Valószínűségszámítás, 5. feladatsor, 2022. október 15.

1. Tegyük fel, hogy a holnapi középhőmérséklet 6 várható értékű 2 szórású normális eloszlású valószínűségi változó. Mennyi a valószínűsége, hogy a holnapi középhőmérséklet legfeljebb 10 fok? Mennyi a valószínűsége, hogy a holnapi középhőmérséklet 8 és 10 fok közé esik? Mennyi a valószínűsége, hogy a holnapi középhőmérséklet 0 és 12 fok közé esik?

   Megoldás

   Jelölje X a holnapi középhőmérsékletet. A feladat szövege alapján X eloszlása $N(6, 2^2)$. Mivel $\mathbb{P}(X = a) = 0$, ezért teljesül, hogy $\mathbb{P}(X \leq a) = \mathbb{P}(X < a)$. Ezt felhasználva kapjuk, hogy

$$\mathbb{P}(X \leq 10) = \mathbb{P}(X < 10) = \mathbb{P}(X - 6 < 10 - 6) = \mathbb{P}\Big(\frac{X-6}{2} < \frac{10-6}{2}\Big) = \int_{-\infty}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx =$$
$$= \Phi\Big(\frac{10-6}{2}\Big) = \Phi(2) = 0,9772.$$

   Itt $(X-6)/2$ standard normális (0 várható értékű, 1 szórású) eloszlású, ezért a sűrűségfüggvénye $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. A $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$ függvény értékeit táblázatból olvashatjuk ki.

$$\mathbb{P}(8 < X < 10) = \mathbb{P}(X < 10) - \mathbb{P}(X < 8) = \Phi(2) - \Phi(1) = 0,9772 - 0,8413 = 0,1359$$

   Az utolsó kérdés megválaszolásához használjuk fel, hogy $\Phi(-x) = 1 - \Phi(x)$.

$$\mathbb{P}(0 < X < 12) = \mathbb{P}(X < 12) - \mathbb{P}(X < 0) = \Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) = 2\Phi(3) - 1 = 0,9974$$

   Legyen a készülék élettartama X év, ekkor $X \sim \mathcal{N}(10, 4)$. Keressük azt az a értéket, amelyre: $\mathbb{P}(X < a) = 0,1$. Azaz $\Phi\big(\frac{a-10}{2}\big) = 0,1$. Felhasználva, hogy $\Phi(1,29) \approx 0,9$, és $\Phi(-x) = 1 - \Phi(x)$. Azt kapjuk, hogy $\frac{a-10}{2} = -1,29$, ahonnét $a = 7,42$ év. Ennyi garanciát adhatunk.

2. Egy szoftver frissítéséhez 68 fájlt kell telepíteni, amik egymástól függetlenül 10mp várható értékű és 2mp szórású, normális eloszlású ideig töltődnek.
   a) Mennyi a valószínűsége, hogy a teljes frissítés lezajlik 12 percen belül?
   b) A cég a következő frissítésnél azt ígéri, hogy az már 95% valószínűséggel 10 percen belül betöltődik. Hány fájlból állhat ez a frissítés?
   $(\Phi(2,42) = 0,992, \Phi(1,645) = 0,95)$

   Megoldás

   Legyen X egy fájl telepítési ideje $\mu = 10$ mp várható értékkel és $\sigma = 2$ mp szórással. Jelölje S_n az n db fájl telepítési idejének az összegét ($n = 68$). Mivel független normális eloszlások összege is normális eloszlású, S_n is normális eloszlású, $n\mu$ várható értékkel és $\sqrt{n}\sigma$ szórással. Így, felhasználva, hogy m várható értékű és σ szórású normális eloszlás esetén $(X - m)/\sigma$ standard normális eloszlású valószínűségi változó, és így az eloszlásfüggvénye Φ:
   a)

$$P(\text{teljes frissítés lezajlik 12 percen belül}) = P(S_n < 720) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{720 - 680}{2\sqrt{68}}\right) = \Phi(2,42) = 99,2\%$$

   b)

$$0,95 = P(S_n < 600) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{600 - 10n}{2\sqrt{n}}\right) = \Phi\left(\frac{600 - 10n}{2\sqrt{n}}\right)$$

   Mivel tudjuk, hogy $\Phi(1,645) = 0,95$, így $1,645 = \dfrac{600 - 10n}{2\sqrt{n}}$. Ezt megoldva következik, hogy $n = 57,5$, azaz legfeljebb 57 fájlból állhat a frissítés.

