

9. Gyakorlat

Kovariancia folytonos esetben, korreláció, lineáris regresszió

Végeredmények

1. A várható érték lineáris tulajdonsága miatt $\mathbb{E}(5X - 6Y) = 5 \cdot \mathbb{E}(X) - 6 \cdot \mathbb{E}(Y)$. Ebbe behelyettesítve a feladat szövegében adott várható értékeket $\mathbb{E}(5X - 6Y) = 5 \cdot 4 - 6 \cdot 0 = 20$.

Független valószínűségi változók esetén a szorzat várható értéke egyenlő a várható értékek szorzatával. Mivel a feladat szövege szerint X és Y független, így $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y) = 4 \cdot 0 = 0$.

A szórásnégyzet eltolás-invariáns, ezért $\mathbb{D}^2(5X - 6Y + 8) = \mathbb{D}^2(5X - 6Y)$. Független valószínűségi változók összegének, illetve különbségének szórásnégyzete egyenlő a szórásnégyzetek összegével. Mivel X és Y független, ezért $5X$ és $6Y$ is független, így $\mathbb{D}^2(5X - 6Y) = \mathbb{D}^2(5X) + \mathbb{D}^2(6Y)$. Most felhasználjuk, hogy egy valószínűségi változó konstans c -szeresének szórásnégyzete, a szórásnégyzet c^2 -szerese. Ezért $\mathbb{D}^2(5X) = 5^2 \cdot \mathbb{D}^2(X) = 25 \cdot 1 = 25$ és $\mathbb{D}^2(6Y) = 6^2 \cdot \mathbb{D}^2(Y) = 36 \cdot 2 = 72$. A kért szórásnégyzet tehát $\mathbb{D}^2(5X - 6Y + 8) = 25 + 72 = 97$.

A kovariancia is eltolásinvariáns, ezért $\text{cov}(5X + 2Y + 2, X + 6Y - 3) = \text{cov}(5X + 2Y, X + 6Y)$. A bilineáris tulajdonság miatt $\text{cov}(5X + 2Y, X + 6Y) = 5 \cdot \text{cov}(X, X) + 30 \cdot \text{cov}(X, Y) + 2 \cdot \text{cov}(Y, X) + 12 \text{cov}(Y, Y)$. Mivel a két valószínűségi változó független, ezért a kovarianciájuk nulla: $\text{cov}(X, Y) = \text{cov}(Y, X) = 0$. Egy valószínűségi változó önmagával vett kovarianciája az a valószínűségi változó szórásnégyzete, így $\text{cov}(X, X) = \mathbb{D}^2(X) = 1$ és $\text{cov}(Y, Y) = \mathbb{D}^2(Y) = 2$.

A kért kovariancia tehát $\text{cov}(5X + 2Y + 2, X + 6Y - 3) = 5 \cdot 1 + 30 \cdot 0 + 2 \cdot 0 + 12 \cdot 2 = 29$.

2. $(0, 1)$ -en egyenletes eloszlású valószínűségi változó várható értéke $\frac{1}{2}$, szórásnégyzete pedig $\frac{1}{12}$, így $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{1}{2}$ és $\mathbb{D}^2(X) = \mathbb{D}^2(Y) = \frac{1}{12}$.

a) Z_1 definíciója, valamint X és Y függetlensége miatt $\mathbb{E}(Z_1) = \mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Z_2 definíciója, valamint a várható érték linearitása miatt $\mathbb{E}(Z_2) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \frac{1}{2} + \frac{1}{2} = 1$.

b) Egyrészt $\mathbb{D}^2(XY) = \mathbb{E}[(XY)^2] - [\mathbb{E}(XY)]^2$, ahonnan az előző pont alapján $\mathbb{E}(XY) = \frac{1}{4}$, míg $\mathbb{E}[(XY)^2] = \mathbb{E}[X^2Y^2]$, ami a függetlenség miatt $\mathbb{E}(X^2) \cdot \mathbb{E}(Y^2)$, hiszen, ha két valószínűségi változó független, akkor tetszőleges függvényeik, és így a négyzeteik is azok. A két második momentum megegyezik, mert X és Y azonos eloszlású: $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = \mathbb{D}^2(Y) + (\mathbb{E}(Y))^2 = \frac{1}{12} + \left(\frac{1}{2}\right)^2 = \frac{1}{3}$. Tehát $\mathbb{D}^2(XY) = \mathbb{E}[(XY)^2] - [\mathbb{E}(XY)]^2 = \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) - \left(\frac{1}{4}\right)^2 = \frac{1}{3} \cdot \frac{1}{3} - \frac{1}{16} = \frac{7}{144}$.

