Valószínűségszámítás előadás, 5. hét, 2020. október 7.

Nevezetes diszkrét eloszlások, függetlenség, a várható érték és szórás tulajdonságai

1. Nevezetes diszkrét eloszlások

A binomiális eloszláson kívül más eloszláscsaládok is gyakran előfordulnak valós helyzetek matematikai modellezésénél.

1.1. Poisson-eloszlás

Emlékeztetőül:

1.1. Definíció (Binomiális eloszlás). *Az X valószínűségi változó* **binomiális eloszlású** *n renddel és p paraméterrel, ha lehetséges értékei:*

$$0, 1, 2, \dots, n,$$

és minden $0 \le k \le n$ egészre

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

($n \ge 1$ egész, $0 < p < 1$.) Jelölés: $Bin(n, p)$.

Ez jött elő például a visszatevéses mintavételnél is. Általában: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeres kísérletek száma lehet.

Nézzünk egy olyan esetet, amikor a kísérletek száma jóval nagyobb, mint a sikeres kísérletek számának várható értéke, azaz ritkának mondhatók a sikeres kísérletek.

Tegyük fel, hogy egy biztosító $n = 100000$ ügyfelének mindegyike egy év alatt egymástól függetlenül $p = 0,0001$ valószínűséggel okoz balesetet. A balesetet okozó ügyfelek számának (ezt jelöljük X-szel) **várható értéke**:

$$\mathbb{E}(X) = np = 100000 \cdot 0,0001 = \mathbf{10}.$$

Ilyenkor annak valószínűségét, hogy **pontosan k ügyfél okoz balesetet**, felírhatjuk pontosan, de ez alapján egy közelítő számítást is végezhetünk, ha k értéke nem túlságosan nagy, mondjuk $k = 15$:

$$\begin{aligned}
\mathbb{P}(X = k) &= \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} = \\
&= \frac{100000 \cdot 99999 \cdot \ldots (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx \\
&\approx \frac{100000^k \cdot 0,0001^k}{k!} \left(1 - \frac{10}{100000}\right)^{100000} \approx \frac{10^k}{k!} e^{-10},
\end{aligned}$$

felhasználva, hogy $\lim_{n\to\infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$ tetszőleges $x > 0$-ra.

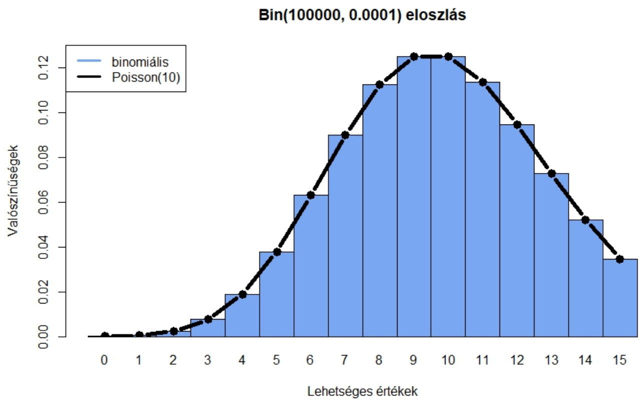
Az 1. ábrán láthatjuk, hogy valóban jó a közelítés: az oszlopok az $n = 100000$ rendű és $p = 0,0001$ paraméterű binomiális eloszlást ábrázolják (vízszintes tengely: k, oszlopok magassága: $\mathbb{P}(X = k)$).
Feketével a $\frac{10^k}{k!} e^{-10}$ függvény látható. Ez utóbbi is egy eloszlás, a gyakran használt Poisson-eloszlás.

1.2. Definíció. *Legyen $\lambda > 0$. Az X valószínűségi változó* **λ paraméterű Poisson-eloszlású**, *ha lehetséges értékei:*

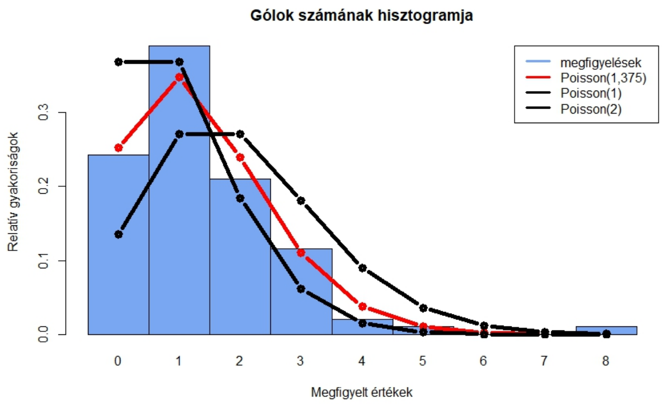
$$k = 0, 1, 2, \dots, \quad \textit{és ekkor } \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