3. Egy elektromos vezetékgyártó cég 40 m-es vezetékeket gyárt 0,2 m szórással. Legfeljebb mennyi annak a valószínűsége, hogy a vezeték hossza legalább 1 m-rel eltér a várható 40 m-es értéktől? Mennyi a pontos érték, ha feltesszük még azt is, hogy a hossz normális eloszlású?

   Megoldás

A Csebisev-egyenlőtlenséget $\varepsilon = 1$ értékre használva

$$P(|X - 40| \geq 1) \leq \frac{D^2(X)}{1^2} = \frac{0,2^2}{1^2} = 0,04$$

Vagyis legfeljebb 0,04 annak a valószínûsége, hogy a vezeték rövidebb, mint 39 m ill. hosszabb, mint 41 m.

Ha X normális eloszlású, a 40 körüli szimmetria miatt:

$$\mathbb{P}(|X-40| \geq 1) = 2\mathbb{P}(X-40 > 1) = 2(1-\mathbb{P}(X < 41)) = 2\left(1-\mathbb{P}\left(\frac{X-40}{0,2} < \frac{1}{0,2}\right)\right) = 2(1-\Phi(5)) = 0,0000006.$$

A táblázat szerint $\Phi(5) \geq 0,9999$, valójában $\Phi(5) = 0,9999997$ (például az R szoftverben `pnorm(5), matlabban normcdf(5))`.

4. Legyen X eloszlása (a) normális eloszlás m és σ paraméterekkel; (b) Poisson-eloszlás $\lambda > 0$ paraméterrel; (c) exponenciális eloszlás $\lambda > 0$ paraméterrel. Határozzuk meg X szórását.

5. Húzzunk egy franciakártya-csomagból két lapot visszatevés nélkül. Jelölje X a kihúzott kárók, Y az ászok számát. (Ötvenkét lap van a csomagban, ebből 13 káró és 4 ász, káró ászból pedig egy van.)

   (a) Adjuk meg X és Y együttes eloszlását.

   (b) Igaz-e, hogy X és Y függetlenek?

   (c) Mennyi $X + Y$ várható értéke? Mennyi $3X + 5Y$ várható értéke?

Megoldás

(a) A csomagban 12 káró nemász, egy káró ász, 3 nemkáró ász, valamint 36 további lap van. Ez alapján például annak a valószínűsége, hogy pontosan 1 káró és 1 ász van: $\frac{\binom{3}{1}\cdot\binom{12}{1}+1\cdot\binom{36}{1}}{\binom{52}{2}} = 5,43\%$. A többi lehetőség ehhez hasonlóan meghatározható:

$A/\diamondsuit$	0	1	2	összesen
0	$47,5\%$	$32,6\%$	5%	$85,1\%$
1	$8,1\%$	$5,4\%$	$0,9\%$	$14,4\%$
2	$0,2\%$	$0,2\%$	0	$0,4\%$
összesen	$55,8\%$	$38,2\%$	$5,9\%$	100%

Vagy például annak valószínűsége, hogy mindkettő 0: $\frac{\binom{36}{2}}{\binom{52}{2}} = 47,5\%$.

(b) Nem igaz, ugyanis például ekkor: $\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 1) = \frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}} \cdot \frac{\binom{13}{1}\binom{39}{1}}{\binom{52}{2}} \approx 5,53\%$ lennne, de ahogyan azt fentebb kiszámoltuk ez a valószínűség $\approx 5,43\%$.

Függetlenség esetén minden k,l-re $\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k) \cdot \mathbb{P}(Y = l)$ teljesülne, sőt $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x) \cdot \mathbb{P}(Y < y)$ minden x, y-ra a pontos definíció.

(c) A várható érték lineráris, azaz $\mathbb{E}(aX+bY) = a\mathbb{E}(X)+b\mathbb{E}(Y)$. Így elég kiszámolni az $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ értékeket.

