11. gyakorlat

Független normálisok összegének tulajdonságai, Csebisev-egyenlőtlenség.
Statisztikai alapfogalmak, rendezett minták, empirikus eloszlásfüggvény

Végeredmények és megoldás(vázlat)ok

1. (a) (6 pont)
(1+1 pont) Sztenderdizálunk (ezt nem muszáj szövegesen odaírni):

$$\mathbb{P}(X > 0{,}5) = \mathbb{P}(\frac{X - \mathbb{E}(X)}{\mathbb{D}(X)} > \frac{0{,}5 - \mathbb{E}(X)}{\mathbb{D}(X)} = \mathbb{P}(\frac{X-1}{3} > \frac{0{,}5-1}{3})$$

(a két lépés összevonható, ha a második jó)
(1 pont) $= \mathbb{P}(Z > -1/6)$, ahol $Z \sim \mathrm{N}(0;1)$ (az 1/6 helyett 4 értékes jegyre kerekített numerikus eredmény is jó)
(1 pont) $= \mathbb{P}(Z < 1/6) = \Phi(1/6)$ (elég a kettő közül az egyik, és nem kell a folytonosságra hivatkozni, mert a sűrűségfüggvény szimmetriája miatt a két integrál ugyanaz)
VAGY: $= 1 - \mathbb{P}(Z \leq -1/6) = 1 - \Phi(-1/6) = 1 - (1 - \Phi(1/6)) = \Phi(1/6)$ (a Φ jelölés nem kötelező, és a folytonosságra itt sem jár pont, mert az egészre összesen 1 pont jár, emiatt bármelyik lépésben szabad váltani a $\leq$ és a $<$ között)
(1 pont) $\approx \Phi(0{,}17)$
(1 pont) $\approx 0{,}5675$.
(Ne vonjunk le pontot, ha csak az a baj, hogy valamelyik $\approx$ helyett $=$-t ír. Az utolsó két lépés összevonható, ha előtte az 1/6 már megjelent ÉS jó a kiolvasott végeredmény.)
(b) (11 pont)
(2 pont) $\mathrm{cov}(X+2Y, X-4Y) = \underbrace{\mathrm{cov}(X,X)}_{=\mathbb{D}^2(X)} + 2\mathrm{cov}(X,Y) - 4\mathrm{cov}(X,Y) - 8\underbrace{\mathrm{cov}(Y,Y)}_{=\mathbb{D}^2(Y)}$.
(Ebből 1 pont, ha a bilinearitást jól használja, ehhez elég 4 helyes tagra bontani az összeget további egyszerűsítések nélkül. Még 1 pont, ha a konstans szorzókat is jól emeli ki, a szimmetriát jól használja és tudja, hogy egy val. változó önmagával vett kovarianciája a szórásnégyzete.
(1 pont) $\mathrm{cov}(X,Y) = 0$,
(1 pont) mert függetlenek.
(1 pont) Ezért (mivel a szórásnégyzet a normális eloszlás második paramétere – erre nem jár pont, mert benne van a nevezetes eloszlás táblázatban) $\mathrm{cov}(X+2Y, X-4Y) = 9 - 8 \cdot 16 = -119$.
(Ha minden szükséges részeredmény szerepelt már, akkor itt elég a végeredmény.)
(0 pont) A korreláció meghatározásához most kiszámítjuk $X+2Y$ és $X-4Y$ szórásait is.
(1 pont) Mivel X és Y függetlenek,
(1 pont) $\mathbb{D}^2(X+2Y) = \mathbb{D}^2(X) + \mathbb{D}^2(2Y)$
(1 pont) $= \mathbb{D}^2(X) + 4\,\mathbb{D}^2(Y) = 9 + 64 = 73$ (Ha a sor eleji lépés megvan, akkor ez a pont az előzővel összevonható, de ha a végeredmény nem jó, akkor csak az előző pont jár.
Ha pedig az előző lépés végéig megvannak a részeredmények és a végeredmény jó, akkor higgyük el, hogy így számol, de rossz végeredmény esetén ez az egy pont nem jár).
(0 pont) Hasonlóan (itt már nem muszáj a függetlenségre hivatkozni)
(1 pont) $\mathbb{D}^2(X-4Y) = \mathbb{D}^2(X) + \mathbb{D}^2(-4Y)$
(1 pont) $= \mathbb{D}^2(X) + 16\,\mathbb{D}^2(Y) = 9 + 16 \cdot 16 = 265$ (ha minden korábbi részletszámolás jó, akkor itt már elég a végeredmény).
(0 pont) arra, hogy a szórás a szórásnégyzet négyzetgyöke. Erre már a 2. (b)-ben adtunk pontot. Aki ott nem kapott rá pontot és itt leírta ezt, utólag kérhet a 2. feladatra 1 pontot.
(1 pont) Tehát $\mathrm{corr}(X+2Y, X-4Y) = \frac{\mathrm{cov}(X+2Y, X-4Y)}{\mathbb{D}(X+2Y)\,\mathbb{D}(X-4Y)} = \frac{-119}{\sqrt{73}\cdot\sqrt{265}} \approx -0{,}8556$ (ha már minden részeredmény megvan, akkor a korreláció általános képlete után elég a numerikus végeredmény, akkor is OK, ha csak a szórásnégyzetek vannak meg és nem a szórások. Az általános képlet helyett pedig elég a konkrét esetbeli képlet, ha azonosítható a benne szereplő számok szerepe).
(c) (3 pont)
(1 pont) Független normálisok összege normális és a paraméterek összeadódnak. Normális lineáris transzformáltja normális (vagy konstans – ezt nem kell odaírni).
Ezekre összesen 1 pont jár. Ezek általános paraméterekkel is kifejezhetőek, de ez a pont csak akkor

