Valószínűségi változók (4. előadás)

események (egy kísérlet eredményéhez **igen vagy nem** tartozik):

- A: holnap lesz csapadék Budapesten
- B: egy véletlenszerűen választott magyar ember Budapesten lakik
- C: egy véletlenszerűen választott magyar ember 500000 forintnál többet keres

valószínűségi változók (egy kísérlet eredményéhez egy **szám** tartozik):

- X: a holnap Budapesten lehulló csapadék mennyisége mm-ben
- Y: egy véletlenszerűen választott magyar ember lakcímének irányító-száma
- Z: egy véletlenszerűen választott magyar ember bruttó havi keresete forintban

Valószínűségi változók: jelölések és definíció

- X: a holnap lehulló csapadék mennyisége mm-ben $\rightarrow \mathbb{P}(X \leq 5)$, azaz mennyi annak valószínűsége, hogy holnap **legfeljebb** 5 **mm** csapadék esik;
- Y: egy véletlenszerűen választott magyar ember lakcímének irányítószáma $\rightarrow \mathbb{P}(Y = 2092)$, azaz mennyi annak valószínűsége, hogy a kiválasztott ember Budakeszin lakik, **irányítószáma pontosan** 2092
- Z: egy véletlenszerűen választott magyar ember bruttó havi keresete forintban $\rightarrow \mathbb{P}(Z \leq 500000)$, azaz mennyi annak valószínűsége, hogy a kiválasztott ember **legfeljebb bruttó** 500000 **forintot** keres havonta

Definíció (Valószínűségi változó)

Egy $X : \Omega \rightarrow \mathbb{R}$ függvény **valószínűségi változó***, ha tetszőleges t valós számra teljesül, hogy*

$$\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A},$$

azaz tetszőleges t valós számra a $\mathbb{P}(X \leq t)$ valószínűség értelmes.

Valószínűségi változó: példa

Valakinek három gyereke születik. Legyen X a fiúk száma. Ekkor az összes lehetőség halmaza Ω, és $X : \Omega \to \mathbb{R}$ függvény az alábbi módon:

$$\Omega = \{FFF, FFL, FLF, FLL, LFF, LFL, LLF, LLL\};$$
$$X(LLL) = 0; \qquad X(LLF) = X(LFL) = X(FLL) = 1;$$
$$X(FFL) = X(FLF) = X(LFF) = 2; \qquad X(FFF) = 3.$$

Az X **valószínűségi változó** **lehetséges értékei**:

$$\{0, 1, 2, 3\} \to \text{véges halmaz}$$

A lehetséges értékekhez tartozó valószínűségek, feltéve, hogy a gyerekek egymástól függetlenül $1/2$ valószínűséggel fiúk:

$$\mathbb{P}(X = 0) = 1/8, \qquad \mathbb{P}(X = 1) = 3/8,$$
$$\mathbb{P}(X = 2) = 3/8, \qquad \mathbb{P}(X = 3) = 1/8.$$

Diszkrét valószínűségi változó és eloszlása: definíciók

Definíció

Egy $X : \Omega \to \mathbb{R}$ *függvény* **valószínűségi változó**, *ha tetszőleges* t *valós számra*

$$\{\omega \in \Omega : X(\omega) \le t\} \in \mathcal{A},$$

azaz tetszőleges t *valós számra a* $\mathbb{P}(X \le t)$ *valószínűség értelmes.*

Ilyenkor a $\mathbb{P}(X = t)$ valószínűség is tetszőleges t-re értelmes.

Az $X : \Omega \to \mathbb{R}$ valószínűségi változó **diszkrét**, ha **lehetséges értékeinek halmaza véges vagy megszámlálhatóan végtelen**.

Diszkrét valószínűségi változó és eloszlása: definíciók

Definíció

Egy $X : \Omega \to \mathbb{R}$ függvény **valószínűségi változó**, *ha tetszőleges t valós számra*

$$\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A},$$

azaz tetszőleges t valós számra a $\mathbb{P}(X \leq t)$ valószínűség értelmes.

