A sztochasztika alapjai

4. feladatsor: véletlen változók és eloszlásuk

Megoldások

14. Egy permetező szakaszoló szelep napokban mért élettartamának sűrűségfüggvénye $f(x) = 3000/x^4$, ha $x \geq 10$, 0 különben. Mennyi a valószínűsége, hogy a permetező szakaszoló szelep 20 napot túlél? Határozzuk meg az szelep élettartamának eloszlásfüggvényét, várható értékét és szórását!

Megoldás. Jelölje X egy permetező szakaszoló szelep napokban mért élettartamát. Világos, hogy ez egy véletlen mennyiség, véletlen változó. A feladat szerint ez egy folytonos véletlen változó f sűrűségfüggvénnyel, ami definíció szerint azt jelenti, hogy

$$\mathbf{P}(X \leq x) = \int_{-\infty}^{x} f(y)\mathrm{d}y.$$

Az, hogy a permetező szakaszoló szelep 20 napot túlél pontosan azt jelenti, hogy $X > 20$. Ennek a valószínűsége

$$\begin{aligned}
\mathbf{P}(X > 20) &= 1 - \mathbf{P}(X \leq 20) = 1 - \int_{-\infty}^{20} f(y)\mathrm{d}y \\
&= 1 - \int_{10}^{20} 3000\, x^{-4}\mathrm{d}x = 1 - 3000\, \left[-3^{-1}x^{-3}\right]_{x=10}^{x=20} \\
&= 1 + 1000\left(\frac{1}{8000} - \frac{1}{1000}\right) = \frac{1}{8}.
\end{aligned}$$

Itt annyit használtunk, hogy a sűrűségfüggvény értéke 0, ha $x \leq 10$, és ezért elég a $[10, 20]$ intervallumon integrálni, valamint hatványfüggvény integrálját számoltuk ki. Hát persze, ha $\alpha \neq 1$ akkor

$$\int x^{\alpha}\mathrm{d}x = \frac{x^{\alpha+1}}{\alpha+1}.$$

Tehát $\mathbf{P}(X > 20) = \frac{1}{8}$.

Az eloszlásfüggvény meghatározásához a definíciót használjuk: $F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{x} f(y)\mathrm{d}y$. Ha $x \leq 10$, akkor a konstans 0 függvényt integráljuk, ha $x > 10$ akkor pedig a fenti számoláshoz hasonlóan

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y = \int_{10}^{x} 3000\, y^{-4}\mathrm{d}y = 3000\left[-\frac{y^{-3}}{3}\right]_{y=10}^{y=x} = 1 - \frac{1000}{x^3}.$$

Tehát

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y = \begin{cases} 0, & \text{ha } x \leq 10, \\ 1 - \frac{1000}{x^3}, & \text{ha } x \geq 10. \end{cases}$$

Mivel folytonos az eloszlás, az eloszlásfüggvény is folytonos, ezért mindegy, hogy $x = 10$-re melyik ágon definiáljuk az értéket. Azt is látjuk, hogy az eloszlásfüggvény 10-ig konstans 0, utána szigorúan monoton nő, és végtelenben 1-be tart, mint minden eloszlásfüggvény.

A várható értéket is definíció alapján számoljuk. Eszerint

$$\begin{aligned}\mathbf{E}(X) &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_{10}^{\infty} x 3000\, x^{-4}\mathrm{d}x \\ &= 3000 \int_{10}^{\infty} x^{-3}\mathrm{d}x = 3000 \left[-\frac{x^{-2}}{2}\right]_{x=10}^{\infty} \\ &= 3000 \left[\frac{1}{200} - \lim_{x\to\infty} \frac{x^{-2}}{2}\right] = 150.\end{aligned}$$

Itt egy végtelenben improprius integrált kellett kiszámolni, ezért ha nagyon precízen akarjuk írni a dolgot akkor a végtelen felső határ helyett $\lim_{N\to\infty}\int_{10}^{N}$ kellene, azaz

$$\int_{10}^{\infty} h(x)\mathrm{d}x = \lim_{N\to\infty}\int_{10}^{N} h(x)\mathrm{d}x.$$

Ezt később is elhagyjuk, és nem akarunk ebben az értelemben nagyon precízek lenni.