jár, ha hivatkozik a függetlenségre is. Ha nem írja le, hogy a paraméterek összeadódnak, de később összeadja őket, akkor ez a pont is jár. Ha nem hivatkozik a lineáris transzformációra, de kimondja, hogy $2X$ normális eloszlású és minden más rendben van, akkor is adjuk meg ezt az 1 pontot.
(1 pont) Így $2X \sim \mathrm{N}(2; 36)$, mivel az eloszlás első paramétere $\mathbb{E}(2X) = 2\,\mathbb{E}(X) = 2$, a második $\mathbb{D}^2(2X) = 4\,\mathbb{D}^2(X) = 36$ és
(indoklás nélkül ez az 1 pont nem jár, de ha annyit odaír, hogy az első paraméter a várható érték és a második a szórás, és jó numerikus eredményt ad, akkor fogadjuk el ezért az 1 pontért a megoldást)
(1 pont) $2X + Y \sim \mathrm{N}(4; 52)$.
Az, hogy $2X$ milyen eloszlású, nem szükséges, amennyiben szerepel, hogy normális lineáris transzformáltja normális; ebben az esetben elég a $2X$ várható értékére és szórásnégyzetére vonatkozó indoklás (ami összevonható ugyanezzel a $2X + Y$-ra vonatkozóan, ha részletesen indokol).

2. Független normális eloszlású valószínűségi változók összege normális eloszlású, és a paraméterek összeadódnak. Ezért $X_1 + X_2 + X_3 \sim \mathrm{N}(3\mu, 3\sigma^2)$. (Megjegyzés: könnyen ellenőrizhetjük a várható érték linearitása alapján, hogy $\mathbb{E}(X_1 + X_2 + X_3) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = 3\mu$, és mivel függetlenek, $\mathbb{D}^2(X_1 + X_2 + X_3) = \mathbb{D}^2(X_1) + \mathbb{D}^2(X_2) + \mathbb{D}^2(X_3) = 3\sigma^2$. A fenti első mondatban ehhez képest csak az az újdonság, hogy az összeg is normális eloszlású.)
Továbbá a várható érték linearitása miatt $\mathbb{E}(X_1 + X_2 + X_3 - 3\mu) = \mathbb{E}(X_1 + X_2 + X_3) - 3\mu = 0$, vagyis $\mathbb{E}(\frac{X_1+X_2+X_3-3\mu}{\sqrt{3}\sigma})$ szintén 0. A szórásnégyzet eltolás-invariáns (additív konstans nem számít), tehát $\mathbb{D}^2(X_1 + X_2 + X_3 - 3\mu) = \mathbb{D}^2(X_1 + X_2 + X_3) = 3\sigma^2$, továbbá $\mathbb{D}^2(\frac{X_1+X_2+X_3-3\mu}{\sqrt{3}\sigma}) = \frac{1}{3\sigma^2}\,\mathbb{D}^2(X_1 + X_2 + X_3 - 3\mu) = 1$.
Így tehát, mivel normális eloszlású val. változó lineáris transzformáltjai is normálisak, ezért

$$\frac{X_1 + X_2 + X_3 - 3\mu}{\sqrt{3}\sigma} \sim \mathrm{N}(0; 1).$$

Továbbá

$$\mathrm{corr}(X_1, Y) = \frac{\mathrm{cov}(X_1, Y)}{\mathbb{D}(X_1)\,\mathbb{D}(Y)} = \frac{\mathrm{cov}(X_1, Y)}{\sigma \cdot 1}.$$

