# Valószínűségszámítás 1. vizsga megoldókulcs

1. A $13, 20, 55, 78, 90$ számokat játszottuk meg a lottón. Legyen $A$ az az esemény, hogy az elsőként kihúzott két számot el fogjuk találni, $B$ pedig az, hogy legalább három találatunk lesz. $\mathbf{P}(A+B) = ?$

   *Megoldás*

   $\mathbf{P}(A+B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$ (5 pont)

   Annak a valószínűsége, hogy az első kettő kihúzott számot eltaláljuk (5 pont):

   $$\mathbf{P}(A) = \frac{5 \cdot 4 \cdot 88 \cdot 87 \cdot 86}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{2}{801} \approx 0.002496$$

   Annak a valószínűsége, hogy 3, 4 vagy 5 találatunk lesz (5 pont):

   $$\mathbf{P}(B) = \frac{\binom{5}{3}\binom{85}{2} + \binom{5}{4}\binom{85}{1} + \binom{5}{5}}{\binom{90}{5}} = \frac{2007}{2441626} \approx 0.00082199$$

   Annak a valószínűsége, hogy az első kettő kihúzott számot eltaláljuk, és ezen felül még lesz legalább egy találatunk (5 pont):

   $$\mathbf{P}(AB) = \frac{5 \cdot 4 \cdot (3 \cdot 3 \cdot 85 \cdot 84 + 3 \cdot 3 \cdot 2 \cdot 85 + 3 \cdot 2 \cdot 1)}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{5483}{21974634} \approx 0.00024951$$

   $$\mathbf{P}(A+B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB) \approx 0.00307$$

2. Egy dobozban három piros, két fehér és egy zöld golyó van. Visszatevéssel tízszer húzunk a dobozból. Jelölje $X$ a pirosak számát! Adja meg a $Z = (X+2)(X-2)$ várható értékét!

   *Megoldás*

   $X \in B(10, \frac{1}{2})$, $n = 10$, $p = \frac{1}{2}$ (10 pont)

   $\mathbf{E}Z = \mathbf{E}((X+2)(X-2)) = \mathbf{E}(X^2 - 4) = \mathbf{E}X^2 - 4$ (3 pont)

   $\mathbf{E}X^2 = \sigma^2 X + (\mathbf{E}X)^2 = np(1-p) + n^2p^2 = \frac{55}{2}$ (5 pont)

   $\mathbf{E}Z = 23.5$ (2 pont)

3. Az $X$ és $Y$ együttes sűrűségfüggvénye

   $$f_{X,Y}(x,y) = \begin{cases} a\left(x^2 + 2xy + 3y^2\right), & \text{ha } 0 < x, y < 1 \\ 0, & \text{egyébként} \end{cases}.$$

   a.) Mennyi az $a$ értéke?
   b.) Független-e $X$ és $Y$ ?
   c.) $\operatorname{cov}(X,Y) = ?$

   *Megoldás*

a.)

$$\int_0^1 \int_0^1 a\left(x^2 + 2\,xy + 3\,y^2\right) dx\, dy = 1 \Rightarrow a = \frac{6}{11} \text{ (4 pont)}$$

b.)

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \frac{6}{11}\,x^2 + \frac{6}{11}\,x + \frac{6}{11} \text{ (2 pont)}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = \frac{2}{11} + \frac{6}{11}\,y + \frac{18}{11}\,y^2 \text{ (2 pont)}$$

$$f_{X,Y}(x,y) = \frac{6}{11}\,x^2 + \frac{12}{11}\,xy + \frac{18}{11}\,y^2$$

Mivel $f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y)$ ezért $X$ és $Y$ nem függetlenek. (2 pont)

c.) $\operatorname{cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y$ (2 pont)

$$\mathbf{E}(XY) = \int_0^1 \int_0^1 xy \cdot f_{X,Y}(x,y)dx\,dy = \frac{13}{33} \text{ (2 pont)}$$

$$\mathbf{E}X = \int_0^1 x \cdot f_X(x)dx = \frac{13}{22} \text{ (2 pont)}$$

$$\mathbf{E}Y = \int_0^1 y \cdot f_Y(y)dy = \frac{15}{22} \text{ (2 pont)}$$

$$\operatorname{cov}(X,Y) = -\frac{13}{1452} \approx -0.0089532 \text{ (2pont)}$$

4. Dobjunk tízszer egy szabályos dobókockával! Jelölje $X$ a hatos, $Y$ pedig a hárommal oszthatό számok dobásainak számát! Számolja ki az $\mathbf{E}(Y \mid X)$ regressziót!