Másrészt $\mathbb{D}^2(X) \cdot \mathbb{D}^2(Y) = \frac{1}{12} \cdot \frac{1}{12} = \frac{1}{144}$, ami $\neq \frac{7}{144}$. Tehát $\mathbb{D}^2(XY) \neq \mathbb{D}^2(X) \cdot \mathbb{D}^2(Y)$.

c) Az összeg szórásnégyzetére vonatkozó képlet alapján $\mathbb{D}^2(Z_1 + 2Z_2) = \mathbb{D}^2(Z_1) + \mathbb{D}^2(2Z_2) + 2\text{cov}(Z_1, 2Z_2)$. Az előző pontban meghatároztuk $\mathbb{D}^2(Z_1)$ -t, hiszen $Z_1 = XY$, tehát $\mathbb{D}^2(Z_1) = \mathbb{D}^2(XY) = \frac{7}{144}$. A szórásnégyzet azon tulajdonsága alapján, hogy egy valószínűségi változó konstans c -szeresének szórásnégyzete, a szórásnégyzet c^2 -szerese: $\mathbb{D}^2(2Z_2) = 4 \cdot \mathbb{D}^2(Z_2)$. Definíció szerint $Z_2 = X + Y$, ezért $\mathbb{D}^2(Z_2) = \mathbb{D}^2(X + Y)$, és mivel X és Y függetlenek, ezért $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$.

Hátravan még a kovariancia meghatározása. A kovariancia linearitása miatt $\text{cov}(Z_1, 2Z_2) = 2 \cdot \text{cov}(Z_1, Z_2)$. $\text{cov}(Z_1, Z_2) = \mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1) \cdot \mathbb{E}(Z_2)$, ahonnan Z_1 és Z_2 várhatóértékét az a) pontban meghatároztuk már. A linearitást és a függetlenséget használva, majd behelyettesítve a már kiszámolt értékeket $\mathbb{E}(Z_1 Z_2) = \mathbb{E}(XY(X + Y)) = \mathbb{E}(X^2Y + XY^2) = \mathbb{E}(X^2) \mathbb{E}(Y) + \mathbb{E}(X) \mathbb{E}(Y^2) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}$. Így tehát $\text{cov}(Z_1, Z_2) = \mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1) \cdot \mathbb{E}(Z_2) = \frac{1}{3} - \frac{1}{4} \cdot 1 = \frac{1}{12}$.

Visszatérve a kért szórásnégyzethez:

$$\mathbb{D}^2(Z_1 + 2Z_2) = \mathbb{D}^2(Z_1) + \mathbb{D}^2(2Z_2) + 2\text{cov}(Z_1, 2Z_2) = \frac{7}{144} + 4 \cdot \frac{1}{6} + 2 \cdot 2 \cdot \frac{1}{12} = \frac{151}{144} = 1,049.$$

3. $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \mathbb{E}(Y)$, amihez $\mathbb{E}(X)$ -et már a múltkori feladatban (2024-ben 8.3.) kiszámoltuk: $\mathbb{E}(X) = \frac{13}{20}$, és mivel Y -nak ugyanaz a marginális eloszlásfüggvénye, mint X -nek, ezért $\mathbb{E}(Y)$ is $\frac{13}{20}$ -dal egyenlő. Tehát elég $\mathbb{E}(XY)$ -t kiszámítani, ami (a val. vektorváltozó transzformáltjára vonatkozó képlet szerint, $g(X, Y) = xy$ választással)

$$\begin{aligned} \mathbb{E}(XY) &= \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy = \int_0^1 xy \cdot 2(x^3 + y^3) dx dy = \int_0^1 2x^4 y + 2y^4 x dx dy = \int_0^1 \left[\frac{2x^5 y}{5} + \frac{2y^4 x^2}{2} \right]_0^1 dy \\ &= \int_0^1 \frac{2}{5} y + y^4 dy = \left[\frac{1}{5} y^2 + \frac{1}{5} y^5 \right]_{y=0}^1 = \frac{2}{5}, \end{aligned}$$

tehát $\text{cov}(X, Y) = \frac{2}{5} - (\frac{13}{20})^2 = -\frac{9}{400} (= -0,0225)$.