A kovariancia azonosságai (bilinearitás, eltolás-invariancia/additív konstans nem számít) alapján

$$\mathrm{cov}(X_1, Y) = \mathrm{cov}(X_1, \frac{X_1 + X_2 + X_3 - 3\mu}{\sqrt{3}\sigma}) = \mathrm{cov}(X_1, \frac{X_1 + X_2 + X_3}{\sqrt{3}\sigma}) = \frac{1}{\sqrt{3}\sigma}(\mathrm{cov}(X_1, X_1) + \mathrm{cov}(X_1, X_2) + \mathrm{cov}(X_1, X_3)).$$

Itt $\mathrm{cov}(X_1, X_1) = \mathbb{D}^2(X_1) = \sigma^2$, a másik két kovariancia pedig 0, mert X_1 független X_2-től és X_3-tól is. Így $\mathrm{cov}(X_1, Y) = \frac{\sigma}{\sqrt{3}}$, a végeredmény pedig $\mathrm{corr}(X_1, Y) = \frac{1}{\sqrt{3}} \approx 0{,}5774$.

3. a) X és Y függetlenek, így $\mathrm{cov}(X, Y) = 0$. Továbbá X és Y standard normális eloszlásúak, így $\mathbb{D}^2(X) = \mathbb{D}^2(Y) = 1$. Így a kovarianciamátrixuk $\begin{pmatrix} \mathbb{D}^2(X) & \mathrm{cov}(X, Y) \\ \mathrm{cov}(X, Y) & \mathbb{D}^2(Y) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
b) A várhatóérték-vektor $\underline{m} = \begin{pmatrix} \mathbb{E}(U) \\ \mathbb{E}(V) \end{pmatrix}$, $\underline{\underline{\Sigma}}$ pedig az előbbi kovarianciamátrixhoz hasonlóan adódik, csak (X, Y) helyett (U, V)-vel.
A várható érték linearitása miatt $\mathbb{E}(U) = \mathbb{E}(3X - 2Y + 1) = 3\,\mathbb{E}(X) - 2\,\mathbb{E}(Y) + 1 = 3 \cdot 0 + 2 \cdot 0 + 1 = 1$ és $\mathbb{E}(V) = \mathbb{E}(7X - 5Y - 3) = 7\,\mathbb{E}(X) - 5\,\mathbb{E}(Y) - 3 = -3$ (mivel X és Y standard normális eloszlásúak, tehát a várható értékük 0).
A kovariancia azonosságai (bilinearitás, eltolás-invariancia/„additív konstans nem számít”) és az a) feladatban kiszámoltak alapján

$$\begin{aligned}\mathrm{cov}(U, V) &= \mathrm{cov}(3X - 2Y + 1, 7X - 5Y - 3) = \mathrm{cov}(3X - 2Y, 7X - 5Y) \\ &= 3 \cdot 7 \cdot \mathrm{cov}(X, X) - 2 \cdot 7 \cdot \mathrm{cov}(Y, X) - 3 \cdot 5 \cdot \mathrm{cov}(Y, X) + 2 \cdot 5 \cdot \mathrm{cov}(Y, Y) \\ &= 21\,\mathbb{D}^2(X) - 0 - 0 + 10\,\mathbb{D}^2(Y) = 31,\end{aligned}$$

valamint mivel független valószínűségi változók összegének szórásnégyzete a szórásnégyzeteik összege, a szórásnégyzet azonosságait használva

$$\mathbb{D}^2(U) = \mathbb{D}^2(3X - 2Y + 1) = \mathbb{D}^2(3X + (-2Y)) = 9\,\mathbb{D}^2(X) + 4\,\mathbb{D}^2(Y) = 9 + 4 = 13$$

és

$$\mathbb{D}^2(V) = \mathbb{D}^2(7X - 5Y - 3) = \mathbb{D}^2(7X + (-5Y)) = 49\,\mathbb{D}^2(X) + 25\,\mathbb{D}^2(Y) = 49 + 25 = 74.$$

Így a kovarianciamátrix $\begin{pmatrix} 13 & 31 \\ 31 & 74 \end{pmatrix}$.