1.1. Állítás. *A λ paraméterű Poisson-eloszlás* **várható értéke** *és* **szórása***:*

$$\mathbb{E}(X) = \lambda; \qquad \mathbb{D}(X) = \sqrt{\lambda}.$$



1. ábra. A binomiális eloszlás közelítése Poisson-eloszlással



2. ábra. A gólok számának hisztogramja $n = 95$ mérkőzésen, és különböző paraméterű Poisson-eloszlások

A 2. ábrán pedig az látható, hogy valós adatokra is jól illeszthető a Poisson-eloszlás, ha a megfelelő paramétert választjuk. Itt az adatok átlaga

- **ritkán bekövetkező események száma adott időszak alatt**:
  - lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az egyik első statisztikai példa a XIX. század második felében)
  - a balesetek száma egy városban egy hét vagy egy hónap alatt;
  - a földrengések száma egy év alatt
- egy rendszerbe beérkező igények száma egy adott időszakban:
  - egy üzletbe beérkező vásárlók száma egy óra alatt
  - egy weboldal letöltéseinek száma egy óra alatt
- **általában:** véletlen időközönként bekövetkező események száma adott időszak alatt

Poisson-folyamat: a t idő alatt bekövetkező események száma Poisson-eloszlású ct paraméterrel.

Ekkor a t idő alatt bekövetkező események számának várható értéke: ct, azaz az intervallum hosszával arányos.

1.2. Hipergeometriai eloszlás

A visszatevés nélküli mintavételben a húzott fekete golyók számát hipergeometriai eloszlással írhatjuk le. Nézzünk először egy példát.

Egy sportcsapat $N = 20$ **tagja** közül $M = 9$ **balkezes**.
A pályán egyszerre $n = 7$ **különböző** játszik.
Tegyük fel, hogy minden hétfős összeállítás egyformán valószínű.

Milyen eloszlású a pályán a **balkezes játékosok száma,** X**?**

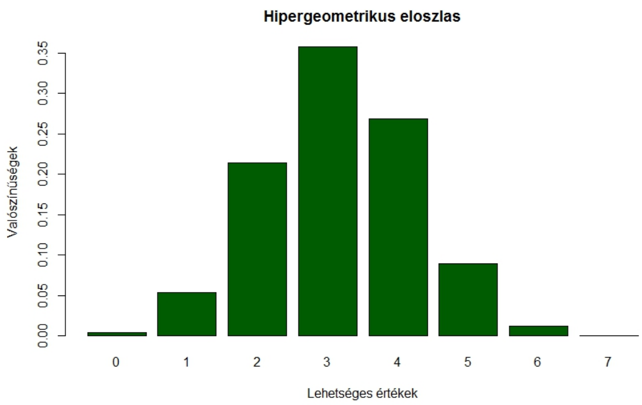
A hétfős összeállítások száma:	$\binom{20}{7}$
k balkezes játékos kiválasztása:	$\binom{9}{k}$
$7-k$ jobbkezes játékos kiválasztása:	$\binom{11}{7-k}$
A jó lehetőségek száma összesen:	$\binom{9}{k} \cdot \binom{11}{7-k} \longleftarrow$

szorzás: bármely balkezes választás bármely jobbkezessel jó

osztás: minden lehetőség egyformán valószínű

$$\boxed{\mathbb{P}(X = k) = \frac{\binom{9}{k} \cdot \binom{11}{7-k}}{\binom{20}{7}}}$$

visszatevés nélküli mintavétel



3. ábra. A kiválasztott balkezes játékosok számának eloszlása hipergeometriai eloszlás, $N = 20$, $M = 9, n = 7$ vízszintes: k, oszlopok magassága: $\mathbb{P}(X = k) = \frac{\binom{9}{k} \cdot \binom{11}{7-k}}{\binom{20}{7}}$.