4. Mivel Y az X -nek egy pozitív, Z pedig egy negatív főegyütthatójú elsőfokú függvénnyel való transzformáltja, azt várjuk, hogy a korreláció -1 lesz.^a Valóban, a kovariancia azonosságai (bilinearitás, eltolás-invariancia/„additív konstans nem számít”) alapján

$$\text{cov}(Y, Z) = \text{cov}(3X + 8, 5 - 2X) = \text{cov}(3X, -2X) = 3 \cdot (-2) \cdot \text{cov}(X, X) = 3 \cdot (-2) \cdot \sigma^2 = -6\sigma^2,$$

mivel $X \sim N(m; \sigma^2)$ és a normális eloszlás második paramétere a szórásnégyzet, továbbá a szórásnégyzet azonosságai szerint

$$\mathbb{D}^2(Y) = \mathbb{D}^2(3X + 8) = \mathbb{D}^2(3X) = 9\mathbb{D}^2(X) = 9\sigma^2$$

és

$$\mathbb{D}^2(Z) = \mathbb{D}^2(5 - 2X) = \mathbb{D}^2(-2X) = 3\mathbb{D}^2(X) = 4\sigma^2,$$

ezért

$$\text{corr}(Y, Z) = \frac{\text{cov}(Y, Z)}{\mathbb{D}(Y)\mathbb{D}(Z)} = \frac{-6\sigma^2}{3\sigma \cdot 2\sigma} = -1.$$

5. $\text{cov}(Y, Z) = \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)$ és mivel X sűrűségfüggvénye 0 és 2π között konstans $\frac{1}{2\pi}$, különben pedig 0 , ezért

$$\begin{aligned} \mathbb{E}(YZ) &= \mathbb{E}(\sin(X)\cos(X)) = \int_{-\infty}^{\infty} \sin(x)\cos(x)f_X(x)dx = \frac{1}{2\pi} \int_0^{2\pi} \sin(x)\cos(x)dx \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2}\sin(2x)dx = \frac{1}{4\pi} \left[-\frac{\cos(2x)}{2} \right]_0^{2\pi} = \frac{1}{2\pi} \left(-\frac{1}{2} + \frac{1}{2} \right) = 0. \end{aligned}$$

Továbbá hasonlóan

$$\mathbb{E}(Y) = \mathbb{E}(\cos(X)) = \frac{1}{2\pi} \int_0^{2\pi} \cos(x)dx = \frac{1}{2\pi} [\sin(x)]_0^{2\pi} = \frac{1}{2\pi} (1 - 1) = 0$$

és

$$\mathbb{E}(Z) = \mathbb{E}(\sin(X)) = \frac{1}{2\pi} \int_0^{2\pi} \sin(x)dx = \frac{1}{2\pi} [-\cos(x)]_0^{2\pi} = \frac{1}{2\pi} (-1 + 1) = 0.$$

Így tehát $\text{cov}(Y, Z) = 0$.

Y és Z nem függetlenek.

1. *érvelés (teljes)*: Tegyük fel ugyanis indirekt, hogy függetlenek, ekkor $g(Y)$ és $h(Z)$ is függetlenek bármilyen $g, h: (0, 2\pi) \rightarrow \mathbb{R}$ folytonos függvényre. Legyen $g(x) = h(x) = x^2$, ekkor $Y^2 = \cos^2(X)$ és $Z^2 = \sin^2(X)$ -ről tudjuk, hogy $Y^2 + Z^2 = 1$ (mert bármilyen 0 és 2π közötti x számra $\cos^2(x) + \sin^2(x) = 1$), tehát

$$g(Y) = \cos^2(X) = 1 - \sin^2(X) = 1 - h(Z).$$

Vagyis $g(Y)$ és $h(Z)$ lineárisan összefüggőek, ahol a lineáris transzformáció főegyütthatója negatív. Továbbá sem $g(Y)$, sem $h(Z)$ nem 1 valószínűséggel konstans (aki nem hiszi, győződjön meg arról, hogy a szórásuk nem 0 ; a szórások pedig léteznek, mert $g(Y)$ és $h(Z)$ is korlátos értékészletű val. változó). Ezért értelmezzük $g(Y)$ és $h(Z)$ korrelációját, ami a változók közötti negatív főegyütthatós lineáris összefüggés miatt -1 lesz. Ez viszont ellentmond annak, hogy $g(Y)$ és $h(Z)$ függetlenek (mert akkor a kovarianciájuk 0 lenne), így az indirekt feltevésünk hamis volt, vagyis X és Y nem függetlenek.