A szórásnégyzet kiszámításához először a második momentumot kell meghatározni. Ez

$$\mathbf{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = 3000 \int_{10}^{\infty} x^{-2}\mathrm{d}x = 3000 \left[-x^{-1}\right]_{x=10}^{x=\infty} = 300.$$

És így a szórásnégyzet

$$\mathbf{D}^2(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = 300 - (15)^2 = 75,$$

azaz a szórás $\mathbf{D}(X) = \sqrt{75}$.

15. Legyen X véletlen változó sűrűségfüggvénye $f(x) = c/x^2$, ha $x > 1$. (a) Határozzuk meg c értékét! (b) Adjuk meg X várható értékét (ha létezik)! (c) Mennyi $\mathbf{P}(X > 4)$? (d) Legyen $Y = 1/X$. Adjuk meg Y eloszlás–, és sűrűségfüggvényét!

Megoldás. (a) Ahhoz, hogy egy f függvény sűrűségfüggvény legyen az kell, hogy ő nemnegatív, és integrálja az egyenesen 1. Mivel a sűrűség értéke cx^{-2} ha $x > 1$, különben pedig 0, ez azt jelenti, hogy $c \geq 0$ és

$$1 = \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{1}^{\infty} cx^{-2}\mathrm{d}x = c\left[-x^{-1}\right]_{x=1}^{\infty} = c.$$

Tehát $c = 1$.

(b) A várható érték definíció szerint

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_{1}^{\infty} x\frac{1}{x^2}\mathrm{d}x = \int_{1}^{\infty} \frac{1}{x}\mathrm{d}x = \infty,$$

azaz a várható érték nem létezik. A várható érték definíciójában feltettük, hogy az $\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x$ improprius integrál létezik és véges. Most ez nem teljesül. Látjuk, hogy vannak olyan véletlen változók, melyeknek nem véges a várható értéke.

(c) A sűrűségfüggvény tulajdonságai szerint $\mathbf{P}(X \in A) = \int_A f(x)\mathrm{d}x$, azaz egy A halmazba esés valószínűsége egyenlő a sűrűségfüggvény A halmazon vett integráljával. Eszerint

$$\mathbf{P}(X > 4) = \int_4^{\infty} f(x)\mathrm{d}x = \int_4^{\infty} x^{-2}\mathrm{d}x = [-x^{-1}]_{x=4}^{\infty} = \frac{1}{4}.$$

Az eloszlásfüggvényt is meghatározzuk. A definíció alapján

$$F(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f(y)\mathrm{d}y = \begin{cases} 0, & \text{ha } x \le 1, \\ \int_1^x y^{-1}\mathrm{d}y = 1 - \frac{1}{x}, & \text{ha } x \ge 1. \end{cases}$$

Ez az 1 paraméterű Pareto-eloszlás.

(d) Legyen $Y = 1/X$. Ezzel a definícióval nincs baj, hiszen $X \ge 1$. Először meghatározzuk Y eloszlásfüggvényét. Mivel $X \ge 1$ ezért $Y = 1/X \in [0,1]$. Ezek szerint

$$G(y) = \mathbf{P}(Y \le y) = \begin{cases} 0, & \text{ha } y \le 0, \\ 1, & \text{ha } y \ge 1. \end{cases}$$

Az érdekes eset amikor $y \in [0,1]$. Ekkor

$$\mathbf{P}(Y \le y) = \mathbf{P}(X^{-1} \le y) = \mathbf{P}(X \ge y^{-1}) = 1 - \mathbf{P}(X \le y^{-1}) = 1 - F(y^{-1}) = y.$$