Ilyenkor a $\mathbb{P}(X = t)$ valószínűség is tetszőleges t-re értelmes.

Az $X : \Omega \to \mathbb{R}$ valószínűségi változó **diszkrét**, ha **lehetséges értékeinek halmaza véges vagy megszámlálhatóan végtelen**.

Legyenek az X **diszkrét valószínűségi változó** lehetséges értékei:

$$\{x_1, x_2, \ldots\}, \qquad \text{és } p_k = \mathbb{P}(X = x_k) \qquad (k = 1, 2, \ldots).$$

Ekkor az $(x_1, p_1), (x_2, p_2), \ldots$ sorozat az X valószínűségi változó **eloszlása**. Ilyenkor

$$p_k \geq 0 \text{ minden } k\text{-ra, és } \sum_{k=1}^{\infty} p_k = 1,$$

Diszkrét valószínűségi változó eloszlása

Három gyerek. Valakinek három gyereke születik, X a fiúk száma, feltesszük, hogy mind a $2^3 = 8$ lehetőség egyformán valószínű. Ekkor X lehetséges értékei:

$$\{0, 1, 2, 3\} \to \text{véges halmaz} \to X \textbf{ diszkrét}.$$

Ahogy láttuk:

$$\mathbb{P}(X = 0) = \frac{1}{8}; \quad \mathbb{P}(X = 1) = \frac{3}{8}; \quad \mathbb{P}(X = 2) = \frac{3}{8}; \quad \mathbb{P}(X = 3) = \frac{1}{8}.$$

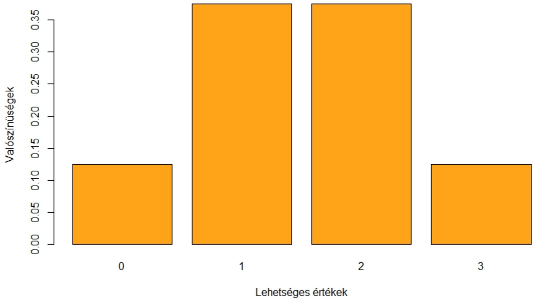
Mindezek alapján X **eloszlása** az alábbi sorozat:

$$(0, 1/8), \qquad (1, 3/8), \qquad (2, 3/8), \qquad (3, 1/8).$$

Szabályos kockadobás. Egyszer dobunk szabályos dobókockával, jelölje Y a dobott számot. Ekkor Y **diszkrét**, és az **eloszlása**:

$$(1, 1/6), \quad (2, 1/6), \quad (3, 1/6), \quad (4, 1/6), \quad (5, 1/6), \quad (6, 1/6).$$

Példa: a fiúk számának eloszlása



A fiúk számának eloszlása: a lehetséges értékek:

$$\{0, 1, 2, 3\}$$

és a hozzájuk tartozó valószínűségek:

$$1/8, \quad 3/8, \quad 3/8, \quad 1/8.$$

Valószínűségi változó eloszlása

Definíció (Valószínűségi változó)

Egy $X : \Omega \to \mathbb{R}$ *függvény* **valószínűségi változó**, *ha tetszőleges* t *valós számra*

$$\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A},$$

vagyis tetszőleges t *valós számra a* $\mathbb{P}(X \leq t)$ *valószínűség értelmes.*

Nem feltétlenül diszkrét $X : \Omega \to \mathbb{R}$ valószínűségi változó **eloszlása**: Q_X mérték, melyre

$$Q_X(B) = \mathbb{P}(X \in B),$$

ahol $B \subseteq \mathbb{R}$ megfelelő feltételeket teljesítő halmaz (Borel-halmaz).

