Gyakorló feladatok 2. zh-ra
Sztochasztika
2017 ősz

1. Egy orvos átlagosan 5, maximum 15 percet foglalkozik egy betegével. Adjunk nagyeltérés-becslést annak a valószínűségére, hogy egy 12 órás műszak nem elég 100 beteg ellátására.

   Megoldás. Legyen X_i az i-edik beteggel töltött idő (percben), és $S_n = X_1 + \cdots + X_n$ az $n = 100$ beteggel töltött idő. A kérdés $\mathbf{P}(S > 720)$ (az $S > 720$ esemény éppen az, hogy 12 óra nem elég a 100 betegre).

   Nagyeltérés-becslés, tehát Cramer vagy Hoeffding alapján kell becsülnünk. Cramer nem alkalmazható, mert nem tudjuk a teljes eloszlását az egy beteggel töltött időnek, anélkül pedig a Cramer-féle rátafüggvényt sem. Hoeffding viszont alkalmazható; a feladat szövege alapján $E(S) = 500$ az összes beteggel töltött idő várható értéke, továbbá $0 \le X_i \le 15$, és így

   $$\mathbf{P}(S > 720) = \mathbf{P}(S > 500 + 220) \le e^{-\frac{2 \cdot 220^2}{100 \cdot (15-0)^2}} \approx 0.0135.$$

2. Egy szabályos érmét addig dobálunk, amíg 100 fejet nem kapunk. Becsüljük meg annak a valószínűségét, hogy ehhez elegendő 190 dobás. Legfeljebb mekkora lehet a becslés hibája a Berry-Esseen tétel szerint?

   Megoldás. Ez nem nagyeltérés-becslés lesz, mivel a szükséges fejek számának várható értéke 200, ehhez képest a 190 nem kirívóan kicsi. Centrális határeloszlás-tétel alapján becsülünk. Kétféle módon is meg lehet válaszolni a kérdést.

   1. módszer. Legyen X_i értéke 1, ha az i-edik dobás fej, 0, ha írás, és legyen $S_n = X_1 + \cdots + X_n$. Ekkor a kérdés $\mathbf{P}(S_{190} \ge 100)$ értéke. Az X_i-k Bernoulli eloszlásúak 1/2 paraméterrel, így $m = \mathbf{E}(X_1) = 1/2$ és $\sigma = \sqrt{\mathbf{E}(X_1^2) - (\mathbf{E}(X_1))^2} = \sqrt{1/2 - 1/4} = 1/2$, így

   $$\mathbf{P}(S_{190} \ge 100) = 1 - \mathbf{P}\left(\frac{S_{190} - nm}{\sqrt{n}\sigma} \ge \frac{100 - 190 \cdot 1/2}{\sqrt{190} \cdot 1/2} \right) \approx 1 - \Phi(0.725) = 1 - 0.764 = 0.236,$$

   ahol a közelítés a centrális határeloszlás-tétel miatt igaz.

   2. módszer. Legyen Y_i az i-edik fejhez szükséges dobások száma (az $i-1$-edik fej után). Legyen $S_n = Y_1 + \cdots + Y_n$. Ekkor a kérdés $\mathbf{P}(S_{100} \le 190)$ értéke. Az Y_i-k geometriai eloszlásúak 1/2 paraméterrel, így $m = \mathbf{E}(Y_1) = 2$ és $\sigma = \frac{\sqrt{1/2}}{1/2} = \sqrt{2}$, így

   $$\mathbf{P}(S_{100} \le 190) = \mathbf{P}\left(\frac{S_{100} - nm}{\sqrt{n}\sigma} \le \frac{190 - 100 \cdot 2}{\sqrt{100} \cdot \sqrt{2}} \right) \approx \Phi(-0.714) = 1 - \Phi(0.714) = 1 - 0.761 = 0.239.$$

   ahol a közelítés a centrális határeloszlás-tétel miatt igaz.