1.3. Definíció (Hipergeometriai eloszlás). *Legyenek N, M, n pozitív egészek úgy, hogy $1 \leq n \leq M \leq N$. Az X valószínűségi változó* **hipergeometriai eloszlású**, *ha*

$$\mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad (k = 0, 1, \ldots, n).$$

- **visszatevés nélküli mintavételnél** a húzott fekete golyók száma: N golyó, ebből M fekete, n-szer húzunk visszatevés nélkül
- lottósorsolásnál a találatok száma, X, hipergeometrikus eloszlású $N = 90$, $M = 5$, $n = 5$ paraméterekkel:
$$\mathbb{P}(X = k) = \mathbb{P}(k \text{ találat}) = \frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}} \qquad k = 0, 1, 2, 3, 4, 5.$$

1.2. Állítás. *Ha az X valószínűségi változó hipergeometriai eloszlású M, N, n paraméterekkel, azaz*

$$\mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad (k = 0, 1, \ldots, n),$$

akkor

$$\mathbb{E}(X) = \frac{M}{N}n; \qquad D(X) = \sqrt{n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}}.$$

Korábbi példa folytatása. Ha $N = 20$ játékos közül $M = 9$ balkezes, és $n = 7$-et választunk visszatevés nélkül, akkor a balkezes játékosok számának, X-nek **várható értéke** és **szórása**:

$$\mathbb{E}(X) = \frac{9}{20} \cdot 7 = \mathbf{3,15}; \quad D(X) = \sqrt{7 \cdot \frac{9}{20} \cdot \left(1 - \frac{9}{20}\right) \cdot \frac{13}{19}} = \mathbf{1,09}.$$

1.3. Geometriai eloszlás

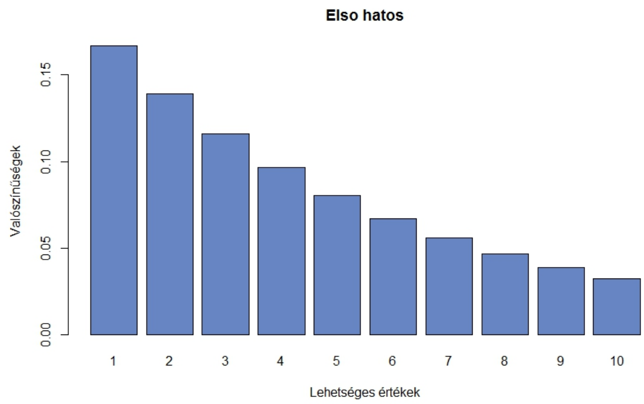
Egy közvéleménykutatásban mindenki a többiektől függetlenül $0,2$ valószínűséggel válaszol egy adott kérdésre. Jelölje Y, hogy hány embert kell megkérdezni, míg találunk egy válaszadót.

$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$

$\mathbb{P}(Y = 2) = \mathbb{P}(\text{az első nem válaszol, a második igen}) = 0,8 \cdot 0,2;$

$\mathbb{P}(Y = 3) = \mathbb{P}(\text{az első kettő nem válaszol, a harmadik igen}) = 0,8^2 \cdot 0,2;$

$\mathbb{P}(Y = k) = \mathbb{P}(\text{az első } k-1 \text{ nem válaszol, a } k. \text{ igen}) = 0,8^{k-1} \cdot 0,2.$



4. ábra. Az első hatos eloszlása: geometriai eloszlás, $p = 1/6$, $k = 10$-ig

A geometriai eloszlást tehát az alábbi helyzetben használhatjuk:

- független kísérleteket végzünk;
- mindegyik p valószínűséggel sikerül;
- Y: hányadik kísérlet az első sikeres.

1.4. Definíció. *Az Y valószínűségi változó* **geometriai eloszlású** *p paraméterrel, ha lehetséges értékei:*

$$1, 2, 3 \ldots$$

és minden $1 \leq k$ egészre

$$\mathbb{P}(Y = k) = (1-p)^{k-1} p.$$

($0 < p < 1$.) Jelölés: Geo(p). Másik elnevezés: Pascal-eloszlás.

Mivel $\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1$, ez valóban valószínűségeloszlás. Ebből az is következik, hogy nnak valószínűsége, hogy sosem sikerül a kísérlet, 0 (de ez nem a lehetetlen esemény).