Például: $B = [a, b]$ intervallum esetén

$$Q_X([a, b]) = \mathbb{P}(X \in [a, b]) = \mathbb{P}(a \leq X \leq b).$$

Vagy $B = (-\infty, t]$ félegyenes esetén

$$Q_X((-\infty, t]) = \mathbb{P}(X \in (-\infty, t]) = \mathbb{P}(X \leq t).$$

Diszkrét valószínűségi változók várható értéke

- Szerencsejáték igazságos ára: a nyeremény várható értéke, például ha 1/1000 valószínűséggel 1000000 forintot nyerünk (különben semmit), akkor

Diszkrét valószínűségi változók várható értéke

- Szerencsejáték igazságos ára: a nyeremény várható értéke, például ha $1/1000$ valószínűséggel 1000000 forintot nyerünk (különben semmit), akkor 1000.
- Annyit nyerünk, amennyit egy szabályos dobókockával dobunk. A nyeremény várható értéke:

Diszkrét valószínűségi változók várható értéke

- Szerencsejáték igazságos ára: a nyeremény várható értéke, például ha 1/1000 valószínűséggel 1000000 forintot nyerünk (különben semmit), akkor 1000.
- Annyit nyerünk, amennyit egy szabályos dobókockával dobunk. A nyeremény várható értéke:

$$\frac{1}{6}(1+2+\ldots+6)=\frac{21}{6}=3,5.$$

- Ha szabálytalan a kocka, például az 1 helyett is 6 van:

Diszkrét valószínűségi változók várható értéke

- Szerencsejáték igazságos ára: a nyeremény várható értéke, például ha 1/1000 valószínűséggel 1000000 forintot nyerünk (különben semmit), akkor 1000.
- Annyit nyerünk, amennyit egy szabályos dobókockával dobunk. A nyeremény várható értéke:

$$\frac{1}{6}(1+2+\ldots+6)=\frac{21}{6}=3,5.$$

- Ha szabálytalan a kocka, például az 1 helyett is 6 van:

$$\frac{1}{6}(6+2+3+4+5+6)=\frac{1}{6}(2+3+4+5)+\frac{1}{3}\cdot 6=\frac{13}{3}=4,33.$$

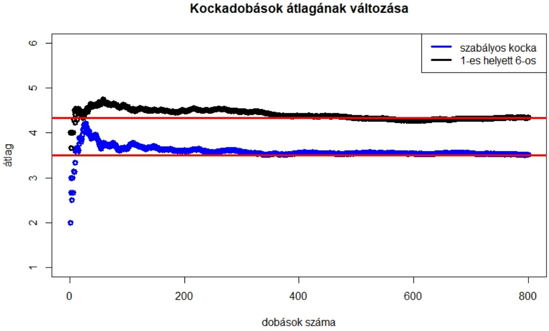
Diszkrét valószínűségi változók várható értéke

- Szerencsejáték igazságos ára: a nyeremény várható értéke, például ha $1/1000$ valószínűséggel 1000000 forintot nyerünk (különben semmit), akkor 1000.
- Annyit nyerünk, amennyit egy szabályos dobókockával dobunk. A nyeremény várható értéke:
$$\frac{1}{6}(1+2+\ldots+6)=\frac{21}{6}=3,5.$$
- Ha szabálytalan a kocka, például az 1 helyett is 6 van:

$$\frac{1}{6}(6+2+3+4+5+6)=\frac{1}{6}(2+3+4+5)+\frac{1}{3}\cdot 6=\frac{13}{3}=4,33.$$

A lehetséges értékeket megszorozzuk a hozzájuk tartozó valószínűséggel, és ezeket összeadjuk.

Kockadobások átlaga



A dobások átlagának változása a dobások számának növelésével, szabályos kocka esetén, illetve ha 1 helyett is 6 szerepel.