2. *érvelés (vázlatos)*: Tekintsük az $\{Y \in [0, 1/2]\}$ és a $\{Z \in [0, 1/2]\}$ eseményeket. Ekkor

$$\{Y \in [0, 1/2]\} = \{\cos(X) \in [0, 1/2]\} = \{X \in [\pi/3, \pi/2] \cup [3\pi/2, 5\pi/3]\}$$

és

$$\{Z \in [0, 1/2]\} = \{\sin(X) \in [0, 1/2]\} = \{X \in [0, \pi/6] \cup [5\pi/6, \pi]\}.$$

Innen látható, hogy a két esemény kizáró, viszont az X -szel kifejezett alakjukból az is látszik, hogy nem 0 valószínűségűek, így nem is lehetnek függetlenek. Ha viszont Y és Z függetlenek lennének, akkor ezeknek az eseményeknek is függetleneknek kellene lenniük.

^aEzt az érvelést rendes bizonyítással is továbbfejleszthetjük, ha megmutatjuk, hogy Z felírható $Z = aY + b$ alakban, ahol $a < 0$ és $b \in \mathbb{R}$, amiből egyből következik, hogy a kért korreláció -1 . Ezt a lehetséges megoldást az olvasóra bízunk, helyette itt kiszámoljuk a korrelációt.

6. $\text{cov}(U, V^2) = \mathbb{E}(UV^2) - \mathbb{E}(U) \mathbb{E}(V^2)$, ahol

$$\begin{aligned}\mathbb{E}(UV^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv^2 f_{U,V}(u, v) dv du = \int_0^1 \int_0^{u^2} \frac{uv^2}{\sqrt{v}} dv du = \int_0^1 \int_0^{u^2} uv^{3/2} dv du \\ &= \int_0^1 \left[\frac{2}{5} v^{5/2} \right]_{v=0}^{u^2} du = \int_0^1 \frac{2}{5} u \cdot u^5 du = \left[\frac{2}{35} u^7 \right]_{u=0}^1 = \frac{2}{35},\end{aligned}$$

$$\mathbb{E}(U) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u f_{U,V}(u, v) dv du = \int_0^1 \int_0^{u^2} \frac{u}{\sqrt{v}} dv du = \int_0^1 [2u\sqrt{v}]_{v=0}^{u^2} du = \int_0^1 2u^2 du = \left[\frac{2u^3}{3} \right]_{u=0}^1 = \frac{2}{3}$$

és

$$\begin{aligned}\mathbb{E}(V^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v^2 f_{U,V}(u, v) dv du = \int_0^1 \int_0^{u^2} \frac{v^2}{\sqrt{v}} dv du = \int_0^1 \int_0^{u^2} v^{3/2} dv du = \int_0^1 \left[\frac{2}{5} v^{5/2} \right]_{v=0}^{u^2} du = \int_0^1 \frac{2}{5} u^5 du \\ &= \left[\frac{2}{5} u^6 / 6 \right]_{u=0}^1 = \frac{1}{15},\end{aligned}$$

$$\text{tehát } \text{cov}(U, V^2) = \frac{2}{35} - \frac{2}{3} \cdot \frac{1}{15} = \frac{18-14}{315} = \frac{4}{315} \approx 0,01270.$$

7. a) Korrelálatlanok a valószínűségi változók, tehát bármely kettőnek a kovarianciája 0. A kovariancia bilinearitása miatt, valamint mert egy valószínűségi változó önmagával vett kovarianciája az a szórásnégyzete $\text{cov}(X_1 + X_2, X_2 + X_3) = \text{cov}(X_1, X_2) + \text{cov}(X_1, X_3) + \text{cov}(X_2, X_2) + \text{cov}(X_2, X_3) = \text{cov}(X_1, X_2) + \text{cov}(X_1, X_3) + \mathbb{D}^2(X_2) + \text{cov}(X_2, X_3) = 0 + 0 + 1 + 0 = 1$.

$$\mathbb{D}^2(X_1 + X_2) = \mathbb{D}^2(X_1) + \mathbb{D}^2(X_2) + 2\text{cov}(X_1, X_2) = 1 + 1 + 2 \cdot 0 = 2. \text{ Ugyanígy } \mathbb{D}^2(X_2 + X_3) = 2.$$

b) A szükséges mennyiségeket a)-ban kiszámoltuk. Definíció szerint

$$\text{corr}(X_1 + X_2, X_2 + X_3) = \frac{\text{cov}(X_1 + X_2, X_2 + X_3)}{\mathbb{D}(X_1 + X_2) \cdot \mathbb{D}(X_2 + X_3)} = \frac{1}{\sqrt{2} \cdot \sqrt{2}} = \frac{1}{2}$$

c) A lineáris regresszió képletében szereplő α, β paraméterek:

$$\beta = \frac{\text{cov}(X_1 + X_2, X_2 + X_3)}{\mathbb{D}^2(X_1 + X_2)} = \frac{1}{2}$$

$\alpha = \mathbb{E}(X_2 + X_3) - \beta \cdot \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_2) + \mathbb{E}(X_3) - \beta \cdot [\mathbb{E}(X_1) + \mathbb{E}(X_2)] = 0$, hiszen mindegyik valószínűségi változó 0 várható értékű.

