Valószínűségszámítás 1 gyakorlat, 5. feladatsor, megoldások

1. Alapfeladatok

1. Legyen X sűrűségfüggvénye $\dfrac{c}{x^4}$ ha $x > 1$, és 0 különben.

   a) Mennyi c értéke?

   b) Számítsuk ki X momentumait minden olyan $k \geq 1$-re, melyre ez véges.

   c) Mennyi X szórása?

   Megoldás.

   a) Általában is,
$$\int_{-\infty}^{\infty} f(x)\,dx = 1,$$
   ha f a sűrűségfüggvény.

   Mivel $1 = \int_1^\infty \frac{c}{x^4}dx = \lim_{t\to\infty}\int_1^t \frac{c}{x^4}dx = \lim_{t\to\infty}\left[\frac{c}{-3\cdot x^3}\right]_1^t = \lim_{t\to\infty}\left[\frac{c}{-3\cdot t^3} - \frac{c}{-3\cdot 1^3}\right] = 0+\frac{c}{3} = \frac{c}{3}$

   $\left(\text{egyszerűbb jelöléssel: } 1 = \int_1^\infty \frac{c}{x^4}dx = \left[\frac{c}{-3\cdot x^3}\right]_1^\infty = \frac{c}{3}\right)$, így következik, hogy $c = 3$.

   b) Vagyis X eloszlásfüggvénye: $F(x) = \int_1^x \frac{3}{t^4}dt = \left[\frac{-1}{t^3}\right]_1^x = -\left(\frac{1}{x^3} - 1\right) = 1 - \frac{1}{x^3}$, ha $x > 1$

   Tehát $\mathbb{E}(X^k) = \int_1^\infty x^k \frac{3}{x^4}\,dx = \left[\frac{-3}{(k-3)\cdot x^{k-3}}\right]_1^\infty = \frac{3}{k-3}$, ha $k = 1, 2$. Ha $k \geq 3$, akkor nem létezik a k. momentum.

   c) $E(X^2) = \int_1^\infty x^2\frac{3}{x^4}\,dx = \left[\frac{-3}{x}\right]_1^\infty = 3$, tehát a szórásnégyzet: $D^2(X) = E(X^2) - E^2(X) = 3 - 1,5^2 = 0,75$, vagyis a szórás $D(X) = \sqrt{0,75} \approx 0,866$

   Ez az eloszlás a Pareto-eloszlások közé tartozik, a Pareto-eloszlásokat például a biztosításmatematikában használják a kárkifizetések modellezésénél.

2. Az X és Y valószínűségi változók együttes sűrűségfüggvénye $h(x,y) = \frac{1}{2}\sin(x+y)$, ha $x, y \in (0, \pi/2)$, és 0 máshol. Határozzuk meg X és Y kovarianciáját, illetve korrelációs együtthatóját.

   Megoldás. (Vázlat)

   Az X (és Y) várható értékéhez meg kell határozni az
$$\mathbb{E}[X] = \int_0^{\pi/2}\int_0^{\pi/2} x\cdot\frac{1}{2}\sin(x+y)\;dx\;dy$$
   integrált.

   A kovariancához pedig az
$$\mathbb{E}[XY] = \int_0^{\pi/2}\int_0^{\pi/2} xy\cdot\frac{1}{2}\sin(x+y)\;dx\;dy$$
   integrált.

   A számolás könnyítése édekében addíciós tétellel a $\sin(x+y)$-t átírjuk $\sin(x)\cos(y)+\cos(x)\sin(y)$ alakba. Az integrálokat tagokra bontjuk, majd kihasználjuk, hogy
$$\int\int f(x)\cdot g(y)\;dx\;dy = \left(\int f(x)\;dx\right)\left(\int g(y)\;dy\right).$$

   Így csak két részfeladatot, az $\int_0^{\pi/2} x\sin(x)\;dx$ és $\int_0^{\pi/2} y\cos(y)\;dy$ integrálokat kell (parciálisan) kiszámolni.

3. Egy szabályos kockával 100-szor dobunk. Jelölje X az első 50 dobás során kapott 6-osok számát, Y pedig az utolsó 75 dobás során dobott páros számok számát. Számoljuk ki $Y-X$ szórásnégyzetét.