Itt felhasználtuk, hogy X folytonos véletlen változó, ezért eloszlásfüggvénye is folytonos, valamint beírtuk a (c) részben meghatározott eloszlásfüggvény pontos alakját. Ezek szerint Y eloszlásfüggvénye

$$G(y) = \mathbf{P}(Y \le y) = \begin{cases} 0, & \text{ha } y \le 0, \\ y, & \text{ha } y \in [0,1], \\ 1, & \text{ha } y \ge 1. \end{cases}$$

A sűrűségfüggvény az eloszlásfüggvény deriváltja (majdnem mindenütt). Az eloszlásfüggvény $G(y)$ szép differenciálható függvény, kivéve a 0 és az 1 pontokban. Tehát

$$g(y) = G'(y) = \begin{cases} 1, & \text{ha } y \in (0,1), \\ 0, & \text{különben.} \end{cases}$$

Ez az egyenletes eloszlás a $(0,1)$ intervallumon.

19. Egy dobókockával 10-szer dobunk. Határozzuk meg (a) a kapott hatosok számának eloszlását! (b) az összeg harmadik momentumát!

Megoldás. (a) Ilyet már láttunk. Jelölje X a hatosok számát. Ekkor X diszkrét véletlen változó lehetséges értékei $0, 1, \ldots, 10$, és

$$\mathbf{P}(X = k) = \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k}.$$

Valóban, kiválasztjuk a 10 dobásból melyik k lett hatos, az a k db hatos, a többi meg bármi, de nem hatos.

(b) Legyen most $Y_1, \ldots, Y_{10}$ az egyes dobások eredménye. Ekkor ezek független véletlen változók, hiszen az egyes dobások egymástól függetlenek. Az összeg

$$S = Y_1 + \ldots + Y_{10}.$$

Ez egy diszkrét véletlen változó, melynek lehetséges értékei $10, 11, \ldots, 60$. A pontos eloszlást nem akarjuk meghatározni. Ehelyett írjuk fel a harmadik momentumot:

$$\mathbf{E}(S^3) = \mathbf{E}\left[\left(\sum_{i=1}^{10} Y_i\right)^3\right].$$

Az összeg harmadik hatványát kifejtve lesznek Y_i^3, $Y_i^2 Y_j$, $Y_i Y_j Y_k$ alakú tagok, ahol i, j, k különbözőek. Azt kell összeszámolni, hogy az egyes típusú tagokból mennyi lesz. Először is, Y_i^3-ból pontosan 1 lesz minden i-re, hiszen mind a három szorzótényezőből az Y_i-t kell kivenni. Így összesen 10 ilyen tag lesz. Az $Y_i^2 Y_j$ alakú tagokból pedig 3 lesz, hiszen kettő tényezőből Y_i-t vesszük ki, a harmadikból pedig Y_j-t. Mivel i és j szerepe nem szimmetrikus, összesen $10 \cdot 9$ féle i, j választás van, így $3 \cdot 10 \cdot 9 = 270$ ilyen tag lesz. Végül $Y_i Y_j Y_k$ alakú tagból 6 lesz, mert az Y_i-t három helyről, Y_j-t kettőről választhatjuk. Az i, j, k értékeket pedig $\binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{2 \cdot 3} = 120$ féleképpen választhatjuk, azaz összesen $6 \cdot 120 = 720$ ilyen tagot kapunk. Összegezve

$$\mathbf{E}(S^3) = 10 \cdot \mathbf{E}(Y_1^3) + 270 \cdot \mathbf{E}(Y_1^2 Y_2) + 720 \cdot \mathbf{E}(Y_1 Y_2 Y_3).$$