1.3. Állítás. *Ha az X valószínűségi változó geometriai eloszlású p paraméterrel, azaz*

$$\mathbb{P}(X = k) = (1-p)^{k-1} p \qquad (k = 1, 2, \ldots),$$

akkor

$$\mathbb{E}(X) = \frac{1}{p}; \qquad D(X) = \sqrt{\frac{1-p}{p^2}}.$$

Példa. Tegyük fel, hogy egy adott pártot mindenki a többiektől függetlenül $p = 0,06$ valószínűséggel támogat. Jelölje X, hogy hány embert kell megkérdezni, míg az első olyan embert megtaláljuk, aki ezt a pártot támogatná. Ekkor

$$\mathbb{E}(X) = \frac{1}{0,06} = 16,67; \qquad D(X) = \sqrt{\frac{0,94}{0,06^2}} = 16,16.$$

1.4. Nevezetes diszkrét eloszlások: összefoglalás

$X : \Omega \to \mathbb{R}$ diszkrét valószínűségi változó eloszlása: a lehetséges értékei és a hozzájuk tartozó valószínűségek.

- binomiális eloszlás: n független kísérlet, mindegyik p valószínűséggel sikerül, X: hány sikerült. $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, ahol $k = 0, 1, 2 \ldots, n$. Ekkor

$$\mathbb{E}(X) = np; \qquad D(X) = \sqrt{np(1-p)}.$$

- Poisson-eloszlás: az X valószínűségi változó Poisson-eloszlású λ paraméterrel, ha

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2 \ldots).$$

  Ekkor

$$\mathbb{E}(X) = \lambda; \qquad D(X) = \sqrt{\lambda};$$

- hipergeometriai eloszlás: visszatevés nélküli mintavételnél a húzott fekete golyók száma; $\mathbb{P}(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$. Ekkor

$$\mathbb{E}(X) = \frac{M}{N} n; \qquad D(X) = \sqrt{n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}}.$$

- geometriai eloszlás: minden kísérlet p valószínűséggel sikerül; Y: hányadik az első sikeres. $\mathbb{P}(Y = k) = (1-p)^{k-1} p$, ahol $k = 1, 2, \ldots$. Ekkor

$$\mathbb{E}(X) = \frac{1}{p}; \qquad D(X) = \sqrt{\frac{1-p}{p^2}}.$$

2. Valószínűségi változók függetlensége

Két valószínűségi változó között sokféle kapcsolat képzelhető el. Ezt valamilyen értelemben mérni is tudjuk majd, de először azt fogalmazzuk meg, hogy mit értünk azon, amikor „nincs kapcsolat” a kettő között, azaz független a két valószínűségi változó. Ehhez az eseményekre megfogalmazott függetlenség fogalmát tudjuk majd kiindulásként használni.

Emlékeztetőül: az A és B események **függetlenek**, ha

$$\mathbb{P}(\mathbf{A} \cap \mathbf{B}) = \mathbb{P}(\mathbf{A}) \cdot \mathbb{P}(\mathbf{B}).$$

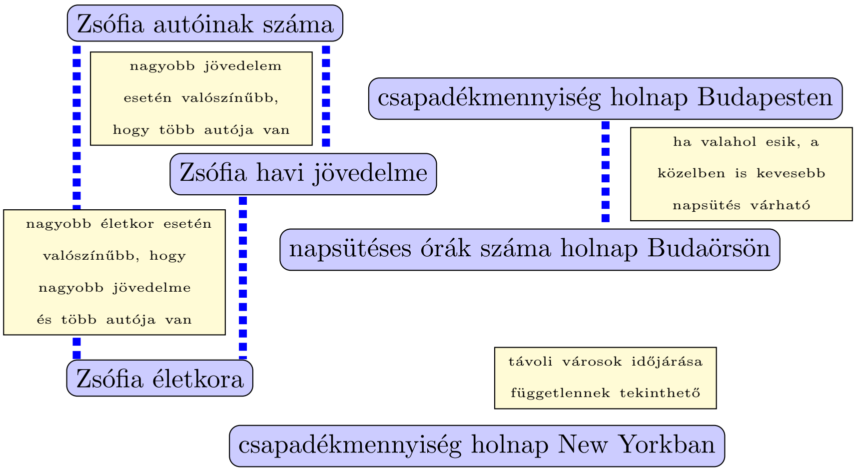
Ha például X a csapadékmennyiség holnap Budapesten (mm-ben), és Y New Yorkban, akkor például

$$\mathbf{A} : X \leq 5; \qquad \mathbf{B} : Y \leq 3$$

esetén ez a feltétel így írható:

$$\mathbb{P}(X \leq 5, Y \leq 3) = \mathbb{P}(X \leq 5) \cdot \mathbb{P}(Y \leq 3).$$