A regressziós egyenes egyenlete tehát: $\{(x, y) \in \mathbb{R}^2 \mid y = \beta x + \alpha\} = \left\{ (x, y) \in \mathbb{R}^2 \mid y = \frac{x}{2} \right\}$

d)

1. *megoldás:* Használjuk a tanult általános képletet. A hiba szórásnégyzete Y -nak Z -re vett lineáris regressziója esetén

$$\mathbb{D}^2(Y - (\beta Z - \alpha)) = \mathbb{D}^2(Y) - \frac{(\text{cov}(Z, Y))^2}{\mathbb{D}^2(Z)} = \mathbb{D}^2(Y)(1 - (\text{corr}(Z, Y))^2).$$

Itt $Y = X_2 + X_3$ és $Z = X_1 + X_2$, ezért a hiba szórásnégyzete

$$\mathbb{D}^2(X_2 + X_3 - (\beta(X_1 + X_2) - \alpha)) = \mathbb{D}^2(X_2 + X_3)(1 - (\text{corr}(X_1 + X_2, X_2 + X_3))^2) = 2 \left(1 - \left(\frac{1}{2} \right)^2 \right) = \frac{3}{2}$$

2. *megoldás:* A kérdést megválaszolhatjuk a négyzetes hibára vonatkozó általános képlet nélkül is:

$$\begin{aligned}\mathbb{D}^2(X_2 + X_3 - (\beta(X_1 + X_2) - \alpha)) &= \mathbb{D}^2(X_2 + X_3 + (-\frac{1}{2}(X_1 + X_2))) \\ &= \mathbb{D}^2(X_2 + X_3) + \mathbb{D}^2(-\frac{1}{2}(X_1 + X_2)) + 2\text{cov}(-\frac{1}{2}(X_1 + X_2), X_2 + X_3) \\ &= \mathbb{D}^2(X_2 + X_3) + \frac{1}{4} \cdot \mathbb{D}^2(X_1 + X_2) - 2 \cdot \frac{1}{2} \cdot \text{cov}(X_1 + X_2, X_2 + X_3) \\ &= 2 + \frac{1}{4} \cdot 2 - 2 \cdot \frac{1}{2} \cdot 1 = \frac{3}{2}\end{aligned}$$

használva az a) részben kiszámolt értékeket.

8. A várható érték linearitása és a kovariancia azonosságai miatt (bilinearitás, „additív konstans nem számít”/eltolás-invariancia, $\text{cov}(X, X) = \mathbb{D}^2(X)$) lineáris regresszió képletében szereplő α, β paraméterekre mindhárom esetben teljesül, hogy

$$\beta = \frac{\text{cov}(X^2 + X + 1, X)}{\mathbb{D}^2(X)} = \frac{\text{cov}(X^2, X) + \text{cov}(X, X)}{\mathbb{D}^2(X)} = \frac{\mathbb{E}(X \cdot X^2) - \mathbb{E}(X) \mathbb{E}(X^2) + \mathbb{D}^2(X)}{\mathbb{D}^2(X)} = \frac{\mathbb{E}(X^3) - \mathbb{E}(X) \mathbb{E}(X^2)}{\mathbb{D}^2(X)} + 1$$

és

$$\alpha = \mathbb{E}(X^2 + X + 1) - \beta \mathbb{E}(X) = \mathbb{E}(X^2) + \mathbb{E}(X) + 1 - \beta \mathbb{E}(X).$$

- a) Ha $X \sim \text{Exp}(\lambda)$ (tekintsük először az általános λ esetét, majd utána helyettesítsünk be $\lambda = 5$ -öt), akkor $\mathbb{E}(X) = \frac{1}{\lambda}$, $\mathbb{D}^2(X) = \frac{1}{\lambda^2}$ és $\mathbb{E}(X^2) = \mathbb{E}(X)^2 + \mathbb{D}^2(X) = \frac{2}{\lambda^2}$. Tehát még $\mathbb{E}(X^3)$ -öt kell meghatároznunk. A transzformált várható értékére vonatkozó képlet szerint

$$\mathbb{E}(X^3) = \int_{-\infty}^{\infty} x^3 f_X(x) dx = \int_0^{\infty} x^3 \cdot \lambda e^{-\lambda x} dx.$$