Tényleg megszámoltunk minden tagot, hiszen $10^3 = 1000 = 10+270+720$. A függetlenség miatt $\mathbf{E}(Y_1^2 Y_2) = \mathbf{E}(Y_1^2)\mathbf{E}(Y_2)$, és

$$\mathbf{E}(Y_1 Y_2 Y_3) = \mathbf{E}(Y_1)\mathbf{E}(Y_2)\mathbf{E}(Y_3) = (\mathbf{E}(Y_1))^3.$$

Tehát

$$\mathbf{E}(S^3) = 10 \cdot \mathbf{E}(Y_1^3) + 270 \cdot \mathbf{E}(Y_1^2)\mathbf{E}(Y_2) + 720 \cdot (\mathbf{E}(Y_1))^3.$$

Végül kiszámoljuk a megfelelő momentumokat. Definíció szerint

$$\mathbf{E}(Y_1) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \ldots + \frac{1}{6} \cdot 6 = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = \frac{7}{2},$$
$$\mathbf{E}(Y_1^2) = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \ldots + \frac{1}{6} \cdot 6^2 = \frac{1}{6} \cdot \frac{6 \cdot 7 \cdot 13}{6} = \frac{91}{6},$$
$$\mathbf{E}(Y_1^3) = \frac{1}{6} \cdot 1^3 + \frac{1}{6} \cdot 2^3 + \ldots + \frac{1}{6} \cdot 6^3 = \frac{1}{6} \cdot \left(\frac{6 \cdot 7}{2}\right)^2 = \frac{147}{2}.$$

Végül ezeket visszahelyettesítve,

$$\mathbf{E}(S^3) = 10 \cdot \frac{147}{2} + 270 \cdot \frac{91}{6}\frac{7}{2} + 720 \cdot \left(\frac{7}{2}\right)^3 = \frac{91875}{2}.$$

A feladat nem túl izgalmas, de valójában a várható érték fontos tulajdonságainak alkalmazásait gyakoroltuk.

20. Egy urnában van 20 piros és 30 fehér golyó. Húzzunk ki 20 golyót visszatevés nélkül. Számoljuk ki a kihúzott piros golyók számának várható értékét és szórásnégyzetét!

Megoldás. Jelölje X a kihúzott piros golyók számát. Világos, hogy X diszkrét véletlen változó, melynek lehetséges értékei $0, 1, 2, \ldots, 20$. Számoljuk ki a $\mathbf{P}(X = k)$ valószínűségeket. Klasszikus valószínűségi mezőn vagyunk, az összes eset száma annyi, ahányféleképpen ki tudunk választani 50 golyóból 20-at. Azaz $\binom{50}{20}$. Az $\{X = k\}$ esemény azt jelenti, hogy pontosan k db piros golyót húztunk, azaz a 20 pirosból k-t, a 30 fehérből $20 - k$-t. Ezt $\binom{20}{k} \cdot \binom{30}{20-k}$ féleképpen tehetjük meg. Tehát

$$\mathbf{P}(X = k) = \frac{\binom{20}{k} \cdot \binom{30}{20-k}}{\binom{50}{20}}, \quad k = 0, 1, \ldots, 20.$$

Ez az X véletlen változó eloszlás. Ő a negatív binomiális eloszlás.

Innen a várható értéket meghatározhatjuk a definíció alapján. Eszerint

$$\mathbf{E}(X) = \sum_{k=0}^{20} k\mathbf{P}(X = k) = \sum_{k=1}^{20} k \frac{\binom{20}{k} \cdot \binom{30}{20-k}}{\binom{50}{20}}.$$

Ezt kellene zárt alakra hozni. Vegyük észre, hogy

$$k\binom{20}{k} = 20\binom{19}{k-1}. \tag{$*$}$$

Valóban, egy 20 fős társaságból kell kiválasztani egy k fős bizottságot, és annak a vezetőjét. Ezt számoljuk meg a bal és jobb oldalon is. A bal oldalon először kiválasztjuk a k fős csapatot, és utána annak a vezetőjét, míg a jobb oldalon először a vezetőt választjuk ki, és utána a maradék 19 főből választunk még $k - 1$-et. Ezt a formulát felhasználva

$$\sum_{k=1}^{20} k\binom{20}{k} \cdot \binom{30}{20-k} = \sum_{k=1}^{20} 20\binom{19}{k-1} \cdot \binom{30}{20-k}.$$