c) Ismét a kovariancia bilinearitását használva

$$\begin{aligned}\mathrm{cov}(-2U + V, 12U - 5V) &= -2 \cdot 12\mathrm{cov}(U, U) + 2 \cdot 5 \cdot \mathrm{cov}(U, V) + 12\mathrm{cov}(U, V) - 5\mathrm{cov}(V, V) \\ &= 22 \cdot 31 - 24\,\mathbb{D}^2(U) - 5\,\mathbb{D}^2(V) = 682 - 24 \cdot 13 - 5 \cdot 74 = 682 - 312 - 370 = 0.\end{aligned}$$

d) Igen, mivel $-2U + V$ és $12U - 5V$ is az X és Y független normális eloszlású valószínűségi változók lineáris kombinációi (ugyanis U és V definíciója szerint $-2U + V = -2(3X - 2Y + 1) + 7X - 5Y - 3 = X - Y - 5$ és $12U - 5V = 12(3X - 2Y + 1) - 5(7X - 5Y - 3) = X + Y + 27$) és a c) alapján korrelálatlanok, ezért függetlenek is.

4. (a) (19 pont)
(1 pont) Tudja, hogy a várhatóérték-vektor koordinátái $\mathbb{E}(U)$ és $\mathbb{E}(V)$ a koordinátái a konkrét val. vektorváltozó esetén.
(1 pont) $\mathbb{E}(U) = \mathbb{E}(X) + \mathbb{E}(Y)$
(1 pont) $= 1 - 2 = -1$ (itt már elég a végeredmény).
(1 pont) $\mathbb{E}(V) = \mathbb{E}(X) - 2\,\mathbb{E}(Y)$
(1 pont) $= 1 - 2 \cdot (-2) = 5$.
(0 pont) Tehát $\underline{\mu} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$.
(1 pont) $\mathbb{D}^2(U) = \mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$
(1 pont) $= 4 + 1 = 5$,
(1 pont) mert függetlenek.
(0 pont) Hasonlóan a függetlenség miatt
(1 pont) $\mathbb{D}^2(V) = \mathbb{D}^2(X - 2Y) = \mathbb{D}^2(X) + \mathbb{D}^2(-2Y)$
(1 pont) $= \mathbb{D}^2(X) + 4\,\mathbb{D}^2(Y) = 8$ (az előzővel összevonható)
(1 pont) (Mivel a kovariancia bilineáris – ezt nem kell odaírni) $\mathrm{cov}(U, V) = \mathrm{cov}(X + Y, X - 2Y) = \mathrm{cov}(X, X) + \mathrm{cov}(X, -2Y) + \mathrm{cov}(Y, X) + \mathrm{cov}(Y, -2Y)$
(1 pont) $= \mathrm{cov}(X, X) - 2\mathrm{cov}(X, Y) + \mathrm{cov}(Y, X) - 2\mathrm{cov}(Y, Y)$ (az előzővel összevonható, indoklás nélkül is).
(1 pont) $\mathrm{cov}(X, X) = \mathbb{D}^2(X)$
(1 pont) $= 4$.
(1 pont) Hasonlóan $\mathrm{cov}(Y, Y) = \mathbb{D}^2(Y) = 1$ (itt már elég a végeredmény és nem jár pont az indoklásra).
(1 pont) $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X) = 0$,
(1 pont) mert függetlenek.
(1 pont) Tehát $\mathrm{cov}(U, V) = 2$.
(1 pont) Tehát a kovarianciamátrix $\underline{\underline{\Sigma}} = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$.
(b) (1 pont) (1 pont) Mivel $\mathrm{cov}(U, V) \neq 0$, ezért U és V nem függetlenek.
(c) (3 pont)
(1+1+1 pont) A regressziós egyenes $\{(u, v) \in \mathbb{R}^2 \colon v = \beta u + \alpha\}$, ahol $\beta = \frac{\mathrm{cov}(U,V)}{\mathbb{D}^2(U)} = \frac{2}{5} (= 0{,}4)$ és $\alpha = \mathbb{E}(V) - \beta\,\mathbb{E}(U) = 5 - \frac{2}{5} \cdot (-1) = \frac{27}{5} (= 5{,}4)$.
Ha a képleteket tudja, de nem vagy rosszul helyettesíti be, akkor max. 2 pont. Nem baj, ha csak az egyenes egyenlete szerepel és nem a ponthalmaz. Viszont ha csak az egyenes képlete van anélkül, hogy β és α meg lenne adva, vagy mindkét paraméter rosszul van megadva, vagy nem derül ki, hogy honnan jön ki, akkor 0 pont. (u, v) helyett (x, y)-nal is jó természetesen.
(d) (6 pont)
(1 pont) $\mathbb{E}(U^2(2V + 1)|U) = U^2\,\mathbb{E}(2V + 1|U)$,
(1 pont) mert... (valahogy hivatkozik a regresszió megfelelő tulajdonságára, pl. ismert információ kiemelése, taking out what is known stb.)