Diszkrét valószínűségi változók várható értéke

Definíció (Várható érték, diszkrét eset)

Legyen $X : \Omega \to \mathbb{R}$ *olyan diszkrét valószínűségi változó, melynek eloszlása* $(x_1, p_1), (x_2, p_2), \ldots$, *azaz* $\mathbb{P}(X = x_i) = p_i$, *ahol* $i = 1, 2, \ldots$. *Ekkor* X *várható értéke:*

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i, \qquad \text{ha} \quad \sum_{i=1}^{\infty} |x_i| p_i < \infty.$$

Diszkrét valószínűségi változók várható értéke

Definíció (Várható érték, diszkrét eset)

Legyen $X : \Omega \to \mathbb{R}$ *olyan diszkrét valószínűségi változó, melynek eloszlása* $(x_1, p_1), (x_2, p_2), \ldots,$ *azaz* $\mathbb{P}(X = x_i) = p_i$*, ahol* $i = 1, 2, \ldots$. *Ekkor* X *várható értéke:*

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i, \qquad \text{ha} \quad \sum_{i=1}^{\infty} |x_i| p_i < \infty.$$

Példa: három gyerek. Legyen X a fiúk száma a három gyerek közül. Ekkor

$$\mathbb{E}(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1,5.$$

Példa: szabályos kockadobás. Legyen Y egy szabályos dobókockával dobott szám. Ekkor

$$\mathbb{E}(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3,5.$$

Diszkrét valószínűségi változók várható értéke

- (elfajult eloszlás) Ha $X = c$ fennáll 1 valószínűséggel: $\mathbb{E}(X) = c \cdot \mathbb{P}(X = c) = c$.
- (korlátosság) Ha $a \leq X \leq b$ valamely $a < b$ számokra, akkor $a \leq \mathbb{E}(X) \leq b$.
- (egyenletes eloszlás) Ha az $x_1, x_2, \ldots, x_n$ számok mindegyikének $1/n$ a valószínűsége, akkor a várható érték a számok számtani közepe: $\overline{x} = \frac{1}{n}(x_1 + \ldots + x_n)$.
- (indikátor) Legyen $\mathbb{I}_A$ az A esemény indikátora, vagyis 1, ha A bekövetkezik, és 0 különben. Ekkor $\mathbb{E}(\mathbb{I}_A) = 1 \cdot \mathbb{P}(\mathbb{I}_A = 1) = \mathbb{P}(A)$.
- (összeg) Ha X, Y valószínűségi változók és $X + Y$ várható értéke létezik, akkor
$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$
- (konstans kiemelése) Ha X várható értéke létezik, és c tetszőleges valós szám, akkor
$$\mathbb{E}(cX) = c\mathbb{E}(X).$$

Függvény várható értéke

Állítás

Legyen $X : \Omega \to \mathbb{R}$ diszkrét valószínűségi változó, melynek lehetséges értékei $x_1, x_2, \ldots$, továbbá $\mathbb{P}(X = x_k) = p_k$ teljesül $k \geq 1$ esetén. Legyen továbbá $g : \mathbb{R} \to \mathbb{R}$ függvény. Ekkor

$$\mathbb{E}(g(X)) = g(x_1)p_1 + g(x_2)p_2 + g(x_3)p_3 \ldots,$$

ha ez a várható érték létezik.

Például: legyen Y egy szabályos dobókockával dobott szám. Ekkor $g(x) = x^2$-tel:

$$\mathbb{E}(Y^2) = 1 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^5 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6} = 15,17.$$

Függvény várható értéke

Állítás

Legyen $X : \Omega \to \mathbb{R}$ diszkrét valószínűségi változó, melynek lehetséges értékei $x_1, x_2, \ldots$, továbbá $\mathbb{P}(X = x_k) = p_k$ teljesül $k \geq 1$ esetén. Legyen továbbá $g : \mathbb{R} \to \mathbb{R}$ függvény. Ekkor

$$\mathbb{E}(g(X)) = g(x_1)p_1 + g(x_2)p_2 + g(x_3)p_3 \ldots,$$

ha ez a várható érték létezik.

