambda+\mu)t} + \underbrace{\frac{\mu}{\lambda+\mu}}_{=\pi_E}\left(1 - e^{-(\lambda+\mu)t}\right)\end{aligned}$$

Jól látszik, hogy ahogy $t \to \infty$ $\pi_E(t)$ úgy lesz egyre közelebb a stacionárius eloszláshoz. $t = 47$ mellett a pontos érték

$$\pi_E(47) = e^{-5{,}17} + \frac{10}{11}\left(1 - e^{-5{,}17}\right) \approx 0.9096.$$

Ehhez képest a stacionárius eloszlás becslése $\frac{10}{11} \approx 0,9091$