Sztochasztika 2 vizsga megoldókulcs Felsőbb matematika tárgy.

2015. január 27. 13:00. Munkaidő: ≤60 perc.

1. *(8 pont)* Bergengócia elektromos hálózatára tízezer fogyasztó kapcsolódik. Közülük 9000-nek 32 amperes biztosítéka van, vagyis az általa felvett teljesítmény legfeljebb $32A \times 230V = 7360W$ lehet. A maradék 1000 fogyasztónak 100 amperes biztosítéka van, igy legfeljebb $100A \times 230V = 23000W$ teljesítményt vehet fel. Bergengóciában a „csúcsidő” délután 2-kor van, ekkor mérik a legnagyobb fogyasztást. A bergengóc elektromos műveknek az egyes fogyasztók csúcsidőbeli fogyasztásának eloszlásáról (a fenti korlátokon túl) fogalma sincs, de azt tudják, hogy az egyes fogyasztók fogyasztásai függetlenek, és hogy az *átlagos összfogyasztás* csúcsidőben $3.2 \cdot 10^7 W$. Mekkora kell legyen az elektromos hálózat K össz-teljesítménye (Watt-ban), ha azt akarják, hogy a csúcsidő-beli össz-fogyasztás $1 - 10^{-8}$ valószínűséggel K alatt maradjon?

   Megoldás:
   Legyen $n = 10000$ és a csúcsidőbeli össz-fogyasztás $S_n = X_1 + \cdots + X_{10000}$, ahol $X_1, \ldots, X_{10000}$ az egyes fogyasztók fogyasztásai Watt-ban. A Hoeffding-egyenlőtlenség szerint minden pozitív t-re

$$\mathbf{P}(S_n > \mathbf{E}S_n + t) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

   ahol $\mathbf{E}S_n = 3.2 \cdot 10^7$ a szöveg szerint, a_i és b_i pedig az i-edik fogyasztás alsó illetve felső korlátja: a konkrét esetben mindegyik $a_i = 0$, a b_i pedig a 9000 kisfogyasztóra 7360, az 1000 nagyfogyasztóra pedig 23000. Így a nevezőbeli szumma

$$\sum_{i=1}^n (b_i - a_i)^2 = 9000 \cdot (7360 - 0)^2 + 1000 \cdot (23000 - 0)^2 = 1.0165264 \cdot 10^{12}.$$

   A 10^{-8}-os biztonsághoz legyen tehát $K = \mathbf{E}S_n + t$, ahol

$$\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) = 10^{-8} \quad \text{(és nem pedig } 1 - 10^{-8}\text{)}.$$

   Ez utóbbit t-re megoldva

$$t = \sqrt{-\frac{\sum_{i=1}^n (b_i - a_i)^2}{2} \ln(10^{-8})} = \sqrt{4 \cdot 1.0165264 \cdot 10^{12} \cdot \ln 10} \approx 3.06 \cdot 10^6,$$

   vagyis $K = 3.2 \cdot 10^7 + 3.06 \cdot 10^6 = 3.506 \cdot 10^7$ jó lesz.

2. *(10 pont)* Pistike az ablakból az utca forgalmát nézi. Személyautók és teherautók mennek arra, mindkettő Poisson-folyamat szerint: személyautóból percenként átlagosan 3, teherautóból percenként átlagosan 1. Pistike csak a teherautókat szereti. Jókedve 5-ös skálán változik (1 és 5 között): ha teherautót lát, 1-gyel felfelé ugrik (hacsak nem már előtte is 5-ös volt), ha pedig személyautót, akkor 1-gyel lefelé (hacsak nem már előtte is 1-es volt). Legyen $X(t)$ Pistike jókedve a t időpillanatban, $t \ge 0$.

   a.) *(3 pont)* Modellezzük a rendszert folytonos idejű Markov lánccal. Írjuk fel $X(t)$ generátorát. Indokoljuk.

   b.) *(3 pont)* Határozzuk meg $(X(t), t \ge 0)$ stacionárius eloszlását. (Szabad észrevenni, hogy X véges állapotterű születési-halálozási folyamat.)

   c.) *(2 pont)* Pistike a nézelődést teljes jókedvvel kezdte. Egy óra elteltével arra jár az apukája. Közelítőleg mennyi annak a valószínűsége, hogy Pistikét teljes rosszkedvben (vagyis 1-es állapotban) találja? Miért?

   d.) *(2 pont)* Hosszú távon az idő hány százalékában lesz Pistikének 5-ös jókedve? Miért?