   (A két közelítés nem pontosan ugyanazt a becslést adja, de azért elég közel vannak.)

3. Egy szabályos érmét addig dobálunk, amíg 100 fejet nem kapunk. Adjunk nagy eltérés-becslést annak a valószínűségére, hogy ehhez elegendő 150 dobás. (Segítség: a Bernoulli(p) eloszlás Cramér-féle rátafüggvénye $I(x) = x \ln((x(1-p)/(p(1-x))) + \ln((1-x)/(1-p))$, és a GEO($p$) eloszlás Cramér-féle rátafüggvénye $I(x) = x \ln((x-1)/(x(1-p))) + \ln((1-p)/(p(x-1))$.)

   Nagyeltérés-becslést kell alkalmaznunk. Az előző feladattal analóg módon kétféle módszerrel lehet.

   1. módszer. Legyen X_i értéke 1, ha az i-edik dobás fej, 0, ha írás, és legyen $S_n = X_1 + \cdots + X_n$. Ekkor $\mathbf{P}(S_{150} \ge 100)$ értékére kell nagyeltérés-becslést adnunk. Hoeffding alapján (figyelembe véve, hogy $0 \le X_i \le 1$ és $\mathbf{E}(S_{150}) = 150 \cdot 1/2 = 75$):

   $$\mathbf{P}(S_{150} \ge 100) \le \mathbf{P}(S_{150} \ge 75 + 25) \le e^{-\frac{2 \cdot 25^2}{150(1-0)^2}} \approx 2.4 \cdot 10^{-4}.$$

   VAGY Cramer alapján:

   $$\mathbf{P}(S_{150} \ge 100) \le \mathbf{P}\left(\frac{S_{150}}{150} \in \left[\frac{100}{150}, 1 \right] \right) \le e^{-n \inf_{x \in [a,b]} I(x)},$$

   ahol $a = 100/150 = 2/3$, $b = 1$, és $I(x) = x \ln((x/(1-x))) + \ln(2(1-x))$ (mivel $p = 1/2$). Figyelembe véve, hogy az $[a, b]$ intervallum $\mathbf{E}(X_1) = 1/2$-től jobbra van és $I(x)$ $x > \mathbf{E}(X_1)$ esetén növő, az infémum $a = 2/3$-ban vétetik fel, és így

   $$\mathbf{P}(S_{150} \ge 100) \le e^{-150 I(2/3)} \approx 2.045 \cdot 10^{-4},$$

   ami némileg jobb a Hoeffding-ből nyert felső korlátnál. (Mindenesetre bármelyik elfogadható felső korlátnak).

4. Egy ipari gép intenzív használatnak van kitéve; minden nap végén 1/10 valószínűséggel (az előzményektől függetlenül) szervizelésre szorul. A szervizelés fix 1 napig tart, utána ismét használják.

   (a) Modellezzük a gép állapotát Markov-lánccal. Mik az állapotok? Írjuk fel az átmenetvalószínűség mátrixot.

   (b) Feltéve, hogy most éppen működik, mekkora a valószínűsége, hogy 2 nap múlva is működik?

   (c) Hosszú idő alatt az idő mekkora részében működik?

   (d) Átlagosan hány naponta kell szervizelni?

   (e) Amíg a gép dolgozik, napi 55000 forint bevételt termel. A szervizelés díja 11000 forint. Hosszú távon mekkora átlagos napi bevételt termel?

   Megoldás.

   (a) Kétféle állapot van: használat vagy szerviz; az átmenetmátrix

   $$P = \begin{pmatrix} 9/10 & 1/10 \\ 1 & 0 \end{pmatrix}.$$

   (b) $v_2 = v_0 \cdot P^2 = (91/100,\, 9/100)$, tehát 91/100.