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_{k=0}^{4} k \cdot \mathbb{P}(X+Y=k) = \sum_{k=0}^{2}\sum_{l=0}^{2}(k+l)\mathbb{P}(X=k,Y=l) = \\ &= \sum_{k=0}^{2}\sum_{l=0}^{2} k\mathbb{P}(X=k,Y=l) + \sum_{k=0}^{2}\sum_{l=0}^{2} l\mathbb{P}(X=k,Y=l) = \\ &= \sum_{k=0}^{2} k\mathbb{P}(X=k) + \sum_{l=0}^{2} l\mathbb{P}(Y=l) = \mathbb{E}(X)+\mathbb{E}(Y)\end{aligned}$$

Ehhez nem kell a függetlenség, ami jó, mert most X és Y nem függetlenek, de így is

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Definíció szerint: $\mathbb{E}(Y) = 1 \cdot \mathbb{P}(Y = 1) + 2 \cdot \mathbb{P}(Y = 2) = 1 \cdot \frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}} + 2 \cdot \frac{\binom{4}{2}}{\binom{52}{2}} \approx 0,1538.$

Hasonlóan számolható $\mathbb{E}(X)$:

$$\mathbb{E}(X) = 1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) = 1 \cdot \frac{\binom{13}{1}\binom{39}{1}}{\binom{52}{2}} + 2 \cdot \frac{\binom{13}{2}}{\binom{52}{2}} = 1 \cdot 38,2\% + 2 \cdot 5,9\% = 0,5$$

Ekkor $\mathbb{E}(X + Y) = 0,65$.

illetve $\mathbb{E}(3X + 5Y) = 3\mathbb{E}(X) + 5\mathbb{E}(Y) = 2,27$, hiszen $3X = X + X + X$, ugyanúgy, mint az előbb.

Általában $\mathbb{E}(cX) = c\mathbb{E}(X)$ minden valós c-re.

A várható értékeket másképp is kiszámíthatjuk: X eloszlása hipergeometrikus eloszlás, ebből kapjuk, hogy a várható értéke $13/52 \cdot 2 = 0,5$, illetve Y eloszlása is hipergeometrikus, így a várható értéke $4/52 \cdot 2 = 2/13$.

6. Egy osztályba 16 fiú és 20 lány jár. Tegyük fel, hogy minden tanítási napon egymástól függetlenül a fiúk $0,04$, a lányok $0,05$ valószínűséggel hiányoznak. Legyen X a jövő hétfőn hiányzó fiúk, Y pedig a jövő hétfőn hiányzó lányok száma.

   (a) Számítsuk ki az összes jövő hétfői hiányzó, vagyis $X + Y$ várható értékét.

   (b) Számítsuk ki X, Y és $X + Y$ szórását.

   (c) Mennyi X és Y kovarianciája?

   (d) Mennyi X és $X + Y$ kovarianciája?

   (e) Mennyi X és $X + Y$ korrelációs együtthatója?

   Megoldás

   (a) X és Y is binomiális eloszlású, így

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 16 \cdot 0,04 + 20 \cdot 0,05 = 1,64.$$

   Itt még nem is használtuk, hogy X és Y függetlenek.

   (b) Binomiális eloszlásnál a szórásnégyzet $np(1 - p)$, így a függetlenség miatt

   $D(X) = \sqrt{16 \cdot 0,04 \cdot 0,96} = 0,784;\quad D(Y) = \sqrt{20 \cdot 0,05 \cdot 0,95} = 0,975;\quad D(X+Y) = \sqrt{D^2(X) + D^2(Y)} = 1,251.$

   (c) Mivel X és Y függetlenek, kovarianciájuk 0.

   (d) A kovariancia bilineáris tulajdonsága és X, Y függetlensége alapján

$$\text{cov}(X, X + Y) = \text{cov}(X, X) + \text{cov}(X, Y) = D^2(X) + 0 = D^2(X) = 0,615.$$

   $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

   (e) Az előzőek és a korrelációs együttható definíciója alapján

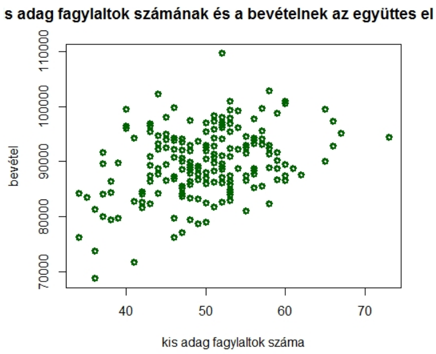
$$R(X, X + Y) = \frac{\text{cov}(X, X + Y)}{D(X)D(X + Y)} = \frac{0,614}{0,784 \cdot 1,251} = 0,626.$$

   A korrelációs együttható mindig a $[-1, 1]$ intervallumba esik. Ez egy közepes erősségű, pozitív irányú (pozitív az előjel, minél nagyobb X, tipikusan annál nagyobb $X + Y$ is), lineáris összefüggésre utal.