Az összegben megjelenő $\binom{19}{k-1} \cdot \binom{30}{20-k}$ kifejezés pontosan olyan, mint amit korábban kaptunk, csak most összesen 49 golyó közül, melyből 19 piros és 30 fehér, választunk ki 19-et. Tehát a fenti összeg az összes kiválasztást adja meg, azaz

$$\sum_{k=1}^{20} \binom{19}{k-1} \cdot \binom{30}{20-k} = \sum_{\ell=0}^{19} \binom{19}{\ell} \cdot \binom{30}{19-\ell} = \binom{49}{19}.$$

Ezt visszahelyettesítve a várható értékre kapott formulába

$$\mathbf{E}(X) = 20\frac{\binom{49}{19}}{\binom{50}{20}} = 20\frac{2}{5} = 8.$$

Itt felhasználtuk azt is, hogy $50\binom{49}{19} = 20\binom{50}{20}$, ami éppen $(*)$ csak más számokkal.

A szórás meghatározásához először a második momentum kell. Ez

$$\mathbf{E}(X^2) = \sum_{k=0}^{20} k^2\mathbf{P}(X=k) = \sum_{k=0}^{20} k^2\frac{\binom{20}{k}\cdot\binom{30}{20-k}}{\binom{50}{20}}.$$

Ez még az előbbinél is bonyolultabb formula. Használjuk, hogy $k^2 = k(k-1)+k$, így a fenti összegben megjelenik a már meghatározott várható érték. A $(*)$ formulához hasonlóan

$$k(k-1)\binom{20}{k} = 20\cdot 19\cdot\binom{18}{k-2}.$$

A bal- és jobboldalon is azt számoljuk össze, hogy hányféleképp lehet egy 20 fős társaságból kiválasztani egy k fős bizottságot, annak a vezetőjét, és egy helyettest. Ezt a formulát felhasználva

$$\begin{aligned}\sum_{k=2}^{20} k(k-1)\binom{20}{k}\binom{30}{20-k} &= \sum_{k=2}^{20} 20\cdot 19\cdot\binom{18}{k-2}\binom{30}{20-k}\\ &= 20\cdot 19\sum_{\ell=0}^{18}\binom{18}{\ell}\binom{30}{18-\ell} = 20\cdot 19\binom{48}{18}.\end{aligned}$$

Itt használtuk, hogy $k(k-1)=0$ ha $k=0$ vagy $k=1$, ezért elegendő 2-től összegezni, valamint a várható érték kiszámításánál is használt formulát. Most 18 piros és 30 fehér golyó közül veszünk visszatevés nélkül 18-at. Ezt visszahelyettesítve a második momentum formulájába

$$\begin{aligned}\mathbf{E}(X^2) &= \sum_{k=0}^{20} k^2\frac{\binom{20}{k}\cdot\binom{30}{20-k}}{\binom{50}{20}}\\ &= \sum_{k=0}^{20} k(k-1)\frac{\binom{20}{k}\cdot\binom{30}{20-k}}{\binom{50}{20}} + \sum_{k=0}^{20} k\frac{\binom{20}{k}\cdot\binom{30}{20-k}}{\binom{50}{20}}\\ &= 20\cdot 19\frac{\binom{48}{18}}{\binom{50}{20}} + \mathbf{E}(X) = 20\cdot 19\frac{20\cdot 19}{50\cdot 49} + 8.\end{aligned}$$

Az utolsó egyenlőségnél felhasználtuk, hogy $20\cdot 19\binom{50}{20} = 50\cdot 49\binom{48}{18}$, amit már láttunk. Tehát a szórásnégyzet

$$\mathbf{D}^2(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \frac{(20\cdot 19)^2}{50\cdot 49} + 8 - 64 = \frac{144}{49},$$

és így $\mathbf{D}(X) = \frac{12}{7}$.