   (c) A stacionárius eloszlásra $v_{\mathrm{st}} = (x_1\, x_2)$ jelöléssel

   $$v_{\mathrm{st}} P = v_{\mathrm{st}}, \qquad x_1 + x_2 = 1,$$

   teljesül, ahonnan $v_{\mathrm{st}} = (10/11,\, 1/11)$, tehát az idő 90%-ában működik a gép hosszú távon.

   (d) A szerviz állapotból indulva az ugyanoda való visszatéréshez szükséges lépésszám várható értéke $1/x_2 = 11$ (nap).

   (e) Az ergodtétel szerint a hosszú távú átlagos napi bevétel

   $$55000 \cdot x_1 + (-11000) \cdot x_2 = 49000.$$

5. Egy betörő átlagosan havi 2 betörést követ el, kivéve, amikor börtönben van. Minden egyes betörésért 1/4 eséllyel kapják el. Ha elkapják, börtönbe küldik. A börtönből átlagosan 4 hónap után szabadul és újból nekilát a betöréseknek.

   (a) Modellezzük a betörő állapotát folytonos idejű Markov-lánccal. Mik az állapotok? Írjuk fel a generátort.

   (b) Feltéve, hogy most éppen szabadlábon van, mekkora a valószínűsége, hogy 10 nap múlva is szabadlábon lesz? (1 hónap tekinthető 30 napnak.)

   (c) Mennyi a valószínűsége, hogy jövőre ilyenkor szabadlábon lesz?

   (d) 10 év alatt az idő mekkora részét tölti börtönben?

   (e) Az okozott kár betörésenként átlagosan 100000 forint. Hosszú távon havonta átlagosan mekkora kárt okoz a betörő?

   (a) Megoldás. Két állapot van (szabadon/börtönben van); a szabad→börtön átmenet rátája 2/4 (1/hó), mivel havi 2 betörést követ el, de mindegyikért csak 1/4 eséllyel kapják el. A börtön→szabad átmenet rátája 1/4, mert a várható érték 4 (és $\mathrm{EXP}(\lambda)$ várható értéke $1/\lambda$). G-t már csak ki kell egészíteni úgy, hogy a sorösszegek nullák legyenek:

   $$G = \begin{pmatrix} -2/4 & 2/4 \\ 1/4 & -1/4 \end{pmatrix}$$

   A modellezés egyébként abban merül ki, hogy feltesszük, hogy a betöréseket Poisson-pontfolyamat szerint követi el, valamint a börtönben töltött időről is feltételezzük, hogy exponenciális eloszlású.

   (b) $v(0) = (1\,0)$, és

   $$v(1/3) = e^{G/3} v(0) \approx (I + G/3) v(0) = \begin{pmatrix} 5/6 & 1/6 \\ 1/12 & 11/12 \end{pmatrix} \cdot (1,\, 0) = (5/6,\, 1/6),$$

   így a kérdéses valószínűség kb. 5/6.

(c) Ez pedig már a stacionárius eloszlással közelíthető, ami $v_{\mathrm{st}} = (x_1\, x_2)$ jelöléssel

$$v_{\mathrm{st}} G = 0, \qquad x_1 + x_2 = 1,$$

ahonnan $v_{\mathrm{st}} = (1/3,\, 2/3)$, így jövő ilyenkor 1/3 valószínűséggel lesz szabadlábos.

(d) Ez is a staci eloszlás, tehát az idejének 2/3 részét tölti a börtönben.

(e) Havonta átlagosan 2 betörést csinál, tehát havonta átlagosan 200000 forint kárt okoz, amíg szabadlábon van - ez viszont csak az idő 1/3-a, így hosszú távon az átlagos havi kár, amit okoz, 66667 forint.

6. Egy jogi iroda 3 jogászt foglalkoztat. Az irodába átlagosan havonta 1 ügy érkezik, amit akkor vállalnak el, ha van legalább egy szabad jogász (aki éppen nem dolgozik egy másik ügyön), és ilyenkor a szabad jogászok közül találomra döntik el, ki vállalja el (minden ügyön egy jogász dolgozik). Egy ügy átlagosan 3 hónapig tart. Jelölje $X(t)$ az iroda által vitt ügyek számát t-kor.