   Megoldás. Legyen $X = X_1 + X_2$, ahol X_1 az első 25, X_2 a második 25 dobás során kapott hatosok száma. Hasonlóképpen $Y = Y_2 + Y_3$, ahol Y_2 a második 25, Y_3 pedig az utolsó 50 dobás során dobott páros számoks száma. Ekkor

$$D^2(Y-X) = D^2(Y_3 + (Y_2 - X_2) - X_1) = D^2(Y_3) + D^2(Y_2 - X_2) + D^2(-X_1),$$

   hiszen ezek a tagok különböző dobásokra vonatkoznak, egymástól függetlenek.

   Az Y_3 eloszlása binomiális, $n = 50$ renddel és $p = 1/2$ paraméterrel. Az $Y_2 - X_2$ éppen a kettesek és négyesek száma 25 dobásból, így ez binomiális eloszlású $n = 25$ renddel és $p = 1/3$ paraméterrel. Végül $D^2(cX) = c^2 D^2(X)$ alapján kapható az utolsó tag, ott is egy binomiális eloszlás jelent meg. Összességében a binomiális eloszlás szórására vonatkozó képlet alapján:

$$D^2(Y-X) = 50 \cdot \frac{1}{2} \cdot \frac{1}{2} + 25 \cdot \frac{1}{3} \cdot \frac{2}{3} + 25 \cdot \frac{1}{6} \cdot \frac{5}{6} = 21,53$$

4. Legyenek $1 < k < m < n$ egészek. Tegyük fel, hogy $X_1, \ldots, X_n$ függetlenek, azonos eloszlásúak és véges szórásúak. Számoljuk ki $X_1 + \cdots + X_m$ és $X_k + \cdots + X_n$ korrelációs együtthatóját.

   Megoldás. Mivel a kovariancia bilineáris, és független valószínűségi változóknak 0 a kovarianciája:

$$\operatorname{cov}(X_1 + \ldots + X_m, X_k + \ldots + X_n) = \sum_{j=k}^{m} \operatorname{cov}(X_j, X_j) = (m-k+1)D^2(X_1),$$

   hiszen az azonos eloszlás miatt minden tagnak ugyanannyi a szórása. Felhasználva a független összeg szórására vonatkozó összefüggést is:

$$R(X_1 + \ldots + X_m, X_k + \ldots + X_n) = \frac{(m-k+1)D^2(X_1)}{\sqrt{m \cdot D^2(X_1)} \cdot \sqrt{(n-k+1) \cdot D^2(X_1)}} = \frac{m-k+1}{\sqrt{m(n-k+1)}}.$$

5. 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 $\mathbb{P}(X \le a) = \mathbb{P}(X < a)$. Ezt felhasználva kapjuk, hogy

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

   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 $\Phi(-x) = 1 - \Phi(x)$.

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

6. Tegyük fel, hogy egy tábla csokoládé tömege normális eloszlású 100 g várható értékkel és 3 g szórással, valamint, hogy az egyes táblák tömege egymástól független. Legalább hány csokoládét csomagoljunk egy dobozba, hogy a dobozban levő táblák átlagos tömege legalább 0,9 valószínűséggel nagyobb legyen 99,5 g-nál?

   Megoldás. Legyen $\overline{X}$ a táblák tömegeinek átlaga. Ennek eloszlása n tábla csokoládé esetén: normális eloszlás $m = 100$ várható értékkel és $\sigma = 3/\sqrt{n}$ szórással. Ugyanis, felhasználva, hogy (a) a szórásból a pozitív konstansok kiemelhetők; (b) független valószínűségi változók összegének szórása a szórásnégyzetek összegének gyöke, valamint hogy (c) minden táblánál ugyanannyi a szórás, és (d) az értékét is tudjuk:

$$D((X_1 + \ldots + X_n)/n) \overset{(a)}{=} \frac{1}{n} D(X_1 + \ldots + X_n) \overset{(b)}{=} \frac{1}{n}\sqrt{D(X_1^2 + \ldots + X_n^2)} \overset{(d)}{=}$$
$$= \frac{\sqrt{nD^2(X_1)}}{n} = \frac{D(X_1)}{\sqrt{n}} \overset{(d)}{=} \frac{3}{\sqrt{n}}.$$

Azt is tudjuk, hogy az átlag normális eloszlású, így az előző feladathoz hasonlóan számolhatunk, az $\overline{X}$ várható értékét és szórását helyettesítve m és σ helyére:

$$\mathbb{P}(\overline{X} > 99,5) = 1 - \mathbb{P}(\overline{X} < 99,5) = 1 - \Phi\left(\frac{99,5 - 100}{3/\sqrt{n}}\right) \geq 0,9 \quad \Leftrightarrow \quad \frac{0,5}{3/\sqrt{n}} > 1,28 \quad \Leftrightarrow \quad n \geq 58,9.$$

Vagyis legalább $n = 59$ táblára van szükség.