(1 pont) $\mathbb{E}(2V+1|U) = 2\,\mathbb{E}(V|U)+1$ (nem kell indokolni, de nyilván a feltételes várható érték/regresszió linearitása miatt).
(1 pont) Végül $\mathbb{E}(V|U) = \frac{2}{5}U + \frac{27}{5}$,
(1 pont) mert független normálisok összegei/lineáris kombinációi (itt: $U = X + Y$, $V = X - 2Y$ – ezt nem kell odaírni) esetén $\mathbb{E}(V|U)$ megegyezik V-nek U-ra vett lineáris regressziójával (elég: a regresszió a lineáris regresszió, viszont ha nem említi a normalitást és függetlenséget, akkor ez a pont nem jár.
(1 pont) Összefoglalva tehát: $\mathbb{E}(U^2(2V+1)|U) = U^2(2 \cdot (\frac{2}{5}U + \frac{27}{5}) + 1)$ (nem kell ennél szebb alakra hozni, de lehet: $\frac{4}{5}U^3 + \frac{59}{5}U^2$).

5.
$$\mathbb{P}(X < Y) = \mathbb{P}(X - Y < 0) = \mathbb{P}(X + (-Y) < 0).$$

Mivel $Y \sim \mathrm{N}(4;3)$, ezért $-Y \sim \mathrm{N}(-4;3)$ (mert normális eloszlású val. változó lineáris transzformáltja normális és $\mathbb{E}(-Y) = -\mathbb{E}(Y)$, $\mathbb{D}^2(-Y) = \mathbb{D}^2(Y)$). Tehát mivel X és Y függetlenek, X és $-Y$ is függetlenek, $X-Y = X+(-Y)$ is normális eloszlású, az összegük is normális eloszlású és a paramétereik összeadódnak (lásd 8. feladat), vagyis

$$X - Y \sim \mathrm{N}(1;5).$$

Így $X - Y$-t a szokásos módon standardizálva és a normális eloszlás táblázatát használva azt kapjuk, hogy

$$\mathbb{P}(X < Y) = \mathbb{P}(X-Y < 0) = \mathbb{P}\,(\underbrace{\frac{X-Y-1}{\sqrt{5}}}_{\sim \mathrm{N}(0;1)} < \frac{0-1}{\sqrt{5}}) = \Phi(-\frac{1}{\sqrt{5}}) = 1-\underbrace{\Phi(1/\sqrt{5})}_{\approx 0{,}4472} \approx 1-\Phi(0{,}45) \approx 1-0{,}6736 = 0{,}3264.$$

---

6. Jelölje X az egy időegység alatti lekérések számát. Vegyük észre, hogy

$$\mathbb{P}(40 < X < 60) = 1 - \mathbb{P}(X \leq 40 \text{ vagy } X \geq 60) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| \geq 10).$$

A feladat szövege alapján $\mathbb{E}(X) = 50$ és $\mathbb{D}(X) = 5$, tehát $\mathbb{D}^2(X) = 25$. Ezért Csebisev-egyenlőtlenséggel

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq 10) \leq \frac{\mathbb{D}^2(X)}{10^2} \leq \frac{25}{100} = \frac{1}{4}.$$

Így tehát $\mathbb{P}(40 < X < 60) \geq \dfrac{3}{4}$.

7. a) Tudjuk, hogy egy kockadobás eredményének várható értéke 3,5, szórásnégyzete pedig 35/12. Így ha $X_1, \ldots, X_{10}$ független kockadobások eredményei, akkor $\mathbb{E}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{E}(X_i) = 35$ és $\mathbb{D}^2(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{D}^2(X_i) = 10\,\mathbb{D}^2(X_1) = \frac{350}{12}$. Legyen $X = \sum_{i=1}^{10} X_i$.

a) Csebisev-egyenlőtlenséggel

$$\mathbb{P}(24 < X < 46) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| \geq 11) \geq 1 - \frac{\mathbb{D}^2(X)}{11^2} = 1 - \frac{350/12}{11^2} \approx 1 - 0{,}2410 = 0{,}7590,$$

ami nem túl jó becslés, mivel kb. 0,2-vel elmarad a CHT által adott, igen pontos közelítéssel helyes értéktől.

b) Ugyanígy számolva Csebisevvel

$$\mathbb{P}(31 < X < 39) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| \geq 4) \geq 1 - \frac{\mathbb{D}^2(X)}{4^2} = 1 - \frac{350/12}{4^2} \approx 1 - 1{,}8229 = -0{,}8229,$$

tehát triviális (negatív) alsó korlátot kapunk a vizsgált valószínűségre.