   (a) Modellezzük az $X(t)$ folyamatot folytonos idejű Markov-lánccal. Írjuk fel a generátort és rajzoljuk fel a gráf-reprezentációt is.

   (b) Feltéve, hogy most éppen 0 ügyet visz az iroda, becsüljük meg annak a valószínűségét, hogy jövő hét ilyenkor is 0 ügyet visznek (1 hét tekinthető 7/30 hónapnak).

   (c) Mennyi a valószínűsége, hogy jövőre ilyenkor 2 ügyet visznek?

   (d) Az idő mekkora részében dolgozik a három jogász közül a legfiatalabb?

   (e) Éppen 2 jogász dolgozik. Mekkora a valószínűsége, hogy valamelyikük előbb végez az aktuális ügyével, minthogy a harmadik is elvállal egy új ügyet?

   Megoldás. Csak utalás szinten: (b) rövid távú viselkedés (lásd előző feladat), (c) staci eloszlás (ez egy születési-halálozási folyamat, ezt ki lehet használni), (d) $0\cdot x_0 + \frac{1}{3}\cdot x_1 + \frac{2}{3}\cdot x_2 + 1\cdot x_3$, ha $v_{\mathrm{st}} = (x_0\, x_1\, x_2\, x_3)$ (mert pl. ha 2 ügyön dolgoznak, akkor ezen belül az idő 2/3 részében dolgozik a legfiatalabb jogász átlagosan), (e) beágyazott ML-ban a $2\to 1$ átmenet valószínűsége a kérdés.

7. Egy szerver terheltségét a bufferben sorban álló igények számából akarjuk megbecsülni. Tudjuk, hogy ha a terheltség $0 < q < 1$, akkor a bufferben lévő igények száma $\mathrm{PGEO}(1-q)$ eloszlású. 6 különböző időpontban ránézve a szerverre, a bufferben lévő igények számára a következő adódott: 0, 3, 1, 2, 1, 0. Adjunk ML-becslést q értékére a minta alapján.

   Megoldás. A likelihood-függvény (figyelve arra, hogy az $1-q$ paraméterű pesszimista geometriai eloszlással dolgozzunk):

$$L(q) = (1-q)q^0 \cdot (1-q)q^3 \cdot (1-q)q^1 \cdot (1-q)q^2 \cdot (1-q)q^1 \cdot (1-q)q^0 = (1-q)^6 \cdot q^7.$$

   A log-likelihood függvény $l(q) = 6\ln(1-q) + 7\ln q$, és az

$$l'(q) = -\frac{6}{1-q} + \frac{7}{q} = 0$$

   egyenlet megoldása $\hat{q} = 7/13$ a ML-becslés.

8. Egy bor alkoholtartalmát mérjük; a mérés eredménye a valódi alkoholtartalomtól eltér egy additív hibával, melynek eloszlása N(0, 0,5) (%-ban mérve). 5 mérés eredményére a következő adódott: 12,4, 13,1, 12,3, 12,2, 13,5. Döntsünk 95%-os szinten arról a hipotézisről, hogy a bor alkoholtartalma 12,5% azon hipotézis ellenében, hogy az alkoholtartalom nem 12,5%.

   Megoldás.

$$u = \frac{\bar{x} - 12.5}{\sigma}\sqrt{n} = \frac{12.7 - 12.5}{0.5}\sqrt{5} = -0.8944, \quad u_{\varepsilon/2} = \Phi^{-1}(0.975) = 1.96.$$

   Mivel $-u_{\varepsilon/2} \le u \le u_{\varepsilon/2}$ teljesül, így elfogadjuk a H_0 hipotézist, azaz azt, hogy az alkoholtartalom 12,5%.