   Megoldás:

a.) Az időt mérjük percben. Az állapottér $S = \{1;2;3;4;5\}$. Ugrani csak szomszédos állapotba lehet, éspedig felfelé 1 rátával (mert a teherautók 1 rátával jönnek), lefelé pedig 3 rátával (mert a személyautók 3 rátával jönnek). Persze az 1-ből csak felfelé, az 5-ből csak lefelé lehet ugrani. Így a generátor

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 3 & -4 & 1 & 0 & 0 \\ 0 & 3 & -4 & 1 & 0 \\ 0 & 0 & 3 & -4 & 1 \\ 0 & 0 & 0 & 3 & -3 \end{pmatrix}.$$

b.) A születési-halálozási folyamat stacionárius eloszlása a szomszédos állapotoknak olyan relatív súlyt ad, ami reciproka az egymásba való átugrások rátái arányának. Vagyis $\pi_1 : \pi_2 = 3 : 1$, $\pi_2 : \pi_3 = 3 : 1$, $\pi_3 : \pi_4 = 3 : 1$, $\pi_4 : \pi_5 = 3 : 1$. Összesítve $\pi_1 : \pi_2 : \pi_3 : \pi_4 : \pi_5 = 81 : 27 : 9 : 3 : 1$. Az aránysort lenormálva

$$\pi = \begin{pmatrix} \frac{81}{121} & \frac{27}{121} & \frac{9}{121} & \frac{3}{121} & \frac{1}{121} \end{pmatrix}.$$

Persze ugyanez jön ki, ha megoldjuk az $A^T\pi^T = 0$ egyenletrendszert (**a transzponálás nagyon fontos**), vagyis azt, hogy

$$\left(\begin{array}{ccccc|c} -1 & 3 & 0 & 0 & 0 & 0 \\ 1 & -4 & 3 & 0 & 0 & 0 \\ 0 & 1 & -4 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 3 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{array}\right).$$

c.) Egy óra hosszú idő. A Markov lánc irreducibilis, véges állapotterű és folytonos idejű, ezért a Markov láncok alaptétele szerint a kiindulási állapottól függetlenül a stacionárius eloszlással közelíthetünk: $\mathbb{P}(X_{60} = 1 \mid X_0 = 5) \approx \pi_1 = \frac{81}{121} \approx 69\%$.

d.) Az $f : S \to \mathbb{R}$ megfigyelhető mennyiség időátlagát keressük, ahol

$$f(i) = \begin{cases} 1, & \text{ha } i = 5 \\ 0, & \text{ha nem} \end{cases},$$

avagy vektor-jelöléssel

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

A Markov lánc irreducibilis és véges állapotterű, így az ergodtétel értelmében az időátlag hosszú távon $\sum_{i\in S} \pi_i f(i) = \pi f = \pi_5 = \frac{1}{121} \approx 0.8\%$.

3. *(7 pont)* Két nagy elektromos ellenállásról szeretnénk eldönteni, hogy melyik a nagyobb. Sajnos az ellenállást mérni csak hibával terhelten tudjuk: a műszerünk által mutatott érték egy valószínűségi változó, aminek a várható értéke a tényleges ellenállás, a szórása pedig $5M\Omega$. Ezért aztán mindkét ellenálláson több mérést is végeztünk, és a következő értékeket kaptuk ($M\Omega$-ban).

A ellenállás	1209	1198	1200	1196	1213	1209	1202	1205	1208	1200
B ellenállás	1198	1202	1191	1198	1192	1201	1193	1193		

Döntsünk 99%-os szinten arról a hipotézisről, hogy az A ellenállás legalább akkora, mint a B.

(Segítség: az „A” adatsor átlaga 1204*, a „B” adatsor átlaga pedig* 1196*.)*

Megoldás:

Kétmintás egyoldali u-próbát végzünk:

- kétmintásat, mert két minta várható értékét kell összehasonlítanunk,
- egyoldalit, mert a hipotézis egy egyenlőtlenség,
- u-próbát, mert a szórások ismertek.

Jelöljük az „A" adatsor hosszát n_1-gyel, elemeit $x_1, \ldots, x_{n_1}$-gyel, a mérés szórását σ_1-gyel. Hasonlóan a „B" adatsor hosszát jelöljük n_2-vel, elemeit $y_1, \ldots, y_{n_2}$-vel, a mérés szórását σ_2-vel. A nullhipotézis H0: $m_x \geq m_y$.

Ezekkel a jelölésekkel $n_1 = 10$, $\bar{x} = 1204$, $n_2 = 8$, $\bar{y} = 1196$, $\sigma_1 = \sigma_2 = 5$.

A teszt-statisztika

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1204 - 1196}{\sqrt{\frac{5^2}{10} + \frac{5^2}{8}}} = \frac{8}{5\sqrt{\frac{1}{10} + \frac{1}{8}}} \approx 3.37.$$

Ezt kell összehasonlítani a $K = u_\varepsilon = \Phi^{-1}(1 - \varepsilon)$ küszöbértékkel, ahol $\varepsilon = 0.01$, mert a hipotérzist 99%-os szinten vizsgáljuk. A táblázat szerint $K = \Phi^{-1}(0.99) \approx 2.33$.

Döntés: Mivel a próbánk egyoldali és a nullhipotézis szerint $m_x \geq m_y$, a nullhipotézist akkor kell elutasítanunk, ha az A adatsor átlaga sokkal kisebb, mint a B adatsor átlaga, vagyis ha u egy túlságosan nagy abszolútértékű negatív szám. Az elutasítás feltétele tehát $u < -K$, ami *nem teljesül*, ezért a nullhipotézist *elfogadjuk.*

Természetesen az u és a K pontos értékének kiszámolása felesleges munka volt: elég annyi, hogy $\bar{x} - \bar{y} > 0$, ezért $u > 0$.