---

8. $\overline{x_n} = \overline{x}_{13} = 3$ (a mintaelemek számtani közepe); $s_n^* = s_{13}^* = \sqrt{\frac{11}{6}} \approx 1.3540$ (az előadáson elhangzott/képletgyűjteményben szereplő $(s_n^*)^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x_n})^2$ képlet szerint, s_n^* ennek a négyzetgyöke); az empirikus medián $x_7^* = 3$ (mivel páratlan sok elem van, definíció szerint az empirikus medián a mintaelemeket növekvő sorrendbe rendezve a középső elem); módusz a 2, a 3 és a 4 is (ezek a leggyakoribb elemek, holtversenyben); $x_8^* = 3$ (a mintaelemeket növekvő sorrendbe rendezve ez a 8. elem); az empirikus eloszlásfüggvény $x \in \mathbb{R}$ pontbeli $F_n^*(x)$ értéke definíció szerint mindig az x-nél kisebb mintaelemek száma leosztva n-nel, vagyis az összes mintaelem számával, esetünkben tehát $F_{12}^*(5) = \dfrac{11}{13}$.

9. a) Az átlagfizetés (a mintaelemek számtani közepe) 437,5 peták, a módusz (a leggyakoribb mintaelem) 200 peták. Mivel a mintaelemek száma páros, az empirikus medián definíció szerint most a nagyság szerinti két középső mintaelem átlaga, azaz szintén 200 peták.
b) Az előzővel analóg számolással kapjuk, hogy az átlagfizetés 625 437,5 petákra növekszik, az empirikus medián és a módusz marad 200 peták.

Az átlag és a medián viselkedésével kapcsolatban az a) feladat alapján azt látjuk, hogy pozitív mintaelemek esetén akár egyetlen kiugróan magas mintaelem is felhúzhatja az átlagot. Az adott alkalmazásban ezért megtévesztő lehet az átlagfizetést úgy interpretálni, mint az egyes dolgozók tipikus fizetését, mert akár az is lehetséges (és a példában igaz is), hogy egy kivétellel mindenki kevesebbet keres ennél. Ezzel szemben a medián valóban azt mutatja, hogy a rendezett mintában a középen álló mintaelem (vagy páros mintaelemszám esetén a két középsőnek az átlaga) mekkora. Tehát ezt az adott példában valóban jogosan interpretálhatjuk tipikus/az adott populációban közepes fizetésként.
A b) feladatban az átlag és a medián hasonló viselkedését láthatjuk, kiegészítve azzal, hogy egy mintaelem megváltozása esetén melyik hogyan változik. A mintaátlag egy mintaelem szélsőséges megnövekedése után maga is szélsőségesen nő, ha a mintaelemszám kicsi. Ezzel szemben a medián a kis mintaelemszám ellenére nem vagy alig változik. Úgy is szokás ezt mondani, hogy a medián robusztusabb *az átlagnál az egyes mintaelemek ingadozására (vagy újonnan érkező szélsőségesen nagy mintaelemekre) nézve.*