2. Megoldás. Megadunk egy másik megoldást, aminek az ötletét már használtuk (pl. Csodaországos tévés feladat) és később is használni fogjuk. Jelölje megint X a kihúzott piros golyók számát. Ekkor persze X diszkrét véletlen változó, de most nem határozzuk meg az eloszlását. Ehelyett X-et felírjuk egyszerűbb változók összegeként. A módszer: *írjuk fel indikátorok összegeként!*

Legyen $i = 1, 2, \ldots, 20$ esetén

$$I_i = \begin{cases} 1, & \text{ha az } i\text{-ediknek húzott golyó piros,} \\ 0, & \text{különben.} \end{cases}$$

Ekkor $X = \sum_{i=1}^{20} I_i$. Egy indikátorváltozó mindig két értéket vehet fel, 0-t és 1-et. Ezért várható értéke a definíció szerint

$$\mathbf{E}(I_i) = 0 \cdot \mathbf{P}(I_i = 0) + 1 \cdot \mathbf{P}(I_i = 1) = \mathbf{P}(I_i = 1).$$

Tehát azt kell meghatároznunk, hogy mennyi a valószínűsége, hogy az i-edik golyó piros. Ez egyszerű, hiszen az i-edik helyre 50 golyó közül választhatunk, abból 20 piros, tehát tetszőleges i esetén

$$\mathbf{P}(I_i = 1) = \mathbf{P}(\, i\text{-edik golyó piros}\,) = \frac{20}{50} = \frac{2}{5}.$$

Tehát

$$\mathbf{E}(X) = \mathbf{E}\left(\sum_{i=1}^{20} I_i\right) = \sum_{i=1}^{20} \mathbf{E}(I_i) = 20 \cdot \frac{2}{5} = 8.$$

Így talán egyszerűbb volt a számolás, mint az előbb. Annyit használtunk, hogy *összeg várható értéke az a várható értékek összege.* Ez mindig teljesül, tetszőleges véletlen változók esetén.

A szórásnégyzet meghatározása is hasonlóan történik, de itt már többet kell számolni. A kovariancia tulajdonságai szerint

$$\begin{aligned} \mathbf{D}^2(X) = \mathbf{Cov}(X, X) = \mathbf{Cov}\left(\sum_{i=1}^{20} I_i, \sum_{i=1}^{20} I_i\right) \\ = \sum_{i=1}^{20}\sum_{j=1}^{20} \mathbf{Cov}(I_i, I_j) = \sum_{i=1}^{20} \mathbf{D}^2(I_i) + 2 \sum_{1 \le i < j \le 20} \mathbf{Cov}(I_i, I_j). \end{aligned}$$

Vegyük észre, hogy I_i eloszlása nem függ i-től, azaz annak a valószínűsége, hogy elsőre pirosat húzunk, pontosan ugyanaz, mint annak a valószínűsége, hogy 7.-re. Ezt talán úgy látjuk legkönnyebben, ha a hetedik húzásra nem úgy gondolunk, ami előtt már volt hat másik, hanem mint a húsz húzás közül az egyikre. Azaz ekkor semmi más nem érdekel minket, csak a hetedik húzás. Ugyanígy látjuk, hogy (I_1, I_2) együttes eloszlása ugyanaz, mint (I_i, I_j) együttes eloszlása tetszőleges $i \neq j$ esetén. Azaz, annak a valószínűsége, hogy az első két húzás piros, pontosan ugyanaz, mint annak a valószínűsége, hogy a hetedik és tizenharmadik piros. Ezek szerint $\mathbf{D}^2(I_i) = \mathbf{D}^2(I_1)$ és $\mathbf{Cov}(I_i, I_j) = \mathbf{Cov}(I_1, I_2)$. Tehát, folytatva a szórásnégyzet kiszámolását

$$\mathbf{D}^2(X) = 20\mathbf{D}^2(I_1) + 20 \cdot 19\mathbf{Cov}(I_1, I_2).$$

A definíció szerint

$$\mathbf{D}^2(I_1) = \mathbf{E}(I_1^2) - (\mathbf{E}(I_1))^2 = \frac{2}{5} - \left(\frac{2}{5}\right)^2 = \frac{2}{5} \cdot \frac{3}{5},$$