Parciális integrálással (x^3 -öt deriválva, $\lambda e^{-\lambda x}$ -et integrálva) azt kapjuk, hogy

$$\int_0^{\infty} x^3 \cdot \lambda e^{-\lambda x} dx = \underbrace{\left[-\frac{1}{\lambda} \lambda e^{-\lambda x} \cdot x^3 \right]_{x=0}^{\infty}}_{=0} + \int_0^{\infty} -\frac{1}{\lambda} \lambda e^{-\lambda x} \cdot 3x^2 dx = \frac{3}{\lambda} \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx.$$

Vegyük észre, hogy a jobb oldalon álló, $3/\lambda$ -val megszorozott integrál éppen $\mathbb{E}(X^2)$ -nek felel meg, amelyről tudjuk, hogy $2/\lambda^2$ -tel egyenlő, így azt kapjuk, hogy $\mathbb{E}(X^3) = \frac{6}{\lambda^3}$. (Ha ezt nem vesszük észre, akkor ugyanazzal a szereposztással további két parciális integrálással is megkaphatjuk ezt az eredményt.)

Tehát az előbbieket szerint

$$\beta = \frac{\frac{6}{\lambda^3} - \frac{1}{\lambda} \cdot \frac{2}{\lambda^2}}{\frac{1}{\lambda^2}} + 1 = \frac{4}{\lambda} + 1 = \frac{9}{5}$$

és

$$\alpha = \frac{2}{\lambda^2} + \frac{1}{\lambda} + 1 - \beta \cdot \frac{1}{\lambda} = \frac{2}{25} + \frac{1}{5} + 1 - \frac{9}{25} = \frac{23}{25}.$$

Így a regressziós egyenes $\{(x, y) \in \mathbb{R}^2 \mid y = \frac{9}{5}x + \frac{23}{25}\}$.

- b) Ha $X \sim N(0; 1)$, akkor $\mathbb{E}(X) = 0$, $\mathbb{D}^2(X) = 1$, így $\mathbb{E}(X^2) = \mathbb{D}^2(X) + \mathbb{E}(X)^2 = 1$. $\mathbb{E}(X^3)$ pedig szintén nulla, mert $x \mapsto x^3$ páratlan függvény és X sűrűségfüggvénye a 0-ra szimmetrikus [és $\mathbb{E}(X^3)$ valóban létezik, mert a 7. feladatsoron láttuk, hogy még $\mathbb{E}(e^{\alpha X})$ is véges minden $\alpha > 0$ -ra, amiből $\mathbb{E}(X^3) < \infty$ is következik az improprius integrálokra vonatkozó majoráns kritériummal]. ($\mathbb{E}(X^3)$ -öt szintén kiszámolhatjuk parciális integrálással is, $\mathbb{E}(X^2)$ -re visszavezetve.) Ezért

$$\beta = \frac{0 - 0}{\mathbb{D}^2(X)} + 1 = 1, \quad \alpha = 1 + 0 + 1 + 1 \cdot 0 = 2.$$

Tehát a regressziós egyenes b) $\{(x, y) \in \mathbb{R}^2 \mid y = x + 2\}$.

- c*) Tudjuk, hogy ha $X \sim \text{Bin}(n; p)$, akkor $\mathbb{E}(X) = np$ és $\mathbb{D}^2(X) = np(1-p)$, tehát $\mathbb{E}(X^2) = \mathbb{E}(X^2) + \mathbb{E}(X)^2 = np(1-p) + n^2p^2$. Ahhoz, hogy β képletét kiszámoljuk, ismét meg kell határoznunk $\mathbb{E}(X^3)$ -t.

Tudjuk, hogy X felírható n darab független, p paraméterű indikátor összegeként, jelölje ezeket az indikátorokat $X_1, \dots, X_n$. Előadáson láttuk, hogy egy p paraméterű indikátorváltozó minden momentuma p : $p = \mathbb{E}(X) = \mathbb{E}(X^2) = \mathbb{E}(X^3) = \dots$. Így a várható érték linearitása alapján, használva azt is, hogy független valószínűségi változók folytonos függvényekkel való transzformáltjai is függetlenek és független valószínűségi változók szorzatának várható értéke a várható értékek szorzata, azt kapjuk, hogy

$$\begin{aligned} \mathbb{E}(X^3) &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^3\right) = \mathbb{E}\left(\sum_{i=1}^n X_i^3 + \sum_{i=1}^n \sum_{j \neq i} (X_j X_i X_i + X_i X_j X_i + X_i X_i X_j) + \sum_{i=1}^n \sum_{j \neq i} \sum_{k \notin \{i, j\}} X_i X_j X_k\right) \\ &= n \mathbb{E}(X_1^3) + 3n(n-1) \mathbb{E}(X_1^2) \mathbb{E}(X_1) + n(n-1)(n-2) \mathbb{E}(X_1)^3 = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3. \end{aligned}$$