---

10. (a) Az előadáson elhangzottak szerint F eloszlásfüggvényű fae. mintaelemek esetén $nF_n^*(x)$ binomiális eloszlású n és $p = F(x)$ paraméterekkel. (Ezt fogjuk a későbbi feladatok megoldásában is használni.) Ha X egy kockadobás értéke és F jelöli az ehhez tartozó eloszlásfüggvényt, akkor $F_X(3) = \mathbb{P}(X < 3) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \frac{1}{3}$. Itt tehát $7F_7^*(x) \sim \mathrm{Bin}(7; 1/3)$.
(b) $F_7^*(6)$ a 6-nál kisebb mintaelemek száma leosztva a minta méretével, vagyis 6/7. Hasonló okból $F_7^*(1) = 0$.
Hasonló elvek alapján készül el az ábra. (Az a szakaszonként konstans, az ugrások helyein is balról folytonos függvény, amely az $x = 1$ helyig 0 és a kockadobás minden lehetséges értékénél annyiszor $\frac{1}{7}$-et ugrik, ahány mintaelem van, ami az adott értékkel egyenlő. Tehát: a függvény értéke 1 és 3 között 2/7 [a 2-ben nincs ugrás, mert nincs 2-vel egyenlő mintaelem], a 3 és a 4 között 3/7, a 4 és az 5 között 5/7, az 5 és a 6 között 6/7, a 6-tól kezdve pedig 1.)
(c) Az (a)-beli jelöléssel $\lim_{n\to\infty} F_n^*(x) = F(x)$ minden rögzített $x \in \mathbb{R}$-re 1 valószínűséggel teljesül. Mivel $F(x) = \mathbb{P}(X < x) = 0$, ha $x \le 1$, $F(x) = 1/6$, ha $1 < x \le 2$, $F(x) = 2/6 = 1/3$, ha $2 < x \le 3$, stb., $F(x) = 5/6$, ha $5 < x \le 6$ és végül $F(x) = 1$, ha $x > 6$, az alábbi határértékeket kapjuk: $\lim_{n\to\infty} F_n^*(1) = F(1) = 0$, $\lim_{n\to\infty} F_n^*(3,5) = F(3,5) = 1/2$, $\lim_{n\to\infty} F_n^*(6) = F(6) = 5/6$ és $\lim_{n\to\infty} F_n^*(7) = F(7) = 1$, mind 1 valószínűséggel.

11. (a) Az előző feladat megoldásában leírtak szerint $8F_8^*(0{,}5)$ binomiális eloszlású $n = 8$ és $p = F_{X_1}(0{,}5)$ paraméterrel, ahol $F_{X_1}(0{,}5)$ jelöli X_1 eloszlásfüggvényét. Mivel

$$F_{X_1}(0{,}5) = \int_{-\infty}^{0{,}5} f_{X_1}(x)\mathrm{d}x = \int_0^{0{,}5} 3x^2\mathrm{d}x = [x^3]_0^{0{,}5} = \frac{1}{8}(= 0{,}125),$$

ezért $8F_8^*(0{,}5) \sim \mathrm{Bin}(8; 0{,}125)$. Ennek a binomiális eloszlású val. változónak a szórásnégyzete $\mathbb{D}^2(F_{X_1}(0{,}5)) = np(1-p) = 8 \cdot \frac{1}{8} \cdot \frac{7}{8} = \frac{7}{8}(= 0{,}875)$. Innen adódik, hogy

$$\mathbb{D}(F_8^*(0{,}5)) = \sqrt{\mathbb{D}^2(F_8^*(0{,}5))} = \sqrt{\mathbb{D}^2(\frac{1}{8} \cdot 8F_8^*(0{,}5))} = \sqrt{\frac{1}{8^2}\,\mathbb{D}^2(8F_8^*(0{,}5))} = \sqrt{\frac{1}{64} \cdot \frac{7}{8}} = \sqrt{\frac{7}{512}} \approx 0{,}1169.$$

(b) Ábra. (A függvény szakaszonként konstans és balról folytonos, 0-ról indul, minden mintaelemnél 1/8-ot ugrik, így a 0,95-nél éri el az 1-et, onnantól marad konstans 1.) $F_8^*(0{,}11)$ a 0,11-nél kisebb mintaelemek száma leosztva az összes mintaelem számával, vagyis 0. Hasonlóan $F_8^*(0{,}95) = \frac{7}{8}$. A minta empirikus mediánja páros mintaelemszám esetén nagyság szerint a két középső elem átlaga, vagyis $(0{,}38 + 0{,}42)/2 = 0{,}4$. A minta átlaga a mintaelemek számtani közepe, azaz 0,4625. Így a korrigált empirikus szórás értéke

$$s_8^* = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_i - 0{,}4625)^2} = 0{,}2793.$$

(c) Az előadáson elhangzottak szerint $\lim_{n\to\infty} F_n^*(0{,}5) = F_{X_1}(0{,}5) = 0{,}125$ (lásd az (a) részt).

12. Az eddigiekhez hasonlóan $\lim_{n\to\infty} F_n^*(3) = F(3)$, ahol F jelöli a mintaelemek eloszlásfüggvényét. Mivel a mintaelemek Poisson-eloszlásúak, $F(3) = \mathbb{P}(X_1 < 3) = \mathbb{P}(X_1 = 0) + \mathbb{P}(X_1 = 1) + \mathbb{P}(X_1 = 2) = (1 + \lambda + \lambda^2/2)e^{-\lambda} = 0{,}9921$ a kérdezett határérték.