7. Egy cukrászdában kétféle terméket árulnak. Tegyük fel, hogy a fagylaltot kérők száma (ez legyen X), Poisson-eloszlású 50 paraméterrel, a süteményt kérők száma Poisson-eloszlású 150 paraméterrel (ez legyen Y), és hogy X és Y függetlenek. A fagylalt ára 300 forint, a süteményé 500.

   (a) Mennyi a napi bevétel várható értéke, illetve szórása?
   (b) Számítsuk ki X-nek és napi bevételnek a korrelációs együtthatóját.

   Megoldás



1. ábra. Az $X + Y$ és $300X + 500Y$ együttes eloszlása kétszáz megfigyelés alapján

(a) Használjuk fel a várható érték linearitását, illetve azt, hogy a Poisson eloszlás paramétere a várható értékkel egyezik meg. $E(300 \cdot X + 500 \cdot Y) = 300 \cdot E(X) + 500 \cdot E(Y) = 300 \cdot 50 + 500 \cdot 150 = 90000$

Mivel X és Y függetlenek, és $D^2(X) = 50$, $D^2(Y) = 150$, ezért $D^2(300 \cdot X + 500 \cdot Y) = 300^2 D^2(X) + 500^2 D^2(Y) = 42000000 \Rightarrow D(300 \cdot X + 500 \cdot Y) = \sqrt{42000000} \approx 6480,74$

(b) A kovariancia tulajdonságai, illetve X és Y függetlensége miatt $cov(X, 300 \cdot X + 500 \cdot Y) = 300 \cdot cov(X, X) + 500 \cdot cov(X, Y) = 300 \cdot D^2(X) + 0 = 300 \cdot 50$. Tehát a korrelációs együttható: $R(X, 300 \cdot X + 500 \cdot Y) = \dfrac{cov(X, 300 \cdot X + 500 \cdot Y)}{D(X) \cdot D(300 \cdot X + 500 \cdot Y)} = \dfrac{300 \cdot 50}{\sqrt{50} \cdot 6480,74} \approx 0,327$

8. Egy szabályos dobókockát kétszer feldobunk. Legyen X a dobások összege, Y a különbségük. Számítsuk ki $\mathrm{cov}(X, Y)$-t és $R(X, Y)$-t! Független-e X és Y? Számítsuk ki $R(X + Y, 2X - Y)$-t is.

Megoldás

Legyen Z_1 az első, Z_2 a második dobás értéke, tehát $X = Z_1 + Z_2$, $Y = Z_1 - Z_2$. Ekkor X és Y kovarianciája a következőképpen írható fel:

$$cov(X, Y) = cov(Z_1 + Z_2, Z_1 - Z_2) = cov(Z_1, Z_1) - cov(Z_1, Z_2) + cov(Z_2, Z_1) - cov(Z_2, Z_2) = D^2(Z_1) - D^2(Z_2) = 0,$$

hiszen Z_1, Z_2 független, azonos eloszlású valószínűségi változók. Mivel a kovariancia 0, ezért a korrelációs együttható is 0, viszont ebből nem következik, hogy X és Y függetlenek, ugyanis

$$0 = P(X = 2, Y = 1) \neq P(X = 2)P(Y = 1) > 0.$$

$R(X + Y, 2X - Y)$ meghatározásához számoljuk ki a $cov(X + Y, 2X - Y)$ értéket. A számolás során használjuk fel, hogy $D^2(X) = D^2(Y)$ és azt, hogy $cov(X, Y) = 0$ (ezt az előbb beláttuk).

$$cov(X + Y, 2X - Y) = 2cov(X, X) - cov(X, Y) + 2cov(X, Y) - cov(Y, Y) = 2D^2(X) - 0 + 0 - D^2(Y) = D^2(X)$$

$$R(X+Y, 2X-Y) = \frac{cov(X + Y, 2X - Y)}{D(X + Y)D(2X - Y)} = \frac{D^2(X)}{\sqrt{D^2(X) + D^2(Y) + 2cov(X, Y)}\sqrt{4D^2(X) + D^2(Y) + 2cov(2X, -Y)}} =$$

$$\frac{D^2(X)}{\sqrt{2D^2(X)}\sqrt{5D^2(X)}} = \frac{D^2(X)}{\sqrt{10}D^2(X)} = \frac{1}{\sqrt{10}} \approx 0,32.$$