A vizsgált konkrét esetben ($n = 10, p = 3$) a jobb oldal $10 \cdot 0,3 + 3 \cdot 10 \cdot 9 \cdot 0,09 + 10 \cdot 9 \cdot 8 \cdot 0,027 = 46,74$ -gyel egyenlő, így

$$\beta = \frac{46,74 - 11,1 \cdot 3}{2,1} + 1 = 7,4,$$

tehát

$$\alpha = 11,1 + 2,1 + 1 - 7,4 \cdot 3 = -7,1.$$

Azaz a regressziós egyenes $\{(x, y) \in \mathbb{R}^2 \mid y = 7,4x - 7,1\}$.

9. A feladat megadja, hogy $\mathbb{D}^2(X) = 0,0775 = \frac{31}{400}$, és mivel X és Y azonos eloszlásúak (azonosak a marginális sűrűségfüggvényeik), ezért $\mathbb{D}^2(Y)$ is $0,0775$.

a) A regressziós egyenes képletében szereplő paraméterek értéke ezért

$$\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = \frac{-9/400}{31/400} = -\frac{9}{31}$$

és

$$\alpha = \mathbb{E}(Y) - \beta \mathbb{E}(X) = \frac{13}{20} + \frac{9}{31} \cdot \frac{13}{20} = \frac{13 \cdot (31 + 9)}{31 \cdot 20} = \frac{13 \cdot 40}{31 \cdot 20} = \frac{26}{31}.$$

Tehát Y -nak az X -re vett regressziós egyenese $\{(x, y) \in \mathbb{R}^2 \mid y = -\frac{9}{31}x + \frac{26}{31}\}$.

A regressziós egyenes átlagos négyzetes hibája (mivel „additív konstans nem számít” a szórásnégyzetnél sem, azaz a szórásnégyzet is eltolás-invariáns)

$$\mathbb{D}^2(Y - (\beta X + \alpha)) = \mathbb{D}^2(Y - \beta X) = \mathbb{D}^2(Y) + \beta^2 \mathbb{D}^2(X) - 2\beta \text{cov}(X, Y) = \frac{31}{400} \left(1 + \frac{81}{961}\right) - 2 \cdot \left(-\frac{9}{31}\right) \cdot \left(-\frac{9}{400}\right) \approx 0,07097.$$

Ugyanez kiszámolható persze az előadáson elhangzott képlettel is: a hiba

$$\mathbb{D}^2(Y) - \frac{\text{cov}(X, Y)^2}{\mathbb{D}^2(X)} = \frac{31}{400} - \frac{(-9/400)^2}{31/400} \approx 0,07097.$$

b) Y -nak ugyanaz a szórásnégyzete és a várható értéke, mint X -nek. Emiatt β és α ebben az esetben is ugyanaz, mint az a) feladatban. Csak az változik, hogy most az egyenes egyenletét úgy kell felírni, hogy x az y lineáris függvénye: $\{(y, x) \in \mathbb{R}^2 \mid x = -\frac{9}{31}y + \frac{26}{31}\}$. A regressziós egyenes átlagos négyzetes hibája is az előzővel analóg módon számolható ki, ezért $0,07097$ -tel egyenlő.

10. Mivel Y regressziós egyenese X -re $\{(x, y) \in \mathbb{R}^2 \mid y = x - 2\}$, ezért a regressziós egyenes képlete alapján $\frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = 1$ és $\mathbb{E}(Y) - \underbrace{\frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)}}_{=1} \mathbb{E}(X) = -2$. Továbbá mivel X regressziós egyenese Y -ra

$\{(y, x) \in \mathbb{R}^2 \mid x = \frac{1}{4}y - \frac{11}{2}\}$, ezért a regressziós egyenes képlete szerint $\frac{\text{cov}(X, Y)}{\mathbb{D}^2(Y)} = \frac{1}{4}$ és

$\mathbb{E}(X) - \underbrace{\frac{\text{cov}(X, Y)}{\mathbb{D}^2(Y)}}_{=\frac{1}{4}} \mathbb{E}(Y) = -\frac{11}{2}$. Ebből az alábbi kétismeretlenes, két egyenletből álló lineáris egyen-

letrendszer adódik:

$$\begin{aligned} \mathbb{E}(Y) - \mathbb{E}(X) &= -2, \\ \mathbb{E}(X) - \frac{1}{4} \mathbb{E}(Y) &= -\frac{11}{2}. \end{aligned}$$