13. Jelöljön X egy olyan valószínűségi változót, ami a napi kihasználtsággal azonos eloszlású, jelölje eloszlásfüggvényét F. Ekkor a feladat szövege szerint $\mathbb{P}(X \geq 0{,}9) = 0{,}3$, vagyis $F(0{,}9) = \mathbb{P}(X < 0{,}9) = 1 - \mathbb{P}(X \geq 0{,}9) = 1 - 0{,}3 = 0{,}7$. Tehát az $85F_{85}^*(0{,}9)$ valószínűségi változó binomiális eloszlású $n = 85$ és $p = F(0{,}9) = 0{,}7$ paraméterekkel.
A $\mathbb{P}(F_{85}^*(0{,}9) > 0{,}75) = \mathbb{P}(85F_{85}(0{,}9) > 63{,}75)$ valószínűséget kell közelítenünk. Mivel a binomiális eloszlás n paramétere elég nagy, a p pedig független n-től, alkalmazhatjuk a De Moivre–Laplace tételt. Legyen $Y := 85F_{85}^*(0{,}9)$, ekkor sztenderdizálva

$$\mathbb{P}(85F_{85}(0{,}9) > 63{,}75) = \mathbb{P}(Y > 63{,}75) = \mathbb{P}\left(\frac{Y - np}{\sqrt{np(1-p)}} > \frac{63{,}75 - np}{\sqrt{np(1-p)}}\right) = \mathbb{P}\left(\frac{Y - 59{,}5}{\sqrt{17{,}85}} > \underbrace{\frac{63{,}75 - 59{,}5}{\sqrt{17{,}85}}}_{\approx 1{,}0059}\right),$$

amit egy Z standard normális eloszlású val. változóval $\mathbb{P}(Z > 1{,}0059) = \mathbb{P}(Z \geq 1{,}0059) = 1 - \mathbb{P}(Z < 1{,}0059) = 1 - \Phi(1{,}0059) \approx 1 - \Phi(1{,}01) \approx 1 - 0{,}8438 = 0{,}1562$-vel közelíthetünk.

14. A mintaelemek eloszlásfüggvényét jelölje F, ekkor megint csak $nF_n^*(x)$-ről tudjuk, hogy az binomiális eloszlású n és $p = F(x)$ paraméterekkel, és $0 < x < 1$ esetén $F(x) = \mathbb{P}(X_1 < x) = \frac{x-0}{1-0} = x$. Egyszerű átalakítással $\mathbb{P}(F_n^*(0{,}3/n) > 1/n) = \mathbb{P}(nF_n^*(0{,}3/n) > 1)$, és itt az $nF_n^*(0{,}3/n)$ binomiális eloszlású valószínűségi változó paramétereire teljesül, hogy $np = n0{,}3/n = 0{,}3 =: \lambda$.
Mivel n végtelenhez, p nullához, np pedig a $\lambda > 0$ konstanshoz tart, Poisson-approximációt alkalmazhatunk: $\mathbb{P}(nF_n^*(0{,}3/n) > 1) \approx \mathbb{P}(X > 1)$, ahol $X \sim \text{Pois}(\lambda)$. A közelítésünk tehát $\mathbb{P}(F_n^*(0{,}3/n) > 1/n) \approx \mathbb{P}(X > 1) = 1 - \mathbb{P}(X \leq 1) = 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1)) = 1 - e^{-\lambda} - \lambda e^{-\lambda} \approx 0{,}0369$.

15. Bizonyításvázlat: az empirikus eloszlásfüggvény definíció szerint balról folytonos és monoton növő. Továbbá $x \leq X_1^*$ esetén $F_n^*(x) = 0$ és $x > X_n^*$ esetén $F_n^*(x) = 1$, így $\lim_{n\to-\infty} F_n^*(x) = 0$ és $\lim_{n\to\infty} F_n^*(x) = 1$ is teljesül, tehát F_n^* egy (a realizációtól függő) eloszlásfüggvény.
Egy F_n^* eloszlásfüggvényű valószínűségi változónak nem lehet sűrűségfüggvénye, mert F_n^* nem folytonos (minden mintaelemnek megfelelő pontban ugrása van, tehát mindig van legalább egy ugrása).