2. Gondolkodtató feladatok

1. Jelölje ϑ_r az r hosszú ciklusok számát az $\{1, 2, \ldots, n\}$ halmaz egy véletlen permutációjában. Számítsuk ki ϑ_r várható értékét és szórásnégyzetét.

   Megoldás. Minden r hosszú c ciklushoz rendeljünk egy $\mathbb{I}_c$ indikátort, aminek az értéke 1, ha az a ciklus része a permutációnak, és 0 különben. Ekkor $X = \sum_c \mathbb{I}_c$ várható értékére és szórására vagyunk kíváncsiak.

   Az r hosszú ciklusok száma $n(n-1) \cdot \ldots \cdot (n-r+1)/r$, mert sorban haladva ennyiféleképpen választhatunk ki elemeket a még hátralévők közül, viszont minden ciklust r-szer számoltunk, aszerint, hogy hol kezdtük el a felsorolást. Annak valószínűsége, hogy $\mathbb{I}_r = 1$: a véletlen permutációban tetszőleges lehet a sorrend, amiben az elemeket képét kiválasztjuk, azaz

$$\mathbb{E}(\mathbb{I}_c) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{n-r+1}$$

   Ebből

$$\mathbb{E}(X) = \sum_c \mathbb{E}(\mathbb{I}_c) = \frac{n(n-1) \cdot \ldots \cdot (n-r+1)}{r} \cdot \frac{1}{n(n-1)\ldots(n-r+1)} = \frac{1}{r}.$$

   A szórásnégyzet kiszámításához, felhasználva, hogy $X = \sum_c \mathbb{I}_c$, és kibontva a zárójelet (a c egy r hosszú ciklust jelöl):

$$\mathbb{E}(X^2) = \sum_c \mathbb{E}(\mathbb{I}_c^2) + \sum_{c \neq c'} \mathbb{E}(\mathbb{I}_c \mathbb{I}_{c'}) = \frac{1}{r} + \frac{n(n-1) \cdot \ldots \cdot (n-2r+1)}{r^2} \cdot \frac{1}{n(n-1)\ldots(n-2r+1)} = \frac{1}{r} + \frac{1}{r^2}.$$

   Az első tagban ugyanaz a számolás, mint a várható értéknél, hiszen valójában $\mathbb{I}_c^2 = \mathbb{I}_c$, mivel csak 1 és 0 lehetnek az indikátor értékei.

   A második tagban: ha c és c' nem diszjunkt, akkor egyszerre nem jöhetnek létre (minden elem pontosan egy ciklusban van benne egy adott permutációban), így $\mathbb{E}(\mathbb{I}_c \mathbb{I}_{c'} = 0)$ ebben az esetben. Olyan c, c' párból, amik diszjuntak, $\frac{n(n-1) \cdot \ldots \cdot (n-2r+1)}{r^2}$ darab van: először kiválasztjuk azt a $2r$ elemet, amiből ezek állnak (egymás után, sorban haladva), viszont mind a két ciklus elemeit r-féle kezdőponttal számoltuk, ezért kell r^2-tel osztani. A kialakulás valószínűsége ugyanúgy számolható, mint az előző esetben, csak most $2r$ elem van összesen.

   Ebből:

$$D^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{r}.$$

   Megjegyzés. Az r hosszú ciklusok számának eloszlása $n \to \infty$ esetén megfelelő értelemben az $1/r$ paraméterű Poisson-eloszláshoz konvergál.

2. Legyenek X_1, $X_2, \ldots$ olyan valószínűségi változók, hogy minden $i \geq 1$-re $0 < D^2(X_i) < \infty$ és minden $1 \leq i < j$ esetén X_i és X_j korrelációs együtthatója r. (Ahol az $-1 \leq r \leq 1$ szám nem függ i-től és j-től.) Mutassuk meg, hogy $r \geq 0$.

   Megoldás. Feltehető, hogy mindegyik valószínűségi változó várható értéke 0 és szórása 1, hiszen különben standardizálhatjuk őket, és az a korrelációs együtthatókon nem változtat.