Például: legyen Y egy szabályos dobókockával dobott szám. Ekkor $g(x) = x^2$-tel:

$$\mathbb{E}(Y^2) = 1 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^5 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6} = 15,17.$$

Valószínűségi változó szórása

Lehetséges motiváció: nem mindegy, hogy a buszok ütemesen (szabályosan) tíz percenként érkeznek, vagy a követési idő várható értéke tíz perc, de hol öt, hol tizenöt percenként jönnek; egy mérőeszköztől a mérési hiba, vagyis a mérés bizonytalansága is fontos, például nem mindegy, hogy adott pontosság eléréséhez hány mérést kell átlagolni.

Definíció (Szórásnégyzet (variancia))

Legyen $X : \Omega \to \mathbb{R}$ *valószínűségi változó, melyre* $\mathbb{E}(X^2)$ *létezik. Ekkor* X *szórásnégyzete:*

$$D^2(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big).$$

Valószínűségi változó szórása

Lehetséges motiváció: nem mindegy, hogy a buszok ütemesen (szabályosan) tíz percenként érkeznek, vagy a követési idő várható értéke tíz perc, de hol öt, hol tizenöt percenként jönnek; egy mérőeszköztől a mérési hiba, vagyis a mérés bizonytalansága is fontos, például nem mindegy, hogy adott pontosság eléréséhez hány mérést kell átlagolni.

Definíció (Szórásnégyzet (variancia))

Legyen $X : \Omega \to \mathbb{R}$ valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor X szórásnégyzete:

$$D^2(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

Definíció (Szórás (standard deviation))

Legyen $X : \Omega \to \mathbb{R}$ valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor X szórásnégyzete:

$$D(X) = \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}.$$

A szórás kiszámítása

Állítás

Legyen X olyan diszkrét valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor

$$D^2(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2.$$

A szórás kiszámítása

Állítás

Legyen X olyan diszkrét valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor

$$D^2(X) = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2.$$

Bizonyítás.

$$\begin{aligned} D^2(X) &= \mathbb{E}\Big((X - \mathbb{E}X)^2\Big) = \mathbb{E}\Big(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\Big) = \\ &= \mathbb{E}(X^2) - \mathbb{E}\big(2X\mathbb{E}(X)\big) + \mathbb{E}(X)^2 = \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2, \end{aligned}$$

ahol felhasználtuk, hogy összeg várható értéke a várható értékek összege, illetve hogy a konstans szorzó kiemelhető.

A szórás kiszámítása

Állítás

Legyen X olyan diszkrét valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor

$$D^2(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2.$$

Bizonyítás.

$$\begin{aligned} D^2(X) &= \mathbb{E}\Big((X - \mathbb{E}X)^2\Big) = \mathbb{E}\Big(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\Big) = \\ &= \mathbb{E}(X^2) - \mathbb{E}\big(2X\mathbb{E}(X)\big) + \mathbb{E}(X)^2 = \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2, \end{aligned}$$

ahol felhasználtuk, hogy összeg várható értéke a várható értékek összege, illetve hogy a konstans szorzó kiemelhető.

Megjegyzés: az $x_1, x_2, \ldots, x_n$ számok tapasztalati szórása

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2}.$$

Diszkrét valószínűségi változó szórása

Állítás

Legyen X olyan diszkrét valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik. Ekkor

$$D^2(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2.$$

Állítás (A szórás kiszámítása egész értékek esetén)

Legyen X olyan diszkrét valószínűségi változó, melyre $\mathbb{E}(X^2)$ létezik, és melynek lehetséges értékei nemnegatív egészek. Ekkor

$$D^2(X) = \sum_{k=0}^{\infty} k^2 \mathbb{P}(X = k) - \mathbb{E}(X)^2 = \sum_{k=0}^{\infty} k^2 \mathbb{P}(X = k) - \left[\sum_{k=0}^{\infty} k \mathbb{P}(X = k)\right]^2.$$

Diszkrét valószínűségi változó szórása

Legyen továbbra is X a fiúk száma három gyerek közül:

$\mathbb{P}(X=0)=1/8;\quad \mathbb{P}(X=1)=3/8;\quad \mathbb{P}(X=2)=3/8;\quad \mathbb{P}(X=3)=1/8.$