Azaz, feltételezve, hogy a két város időjárása egymástól független: annak valószínűsége, hogy **mindkét helyen legfeljebb 5 mm csapadék lesz**, a két esemény **valószínűségének szorzata**.



5. ábra. Példák függetlennek tekinthető (nem összekötött) és nem független (összekötött) valószínűségi változókra

2.1. Definíció (Valószínűségi változók függetlensége). *Valószínűségi változók függetlenségét az alábbi módon definiálhatjuk.*

- **két valószínűségi változóra**: *az* $X, Y : \Omega \to \mathbb{R}$ *valószínűségi változók* **függetlenek**, *ha*

$$\mathbb{P}(X \leq t_1, Y \leq t_2) = \mathbb{P}(X \leq t_1) \cdot \mathbb{P}(Y \leq t_2)$$

  teljesül tetszőleges $t_1, t_2 \in \mathbb{R}$ *valós számokra.*

- **véges sok valószínűségi változóra**: $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *valószínűségi változók* **függetlenek**, *ha*

$$\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n) = \\ = \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \ldots \mathbb{P}(X_n \leq t_n)$$

  teljesül tetszőleges $t_1, t_2, \ldots, t_n$ *valós számokra.*

- **megszámlálható sok valószínűségi változóra**: *az* $X_1, X_2, X_3 \ldots$ *valószínűségi változók* **függetlenek**, *ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.*

2.1. Függetlenség diszkrét esetben

Ha a valószínűségi változók **diszkrétek**, azaz lehetséges értékeik halmaza véges vagy megszámlálhatóan végtelen, akkor a függetlenséget az alábbi módon is ellenőrizhetjük.

Az X és Y **diszkrét** valószínűségi változók pontosan akkor **függetlenek**, ha

az X **minden lehetséges** x_k **értékére** és

az Y **minden lehetséges** y_l **értékére** teljesül, hogy

$$\mathbb{P}(X = x_k, Y = y_l) = \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_l) \qquad (k, l = 1, 2, \ldots).$$

Vagyis: annak valószínűsége, hogy X **értéke** x_k **és** Y **értéke** y_l, ennek a két eseménynek **a valószínűségének a szorzata**.

Példa. Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy az **első dobott szám** és a **második dobott szám** értéke egymástól független?

Tipp. a második dobásnál az első dobás értéke „elfelejtődik”, nincs kapcsolat a két dobás között ⇒ a két dobott szám **független.**

Indoklás. Legyen X az első dobás, Y a második. Legyen például $x_k = 3, y_l = 5$. Ekkor a feltétel teljesül:

$$\frac{1}{36} = \mathbb{P}(X = 3, Y = 5) = \mathbb{P}(X = 3) \cdot \mathbb{P}(Y = 5) = \frac{1}{6} \cdot \frac{1}{6}.$$

Hasonlóképpen tetszőleges (x_k, y_l) lehetséges értékekre (azaz 1 és 6 közötti egészekre) igaz, hogy

$$\frac{1}{36} = \mathbb{P}(X = x_k, Y = y_l) = \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_l) = \frac{1}{6} \cdot \frac{1}{6}.$$

Ezért valóban **a két dobás egymástól független**.

Példa. Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy a dobott számok **összege** és **szorzata** független egymástól?

Tipp: minél nagyobb az összeg, annál valószínűbb, hogy a szorzat értéke is inkább nagy lesz $\Rightarrow$ **nem függetlenek**.

Indoklás: legyen X az összeg, Y a szorzat. Ha például $X = 2$: ez csak úgy lehet, hogy mindkét dobás 1-es, vagyis ekkor Y értéke biztosan 1. Ezért ha például $x_1 = 2$ és $y_2 = 2$-t választunk, $X = 2$ és $Y = 2$ egyszerre nem következhetnek be, és így:

$$\begin{aligned} 0 = \mathbb{P}(X = 2, Y = 2) \neq \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 2) &= \\ = \mathbb{P}(11) \cdot \mathbb{P}(12 \text{ vagy } 21) &= \frac{1}{36} \cdot \frac{1}{18} > 0. \end{aligned}$$

Vagyis az $x_1 = 2$ és $y_2 = 2$ párra nem teljesül az előírt feltétel, az **összeg és szorzat nem függetlenek**.