   Írjuk fel, hogy minden n-re

$$0 \leq D^2(X_1 + \ldots + X_n) = \sum_{j=1}^{n} D^2(X_j) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j) = n + n(n-1) \cdot r,$$

   hiszen standardizált valószínűségi változóknál a kovariancia és a korrelációs együttható ugyanaz. Átrendezve:

$$-\frac{1}{n-1} \leq r$$

   minden n-re, tehát $r \geq 0$. Vagy: a jobb oldalon egy másodfokú polinom szerepel, aminek főegyütthatója r. Ha ez negatív lenne, a polinom értéke minden elég nagy n-re negatív lenne, ez ellentmondás. Tehát $r \geq 0$. Kérdés: mutassunk példát $n = 3$-ra és negatív r-re.

3. Adjunk példát három olyan valószínűségi változóra, amiknek páronként ugyanannyi a korrelációjuk, és ez a közös érték negatív.

   Megoldás ötlet. Egy háromszögből sorsoljunk egy pontot véletlenszerűen, X_j legyen a j. csúcs együtthatója a pont vektoros előállításában.

3. Gyakorló feladatok

1. Legyen X egyenletes eloszlású az $[1, 4]$ intervallumon. Számítsuk ki $(X-1)^2$ várható értékét!

   Megoldás. Az $X - 1$ valószínűségi változó egyenletes eloszlású a $[0, 3]$ intervallumon. Tehát a sűrűségfüggvénye $f(y) = \mathbb{I}(0 < y < 3)\frac{1}{3}$.

$$\mathbb{E}((X-1)^2) = \int_{-\infty}^{\infty} y^2 f(y)\, dy = \int_0^3 y^2 \cdot \frac{1}{3} dy = 3.$$

   Más megoldás:

$$E((X-1)^2) = D^2(X-1) + E^2(X-1) = \frac{(3-0)^2}{12} + \left(\frac{0+3}{2}\right)^2 = \frac{3}{4} + \frac{9}{4} = 3$$

   Harmadik megoldás: ha g az X sűrűségfüggvény, akkor

$$\mathbb{E}((X-1)^2) = \int_{-\infty}^{\infty} (x-1)^2 g(x)\, dx = \int_1^4 (x-1)^2 \cdot \frac{1}{3}\, dx = 3.$$

2. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} 0, & x \leq 0; \\ c \cdot x, & 0 \leq x \leq 1; \\ 0, & x > 1. \end{cases}$$

   Mennyi c értéke? Mennyi annak valószínűsége, hogy X értéke $1/4$ és $1/2$ közé esik? Mennyi a valószínűsége, hogy $1/2$ és $3/4$ közé esik? Határozzuk meg X eloszlásfüggvényét, várható értékét és szórásnégyzetét.

   Megoldás. Minden sűrűségfüggvényre igaz, hogy

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

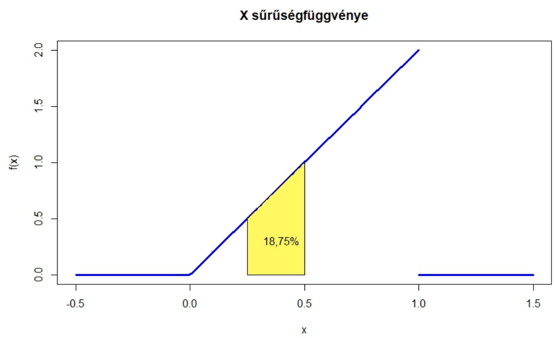
Most
$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 c\cdot x\,dx = c\int_0^1 x\,dx = c\left[\frac{x^2}{2}\right]_{x=0}^{1} = c\cdot\frac{1}{2} = 1 \quad\Rightarrow\quad c = 2.$$
Azokba az intervallumokba, ahol azonosan 0 a sűrűségfüggvény, X nulla valószínűséggel esik, tehát most 1 valószínűséggel a $[0,1]$ intervallumból veszi fel az értékét. Általában, ha az X valószínűségi változó sűrűségfüggvénye f, akkor

$$\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,dx$$

teljesül tetszőleges $a < b$ esetén. Ezért annak valószínűsége, hogy X értéke $1/4$ és $1/2$ közé esik:

$$\mathbb{P}(1/4 \le X \le 1/2) = \int_{1/4}^{1/2} f(x)\,dx = \int_{1/4}^{1/2} 2x\,dx = \left[x^2\right]_{x=1/4}^{1/2} = \left(\frac{1}{2}\right)^2 - \left(\frac{1}{4}\right)^2 = \frac{3}{16} = 18,75\%.$$