Diszkrét esetben így számolhatunk:

$$\mathbb{E}(X^2)=\sum_{k=0}^{3}k^2\mathbb{P}(X=k)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+4\cdot\frac{3}{8}+9\cdot\frac{1}{8}=\frac{24}{8}=3.$$

Ebből és a korábbi számolásból

$$D^2(X)=\mathbb{E}(X^2)-\left[\mathbb{E}(X)\right]^2=3-1,5^2=3-2,25=0,75=\frac{3}{4}.$$

Végül pedig a fiúk számának szórása:

$$D(X)=\sqrt{\frac{3}{4}}=0,866.$$

A kockadobás szórása

Legyen X egy szabályos dobókockával dobott szám. Ekkor

$$\mathbb{E}(X^2) = \frac{1}{6}\cdot 1^2 + \frac{1}{6}\cdot 2^2 + \frac{1}{6}\cdot 3^2 + \frac{1}{6}\cdot 4^2 + \frac{1}{6}\cdot 5^2 + \frac{1}{6}\cdot 6^2 = \frac{91}{6}.$$

Másrészt

$$\mathbb{E}(X) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = \frac{7}{2}.$$

Ebből

$$D^2(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = 2,92.$$

A kockadobás szórása: $D(X) = \sqrt{2,92} = 1,71.$

Általában n oldalú dobókocka esetén: $D(X) = \sqrt{\frac{n^2-1}{12}}.$

Binomiális eloszlás: példa

Egy munkahelyi csapatban **hatan** dolgoznak együtt.
Tegyük fel, hogy egy tetszőleges napon **egymástól függetlenül** mindannyian $p = 0,03$ valószínűséggel **hiányoznak**.
Mennyi a valószínűsége, hogy egy munkanapon **pontosan ketten hiányoznak a csapatból?**

Binomiális eloszlás: példa

Egy munkahelyi csapatban **hatan** dolgoznak együtt.
Tegyük fel, hogy egy tetszőleges napon **egymástól függetlenül** mindannyian $p = 0,03$ valószínűséggel **hiányoznak**.
Mennyi a valószínűsége, hogy egy munkanapon **pontosan ketten hiányoznak a csapatból?**

A B C D E F

néhány jó lehetőség és a valószínűsége:

szorzás

0,03	0,03	0,97	0,97	0,97	0,97	$\to 0,03^2 \cdot 0,97^4$
0,03	0,97	0,03	0,97	0,97	0,97	$\to 0,03^2 \cdot 0,97^4$
0,03	0,97	0,97	0,03	0,97	0,97	$\to 0,03^2 \cdot 0,97^4$

...

0,97	0,03	0,03	0,97	0,97	0,97	$\to 0,03^2 \cdot 0,97^4$

...

0,97	0,97	0,97	0,97	0,03	0,03	$\to 0,03^2 \cdot 0,97^4$

Binomiális eloszlás: példa

Egy munkahelyi csapatban **hatan** dolgoznak együtt.
Tegyük fel, hogy egy tetszőleges napon **egymástól függetlenül**
mindannyian $p = 0,03$ valószínűséggel **hiányoznak**.
Mennyi a valószínűsége, hogy egy munkanapon
pontosan ketten hiányoznak a csapatból?

$0,97$	$0,03$	$0,03$	$0,97$	$0,97$	$0,97$

$\to 0,03^2 \cdot 0,97^4$

a jó lehetőségek száma, azaz
hányféleképpen választhatjuk ki a két hiányzót:

egy jó lehetőség valószínűsége:
tehát a valószínűség:

Binomiális eloszlás: példa

Egy munkahelyi csapatban **hatan** dolgoznak együtt.
Tegyük fel, hogy egy tetszőleges napon **egymástól függetlenül** mindannyian $p = 0,03$ valószínűséggel **hiányoznak**.
Mennyi a valószínűsége, hogy egy munkanapon **pontosan ketten hiányoznak a csapatból?**