9. Tegyük fel, hogy egy ember (szisztolés) vérnyomása minden mérésnél 120 Hgmm várható értékű, 10 szórású valószínűségi változó. Legyen X és Y két vérnyomásmérés eredménye, és tegyük fel, hogy elég sok idő eltelt a két mérés között ahhoz, hogy feltehessük, hogy a mérési eredmények egymástól függetlenek.

(a) Mennyi a két mérés átlagának várható értéke és szórása?

(b) Mennyi lenne a mérések átlagának várható értéke és szórása $n = 10$, illetve $n = 100$ mérés esetén?

(c) Ha n mérés van, mennyi az első k mérés átlagának és az összes mérés átlagának a korrelációs együtthatója?

Megoldás

(a)

$$\mathbb{E}\left(\frac{X+Y}{2}\right) = 120; \qquad D\left(\frac{X+Y}{2}\right) = \frac{10}{\sqrt(2)} = 7,07.$$

(b) A előző számolás 2 helyett n tagra ugyanúgy működik, így az átlag várható értéke 120, a szórása $\frac{D(X_1)}{\sqrt{n}} = \sqrt{10} = 3,16$, illetve 1.

(c) Az első k mérés átlagának és az összes n mérés átlagának korrelációs együtthatója $\sqrt{\frac{k}{n}}$, ugyanis legyen X_j a j. mérés eredménye, ekkor a kovariancia bilineáris tulajdonsága és a függetlenség miatt

$$\operatorname{cov}(X_1 + \ldots + X_k, X_1 + \ldots + X_n) = kD^2(X),$$

így a két átlag kovarianciája $D^2(X)/n$. Ebből a (b) felhasználásával

$$R\left(\frac{X_1 + \ldots + X_k}{k}, \frac{X_1 + \ldots + X_n}{n}\right) = \frac{\frac{D^2(X)}{n}}{\frac{D(X_1)}{\sqrt{k}} \cdot \frac{D(X_1)}{\sqrt{n}}} = \sqrt{\frac{k}{n}}.$$

10. Egy csoportban 25-en tanulnak. Tegyük fel, hogy a tanulók születésnapjai függetlenek és az év tizenkét hónapjában egyenletes eloszlásúak. Számítsuk ki azon hónapok számának a várható értékét és szórását, amelyekre egy születésnap sem esik.

Megoldás

Legyen $\mathbb{I}_j$ annak indikátora, hogy a j. hónapra egyetlen születésnap sem esik. Ekkor

$$\mathbb{E}(\mathbb{I}_j) = \mathbb{E}(\mathbb{I}_j^2) = \left(\frac{11}{12}\right)^{25}; \qquad \mathbb{E}(\mathbb{I}_j\mathbb{I}_k) = \left(\frac{10}{12}\right)^{25},$$

az első ugyanis annak valószínűsége, hogy a j. hónapban senki nem született, a második várható érték pedig annak valószínűsége ($j \neq k$ esetén), hogy sem a j., sem a k. hónapban nem született a tanulók egyike sem, a születések pedig függetlenek, egyenletes eloszlásúak. Legyen X az olyan hónapok száma, ahol egyetlen tanuló sem született. Ekkor, mivel X az $\mathbb{I}_j$-k összege, használhatjuk az összeg várható értékére és szórására vonatkozó összefüggéseket:

$$\mathbb{E}(X) = \sum_{j=1}^{12} \mathbb{E}(\mathbb{I}_j) = 12 \cdot \left(\frac{11}{12}\right)^{25} = 1,36$$

$$D^2(X) = \sum_{j=1}^{12} D^2(\mathbb{I}_j) + 2\sum_{j<k} \operatorname{cov}(\mathbb{I}_j, \mathbb{I}_k) = 12 \cdot \left(\left(\frac{11}{12}\right)^{25} - \left(\frac{11}{12}\right)^{50}\right) + 12 \cdot 11 \cdot \left(\left(\frac{10}{12}\right)^{25} - \left(\frac{11}{12}\right)^{50}\right) = 1,03.$$