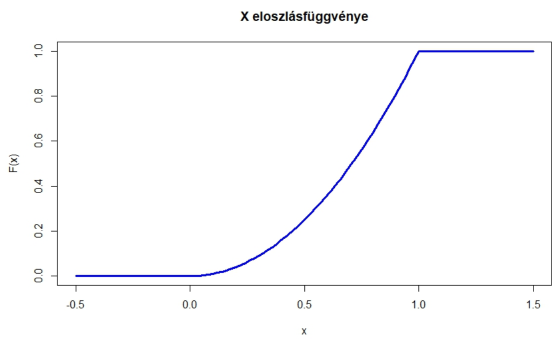
Hasonlóképpen annak valószínűsége, hogy X értéke $1/2$ és $3/4$ közé esik:



$$\mathbb{P}(1/2 \le X \le 3/4) = \int_{1/2}^{3/4} f(x)\,dx = \int_{1/2}^{3/4} 2x\,dx = \left[x^2\right]_{x=1/2}^{3/4} = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16} = 31,25\%.$$

Az eloszlásfüggvény meghatározásához az $F(t) = \mathbb{P}(X < t) = \int_{-\infty}^{t} f(x)\,dx$ összefüggést használhatjuk. Ez a függvény 0, ha $t < 0$, hiszen azonosan nullát integrálunk. Másrészt, ha $0 < t < 1$, akkor
$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \int_0^t 2xdx = \left[x^2\right]_{x=0}^{t} = t^2.$$
Ha pedig $t > 1$, akkor

$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \int_0^1 2xdx = \left[x^2\right]_{x=0}^{1} = 1.$$



Ha két számot választanánk a $(0,1)$ intervallumból egyenletes eloszlással, egymástól függetlenül, és vennénk a nagyobb számot, annak a sűrűségfüggvénye pont a fenti f lenne.

3. Legyen az X és Y valószínűségi változók együttes sűrűségfüggvénye $c(x+y)$, ha $0 \leq x \leq 1$ és $0 \leq y \leq 1$, és 0 különben. Határozzuk meg c értékét, az együttes eloszlásfüggvényt, X sűrűségfüggvényét, valamint X és Y korrelációs együtthatóját.

4. Legyenek A és B tetszőleges események.

   (a) Számoljuk ki $\mathbb{I}_A$ és $\mathbb{I}_B$ korrelációs együtthatóját (itt $\mathbb{I}$ az indikátort jelöli).

   (b) Mutassuk meg, hogy $\mathbb{I}_A$ és $\mathbb{I}_B$ pontosan akkor korrelálatlan, ha A és B függetlenek.

   Megoldás.

$$\mathbb{E}(\mathbb{I}_A) = \mathbb{E}(\mathbb{I}_A^2) = \mathbb{P}(A) \Rightarrow D^2(\mathbb{I}_A) = \mathbb{P}(A)(1-\mathbb{P}(A));$$

$$\mathbb{E}(\mathbb{I}_B) = \mathbb{E}(\mathbb{I}_B^2) = \mathbb{P}(B) \Rightarrow D^2(\mathbb{I}_B) = \mathbb{P}(B)(1-\mathbb{P}(B))$$

$$\mathrm{cov}(\mathbb{I}_A, \mathbb{I}_B) = \mathbb{E}(\mathbb{I}_A\mathbb{I}_B) - \mathbb{E}(\mathbb{I}_A)\mathbb{E}(\mathbb{I}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$$

$$R(\mathbb{I}_A, \mathbb{I}_B) = \frac{\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)}{\sqrt{\mathbb{P}(A)(1-\mathbb{P}(A))\mathbb{P}(B)(1-\mathbb{P}(B))}},$$

   ha pedig valamelyik eseménynek 0 a valószínűsége, akkor a korrelációs együttható 0.

   Az indikátorok pontosan akkor korrelálatlanok, ha a kovarianciájuk 0, a fenti számolásból látható, hogy ez ekvivalens a függetlenség definíciójával.

5. Tegyük fel, hogy Anna busszal és metróval megy a munkahelyére. A buszozás időtartama normális eloszlású 20 várható értékkel és 2 szórással, a metrózás időtartama normális eloszlású 15 várható értékkel és 1 szórással.

   a) Mennyi annak valószínűsége, hogy Anna utazása legfeljebb 40 percig tart?

   b) Mikor kell elindulnia otthonról, hogy legalább 95% valószínűséggel odaérjen 9-re a munkahelyére?

   c) Számítsuk ki a buszozással töltött idő és a teljes utazás idejének korrelációs együtthatóját.

   Felhasználhatjuk, hogy független normális eloszlású valószínűségi változók összege is normális eloszlású.