$0,97$	$0,03$	$0,03$	$0,97$	$0,97$	$0,97$

$\to 0,03^2 \cdot 0,97^4$

a jó lehetőségek száma, azaz
hányféleképpen választhatjuk ki a két hiányzót: $\binom{6}{2}$

egy jó lehetőség valószínűsége: $0,03^2 \cdot 0,97^4$

tehát a valószínűség:

$$\boxed{\mathbb{P}(\textbf{pontosan két hiányzó}) = \binom{6}{2} \cdot 0,03^2 \cdot 0,97^4 = 1,2\%.}$$

Binomiális eloszlás

- n független kísérletet végzünk;
- mindegyik p valószínűséggel sikerül;
- X a sikeres kísérletek száma.

Például:

- Visszatevéses mintavétel, n húzás, p a fekete golyók aránya.
- Egy felmérésben $n = 1500$ embert kérdezünk meg, egy adott kérdésre mindenki egymástól függetlenül $p = 0,8$ valószínűséggel válaszol. A válaszok száma binomiális eloszlású.
- Egy biztosító $n = 60000$ ügyfelének mindegyike egymástól függetlenül $p = 0,0001$ valószínűséggel okoz balesetet egy adott évben. A balesetet okozó ügyfelek száma binomiális eloszlású.
- Tegyük fel, hogy a nyár $n = 92$ napjának mindegyikén egymástól függetlenül $p = 0,02$ valószínűséggel lesz jégeső egy adott helyen. A nyári jégesős napok száma binomiális eloszlású.

Binomiális eloszlás

- n **független** kísérletet végzünk;
- mindegyik p **valószínűséggel** sikerül;
- X a sikeres kísérletek száma.

Mennyi a valószínűsége, hogy **pontosan k kísérlet sikerül**, azaz $X = k$? Ahogyan a korábbi példában láttuk:

- A jó lehetőségek száma, azaz hányféleképpen választhatjuk ki, hogy melyik k kísérlet sikeres: $\binom{n}{k}$.
- Egy jó lehetőség valószínűsége: $p^k(1-p)^{n-k}$, hiszen a kísérletek függetlenek, ezért az együttes bekövetkezés (metszet) valószínűsége a valószínűségek szorzata, és k kísérlet sikerül, a többi $n-k$ nem.
- Mivel minden jó lehetőség ugyanolyan valószínű, az $X = k$ valószínűsége a lehetőségek számának és egy lehetőség valószínűségének szorzata:

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

Binomiális eloszlás: definíció

- n független kísérletet végzünk;
- mindegyik p valószínűséggel sikerül;
- X a sikeres kísérletek száma.

Az X valószínűségi változó **binomiális eloszlású** n renddel és p paraméterrel, ha lehetséges értékei:

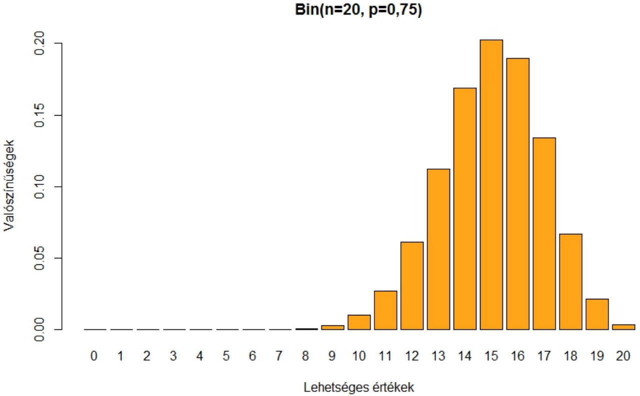
$$0, 1, 2, \ldots, n,$$

és minden $0 \leq k \leq n$ egészre

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

($n \geq 1$ egész, $0 < p < 1$.) Jelölés: $\mathrm{Bin}(n, p)$.

Példa: binomiális eloszlás



Binomiális eloszlás, $n = 20$, $p = 0,75$. Vízszintes tengely: lehetséges értékek, azaz $k = 0, 1, \ldots, 20$, oszlopok magassága: a $\mathbb{P}(X = k)$ valószínűségek.