2.2. A várható érték és a szórás tulajdonságai

2.3. A várható érték tulajdonságai

A várható érték az alábbi tulajdonságokkal rendelkezik.

- (összeg várható értéke) Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- (konstans kiemelése) Ha az X valószínűségi változó várható értéke létezik, és c tetszőleges valós szám, akkor

$$\mathbb{E}(c \cdot X) = c \cdot \mathbb{E}(X).$$

- (szorzat várható értéke független esetben) Ha az X és Y valószínűségi változók függetlenek, és $X, Y, X \cdot Y$ várható értéke létezik, akkor

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

- (függvény várható értéke) Ha $g : \mathbb{R} \to \mathbb{R}$ olyan függvény, hogy $\mathbb{E}(X)$ létezik, akkor

$$\mathbb{E}\big(g(X)\big) = \sum_{i=1}^{\infty} g(x_i)\mathbb{P}(X = x_i),$$

  ahol az X lehetséges értékei $x_1, x_2, \ldots$.

Ebből két állításnak a bizonyítását is megnézzük.

2.1. Állítás. *Ha X, Y valószínűségi változók, és $X, Y, X + Y$ várható értéke létezik, akkor*

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_{k,m} (x_k + y_m)\mathbb{P}(X = x_k, Y = y_m) = \\ &= \sum_{k,m} x_k \mathbb{P}(X = x_k, Y = y_m) + \sum_{k,m} y_m \mathbb{P}(X = x_k, Y = y_m) = \\ &= \sum_{k} x_k \mathbb{P}(X = x_k) + \sum_{m} y_m \mathbb{P}(Y = y_m) = \mathbb{E}(X) + \mathbb{E}(Y),\end{aligned}$$

ahol azt használtuk, hogy az $\{X = x_k, Y = y_m\}$ események kizáróak, uniójuk $\{X = x_k\}$, és hasonlóképpen a másik tagban az Y esetén.

Következmény: $\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)$.

2.2. Állítás. *Ha X, Y független valószínűségi változók, és X, Y várható értéke létezik, akkor*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{k,m} x_k \cdot y_m \mathbb{P}(X = x_k, Y = y_m) = \\ &\overset{(*)}{=} \sum_{k,m} x_k y_m \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_m) \\ &= \left(\sum_{k} x_k \mathbb{P}(X = x_k)\right) \cdot \left(\sum_{m} y_m \mathbb{P}(Y = y_m)\right) = \mathbb{E}(X) \cdot \mathbb{E}(Y),\end{aligned}$$

ahol a $(*)$ lépésben használtuk a függetlenségnek a diszkrét valószínűségi változókra vonatkozó alakját.

2.4. A szórásnégyzet tulajdonságai

A szórásnégyzet az alábbi tulajdonságokkal rendelkezik.

- (nemnegativitás) $D^2(X) \geq 0$ és $D(X) \geq 0$ mindig teljesül
- (szorzás és eltolás) ha a, b valós számok, X véges szórású valószínűségi változó, akkor

$$D^2(aX + b) = a^2 D^2(X) \quad \Rightarrow \quad D(aX + b) = |a| D(X).$$

- (összeg szórása független esetben) ha az X, Y valószínűségi változók függetlenek és szórásuk létezik, akkor

$$D^2(X+Y) = D^2(X) + D^2(Y) \quad \Rightarrow \quad D(X+Y) = \sqrt{D^2(X) + D^2(Y)}.$$

- van olyan valószínűségi változó, melynek várható értéke véges, de a szórása nem létezik (például: $\mathbb{P}(X = k) = c/k^3$ megfelelő c-vel)

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X-Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X+3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X-3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;

- $D(X+Y) = \sqrt{D^2(X)+D^2(Y)} = \sqrt{1+2^2} = \sqrt{5}$;
- $D(X-Y) = \sqrt{D^2(X)+(-1)^2D^2(Y)} = \sqrt{1+2^2} = \sqrt{5}$;
- $D(2X+3Y) = \sqrt{2^2D^2(X)+3^2D^2(Y)} = \sqrt{4\cdot 1+9\cdot 4} = \sqrt{40}$;
- $D(2X-3Y) = \sqrt{2^2D^2(X)+(-3)^2D^2(Y)} = \sqrt{4\cdot 1+9\cdot 4} = \sqrt{40}$.