   *Megoldás*

   a.)

$$r(t) = \mathbf{E}(Y|X=t) = \sum_{\forall i} i \cdot \mathbf{P}(Y=i|X=t) \text{ (2 pont)}$$

Arra a valószínűségre vagyunk kíváncsiak, hogy $10-t$-szer dobva milyen valószínűséggel kapunk $i-t$ darab hárommal osztható számot az $\{1,2,3,4,5\}$ számok közül:

$$\mathbf{P}(Y=i|X=t) = \binom{10-t}{i-t}\left(\frac{1}{5}\right)^{i-t}\left(\frac{4}{5}\right)^{10-t-(i-t)} \quad \text{(5 pont)}$$

Behelyettesítve:

$$\begin{aligned} r(t) &= \sum_{i=t}^{10} i \cdot \binom{10-t}{i-t}\left(\frac{1}{5}\right)^{i-t}\left(\frac{4}{5}\right)^{10-t-(i-t)} \\ &= \sum_{i=t}^{10} (i-t)\binom{10-t}{i-t}\left(\frac{1}{5}\right)^{i-t}\left(\frac{4}{5}\right)^{10-t-(i-t)} + \sum_{i=t}^{10} t\binom{10-t}{i-t}\left(\frac{1}{5}\right)^{i-t}\left(\frac{4}{5}\right)^{10-t-(i-t)} \quad \text{(5 pont)} \end{aligned}$$

Elvégezve a $k = i - t$ behelyettesítést:

$$r(t) = \sum_{k=0}^{10-t} k \binom{10-t}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-t-k} + t \sum_{k=0}^{10-t} \binom{10-t}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-t-k} \quad \text{(3 pont)}$$

Az első szumma után egy $B(10-t, \frac{1}{5})$ valószínűségi változó várható értéke áll, a második után pedig a biztos esemény valószínűsége.

$$r(t) = (10-t)\frac{1}{5} + t \cdot 1 = 2 + \frac{4}{5}t \text{ (3 pont)}$$

$$\mathbf{E}(Y|X) = r(X) = 2 + \frac{4}{5}X \text{ (2 pont)}$$

b.) Legyen $Z = Y - X$. $Z$ éppen a 3-asok száma azon dobások között, amikor nem 6-ost dobtunk. Így $Z|X = t \in B(10-t, \frac{1}{5})$ eloszlású, azaz $\mathbf{E}(Z|X = t) = (10-t)\frac{1}{5}$. Tehát

$$\mathbf{E}(Y|X=t) = \mathbf{E}(Z+X|X=t) = \mathbf{E}(Z|X=t) + \mathbf{E}(X|X=t) = (10-t)\frac{1}{5} + t = 2 + \frac{4}{5}t,$$

azaz

$$\mathbf{E}(Y|X) = 2 + \frac{4}{5}X.$$

5. Igazolja, hogy $n \geq 50000$ megfigyelés esetén egy 2-nél nem nagyobb szórású valószínűségi változó értékeinek átlaga legalább 80%-os valószínűséggel a várható érték $0.02$ sugarú környezetébe esik!

   *Megoldás*

   a.) A Csebisev-egyenlőtlenségből, mivel $\sigma^2 \bar{X}_n = \frac{\sigma^2}{n}$ (5 pont):

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| < 0.02\right) \geq 1 - \frac{\sigma^2}{n \cdot (0.02)^2} \text{ (10 pont)} \geq 1 - \frac{4}{50000 \cdot \frac{4}{10000}} = 0.8 \text{ (5 pont)}$$

   b.) Legyen $\mathbf{E}X_i = \mu$, $\sigma X_i = \sigma$. A kérdés az, hogy

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| < 0.02\right) \overset{?}{\gtrsim} 0.8$$

   A centrális határeloszlás tételből

$$\mathbf{P}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} < x\right) \approx \Phi(x) \text{ (5 pont)}$$

$$\mathbf{P}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} < x\right) = \mathbf{P}\left(\sum_{i=1}^n X_i - n\mu < x\sqrt{n}\sigma\right) = \mathbf{P}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu < \frac{\sigma x}{\sqrt{n}}\right) \approx \Phi(x)$$

   A kérdéses valószínűség:

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| < 0.02\right) = \mathbf{P}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu < 0.02\right) - \mathbf{P}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu < -0.02\right) = (*) \text{ (5 pont)}$$

Mivel $0.02 = \frac{\sigma x}{\sqrt{n}}$:

$$(*) \approx \Phi\left(\frac{0.02\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-0.02\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{0.02\sqrt{n}}{\sigma}\right) - 1 \geq \text{(5 pont) } 2\Phi\left(\frac{0.02\sqrt{50000}}{2}\right) - 1 \geq$$

$$\geq 2\Phi\left(2.2\right) - 1 \text{ (5 pont)} \approx 2 \cdot 0.986 - 1 = 0.972 \geq 0.8$$

Megjegyzés: A centrális határeloszlás-tételen alapuló bizonyítás nem teljesen korrekt, hiszen nem vizsgáltuk, hogy a közelítés mennyire pontos. Ez egyébként nem is olyan egyszerű, mert a pontosságra vonatkozó standard korlátok (az ún. Berry-Esseen-tétel és annak különböző javított változatai) többnyire a 3. (centrális) momentumot használják, amely esetünkben akár végtelen is lehet.