és

$$\mathbf{Cov}(I_1, I_2) = \mathbf{E}(I_1 I_2) - \mathbf{E}(I_1) \cdot \mathbf{E}(I_2) = \mathbf{P}(I_1 = 1, I_2 = 1) - \left(\frac{2}{5}\right)^2 = \frac{20 \cdot 19}{50 \cdot 49} - \left(\frac{2}{5}\right)^2,$$

hiszen annak a valószínűsége, hogy az első és második golyó is piros $(20 \cdot 19)/(50 \cdot 49)$. Összegezve,

$$\mathbf{D}^2(X) = 20\frac{2}{5} \cdot \frac{3}{5} + 20 \cdot 19 \cdot \left(\frac{20 \cdot 19}{50 \cdot 49} - \left(\frac{2}{5}\right)^2\right) = \frac{144}{49},$$

és így $\mathbf{D}(X) = \frac{12}{7}$.

21. Egy urnában van 20 piros és 30 fehér golyó. Húzzunk ki 20 golyót visszatevéssel. Számoljuk ki a kihúzott piros golyók számának várható értékét és szórásnégyzetét!

Megoldás. Ezt a feladatot is megoldhatjuk mindkét fenti módszerrel. Mivel visszatevéssel húzunk, ezért minden húzás előtt pontosan ugyanaz van az urnában, azaz igazából egy kísérletet ismétlek 20-szor, ahol a siker valószínűsége 2/5, hiszen a piros golyókat figyelem, annak a kihúzása lesz most a siker. Tehát, ha Y jelöli a kihúzott pirosak számát, akkor Y lehetséges értékei $0, 1, \ldots, 20$, és

$$\mathbf{P}(Y = k) = \binom{20}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{20-k}.$$

Vagyis Y binomiális eloszlású 20 és 2/5 paraméterekkel. Binomiális eloszlás várható értékét, szórását már számoltuk,

$$\mathbf{E}(Y) = 20 \cdot \frac{2}{5} = 8, \quad \mathbf{D}^2(Y) = 20 \cdot \frac{2}{5} \cdot \frac{3}{5} = \frac{24}{5}.$$

Felbonthatom Y-t is indikátorok összegére. Ha $J_i = 1$, ha az i-edik golyó piros, 0 különben, akkor

$$Y = J_1 + \ldots + J_{20}.$$

Eszerint a várható érték és a szórásnégyzet is könnyen kiszámolható. Most a változók függetlenek, ezért az összeg szórásnégyzete megegyezik a szórásnégyzetek összegével.

Az előző feladattal összevetve nagyon fontos különbség, hogy itt $J_1, \ldots, J_{20}$ *független* véletlen változók, hiszen az első 5 húzás semmilyen módon nem befolyásolja a hatodikat. Ezzel ellentétben, ha visszatevés nélkül húzunk, akkor $I_1, \ldots, I_{20}$ nem függetlenek (persze a kovarianciájuk sem 0), mert ha az első 5 húzás piros, akkor a hatodik húzás előtt az urnában már csak 15 piros van és 30 fehér.

Vegyük észre, hogy a várható érték mindkét modell esetén ugyanaz, míg $\mathbf{D}^2(X) < \mathbf{D}^2(Y)$, azaz visszatevés nélkül a szórásnégyzet jóval kisebb. Gondoljuk végig, mért természetes ez!