2.5. A binomiális eloszlás várható értéke és szórása

2.3. Állítás. *Ha az X valószínűségi változó binomiális eloszlású n renddel és p paraméterrel, azaz*

$$\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k} \qquad (k=0,1,\ldots,n),$$

akkor

$$\mathbb{E}(X) = np; \qquad D(X) = \sqrt{np(1-p)}.$$

Példa. Egy kérdőív egy kérdésére $n = 1000$ megkérdezett közül mindenki a többiektől függetlenül $p = 0,65$ valószínűséggel válaszol. Ekkor a válaszadók számának (melyet jelöljünk X-szel) a várható értéke:

$$\mathbb{E}(X) = np = 1000\cdot 0,65 = 650,$$

míg a szórása:

$$D(X) = \sqrt{np(1-p)} = \sqrt{1000\cdot 0,65\cdot 0,35} = 15,08.$$

Bizonyítás. Legyen X binomiális eloszlású n renddel és p paraméterrel. Azaz: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeresek száma. Vezessük be az alábbi indikátorokat $j = 1,2,\ldots,n$ esetén:

$$X_j = \begin{cases} 1 & \text{ha a } j\text{. kísérlet sikeres;} \\ 0 & \text{ha a } j\text{. kísérlet nem sikeres.} \end{cases}$$

Ekkor X éppen az indikátorok összege (az egyesek száma):

$$X = X_1 + X_2 + \ldots + X_n = \sum_{j=1}^{n} X_j.$$

Mivel bármely j-re

$$\mathbb{E}(X_j) = 1\cdot\mathbb{P}(X=1) + 0\cdot\mathbb{P}(X=0) = p,$$

ezért

$$\mathbb{E}(X) = \mathbb{E}(X_1+X_2+\ldots+X_n) = \mathbb{E}(X_1)+\mathbb{E}(X_2)+\ldots+\mathbb{E}(X_n) = p+p+\ldots+p = np.$$

A szórás kiszámításához: $X_j = X_j^2$, hiszen $0^2 = 0$ és $1^2 = 1$, és már láttuk, hogy $\mathbb{E}(X_j) = p$. Ezért

$$D^2(X_j) = \mathbb{E}(X_j^2) - \mathbb{E}(X_j)^2 = \mathbb{E}(X_j) - \mathbb{E}(X_j)^2 = p - p^2 = p(1-p),$$

Mivel az X_j indikátorok **függetlenek**, és az összegük X:

$$D(X) = \sqrt{D^2(X_1+X_2+\ldots+X_n)} = \sqrt{D^2(X_1)+D^2(X_2)+\ldots+D^2(X_n)} =$$
$$= \sqrt{p(1-p)+p(1-p)+\ldots+p(1-p)} = \sqrt{np(1-p)}.$$

□

A hipergeometrikus eloszlás várható értékére vonatkozó összefüggés szintén indikátorokkal igazolható.

Házi feladat október 14., 8:00-ig Tegyük fel, hogy egy városban az egy nap alatt bekövetkező közlekedési balesetek száma Poisson-eloszlású, és azt is tudjuk, hogy annak valószínűsége, hogy legalább egy baleset történik, 0,4. Mennyi annak valószínűsége, hogy pontosan három baleset történik egy nap alatt? Mennyi az egy nap alatt bekövetkező balesetek számának várható értéke és szórása?

Válasszunk egy $n \geq 100$ számot, és sorsoljunk R-ben (vagy pythonban, más hasonló szoftver segítségével) n elemű Poisson-eloszlású mintát (R-ben rpois), melynek paramétere megegyezik a balesetek számára jellemző paraméterrel a feladat első része alapján. Készítsünk a mintából hisztogramot (hist), számítsuk ki a mintaátlagot (mean), és a korrigált tapasztalati szórást (sd). Hasonlítsuk össze a feladat első részében kiszámított értékeket a mintából kapott tapasztalati értékekkel.