Ezt megoldhatjuk pl. Gauss-eliminációval is, de talán egyszerűbb a szokásos „középiskolás módszerrel”: az első egyenletből $\mathbb{E}(Y) = \mathbb{E}(X) - 2$, és így a másodikból

$$\mathbb{E}(X) - \frac{1}{4}(\mathbb{E}(X) - 2) = -\frac{11}{2},$$

vagyis

$$\frac{3}{4} \mathbb{E}(X) = -6,$$

tehát

$$\mathbb{E}(X) = -8 \quad \text{és} \quad \mathbb{E}(Y) = -10.$$

11. A regressziós egyenesekhez X és Y várható értékének és szórásnégyzetének, valamint a kettőjük kovarianciájának a meghatározása szükséges. Vegyük észre, hogy X és Y értékkészlete is $\{0, 1, 2\}$, és Y értéke mindig legalább annyi, mint X -é (mert minden hatos dobás egyben páros is). Így $\mathbb{E}(XY)$ -t könnyen kiszámíthatjuk (ignorálva azokat az eseteket, amikor $XY = 0$, vagyis amikor $X = 0$):

$$\mathbb{E}(XY) = 2 \cdot 2 \cdot \mathbb{P}(X = Y = 2) + 1 \cdot 2 \cdot \mathbb{P}(X = 1, Y = 2) + 1 \cdot 1 \cdot \mathbb{P}(X = Y = 1).$$

Az $\{X = Y = 2\}$ esemény pontosan akkor következik be, ha két hatost dobunk, ennek a valószínűsége $\frac{1}{36}$. Az $\{X = 1, Y = 2\}$ esemény pontosan akkor következik be, ha egy hatost és egy nem hatos párosat dobunk, vagyis vagy az első dobás hatos és a második kettő vagy négyes, vagy fordítva. Ez négy kedvező eset, tehát $\mathbb{P}(X = 1, Y = 2) = \frac{4}{36} = \frac{1}{9}$. Végül az $\{X = Y = 1\}$ esemény akkor következik be, ha csak egy hatost dobunk és egy páratlant, ami hatféleképpen következhet be: vagy az első dobás hatos és a második egyes, hármas vagy ötös, vagy fordítva, ezért $\mathbb{P}(X = Y = 1) = \frac{6}{36} = \frac{1}{6}$. Tehát

$$\mathbb{E}(XY) = 4 \cdot \frac{1}{36} + 2 \cdot \frac{4}{36} + 1 \cdot \frac{6}{36} = \frac{18}{36} = \frac{1}{2}.$$

A két dobás függetlensége miatt X binomiális eloszlású $n = 2$ és $p = \frac{1}{6}$ paraméterekkel, tehát $\mathbb{E}(X) = np = \frac{1}{3}$ és $\mathbb{D}^2(X) = np(1-p) = \frac{5}{18}$. Hasonlóan Y binomiális eloszlású $n = 2$ és $p = \frac{3}{6} = \frac{1}{2}$ paraméterekkel, tehát $\mathbb{E}(Y) = np = 1$ és $\mathbb{D}^2(Y) = \frac{1}{2}$. Ezeket a várható értékeket és szórásnégyzeteket persze elemi módszerekkel is kiszámíthatjuk. Így tehát $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{2} - \frac{1}{3} \cdot 1 = \frac{1}{6}$, és így $\frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = \frac{\frac{1}{6}}{\frac{5}{18}} = \frac{3}{5}$, tehát $\mathbb{E}(Y) - \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} \mathbb{E}(X) = 1 - \frac{3}{5} \cdot \frac{1}{3} = \frac{4}{5}$. Vagyis Y regressziós egyenese X -re $\{(x, y) \in \mathbb{R}^2 \mid y = \frac{3}{5}x + \frac{4}{5}\}$.

Hasonlóan $\frac{\text{cov}(X, Y)}{\mathbb{D}^2(Y)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$, ezért $\mathbb{E}(X) - \frac{\text{cov}(X, Y)}{\mathbb{D}^2(Y)} \mathbb{E}(Y) = \frac{1}{3} - \frac{1}{3} \cdot 1 = \frac{1}{3} - \frac{1}{3} = 0$. Így X regressziós egyenese Y -ra $\{(y, x) \in \mathbb{R}^2 \mid x = \frac{y}{3}\}$.

(Mivel itt X -nek az Y -ra vonatkozó regressziós egyeneséről van szó, itt x -et írjuk y lineáris függvényeként és (y, x) pontokat tekintünk, ahogy a 10-es feladatban is tettük hasonló helyzetben.)