Példa: binomiális eloszlás

Egy felmérésben $n = 1500$ embert kérdezünk meg. Egy adott kérdésre minden résztvevő **egymástól függetlenül** $p = 0,8$ valószínűséggel válaszol. Jelölje X, hogy hányan válaszoltak erre a kérdésre. Ekkor

- X **binomiális eloszlású** $n = 1500$ renddel és $p = 0,8$ paraméterrel.
- Tetszőleges $0 \leq k \leq 1500$ esetén

$$\mathbb{P}(k \text{ válasz}) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} =$$
$$= \binom{1500}{k} 0,8^k \cdot 0,2^{1500-k}.$$

- Például annak valószínűsége, hogy pontosan $k = 1200$-an válaszolnak a kérdésre:

$$\mathbb{P}(1200 \text{ válasz}) = \mathbb{P}(X = 1200) = \binom{1500}{1200} 0,8^{1200} \cdot 0,2^{300} = 2,57\%.$$

A binomiális eloszlás várható értéke és szórása

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

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

akkor X **várható értéke**, illetve **szórása**:

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

A binomiális eloszlás várható értéke: bizonyítás

Állítás

Legyen X valószínűségi változó n renddel és p paraméterrel. Ekkor X várható értéke np.

A binomiális eloszlásra úgy is gondolhattunk, mint a sikeres kísérletek számára, n független, p valószínűséggel bekövetkező eseményből. Tekintsük az alábbi indikátor valószínűségi változókat minden $j = 1, 2, \ldots, n$-re:

$$\mathbb{I}_j = \begin{cases} 1, & \text{ha a } j\text{. kísérlet sikerül;} \\ 0, & \text{különben.} \end{cases}$$

A binomiális eloszlás várható értéke: bizonyítás

Állítás

Legyen X valószínűségi változó n renddel és p paraméterrel. Ekkor X várható értéke np.

A binomiális eloszlásra úgy is gondolhattunk, mint a sikeres kísérletek számára, n független, p valószínűséggel bekövetkező eseményből. Tekintsük az alábbi indikátor valószínűségi változókat minden $j = 1, 2, \dots, n$-re:

$$\mathbb{I}_j = \begin{cases} 1, & \text{ha a } j\text{. kísérlet sikerül;} \\ 0, & \text{különben.} \end{cases}$$

Ekkor az $\mathbb{I}_j$ indikátorok összege éppen X lesz. Így a várható érték additív tulajdonsága alapján

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{n} \mathbb{I}_j\right) = \sum_{j=1}^{n} \mathbb{E}(\mathbb{I}_j) = \sum_{j=1}^{n} 1 \cdot \mathbb{P}(\mathbb{I}_j = 1) = np. \qquad \square$$

Példa: binomiális eloszlás

Egy felmérésben n = 1500 embert kérdezünk meg. Egy adott kérdésre minden résztvevő **egymástól függetlenül** p = 0,8 valószínűséggel válaszol. Jelölje X, hogy hányan válaszoltak erre a kérdésre. Ekkor

- A válaszadók számának **várható értéke**:

$$\mathbb{E}(X) = np = 1500 \cdot 0,8 = 1200.$$

- A válaszadók számának **szórása**:

$$D(X) = \sqrt{np(1-p)} = \sqrt{1500 \cdot 0,8 \cdot 0,2} = 15,5.$$

A binomiális eloszlás közelítése

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**:

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

Annak valószínűsége, hogy **pontosan k ügyfél okoz balesetet**:

$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

A binomiális eloszlás közelítése

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**:

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

Annak valószínűsége, hogy **pontosan k ügyfél okoz balesetet**:

$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

$$= \frac{100000 \cdot 99999 \cdot \ldots (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx$$

A binomiális eloszlás közelítése

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**:

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

Annak valószínűsége, hogy **pontosan k ügyfél okoz balesetet**:

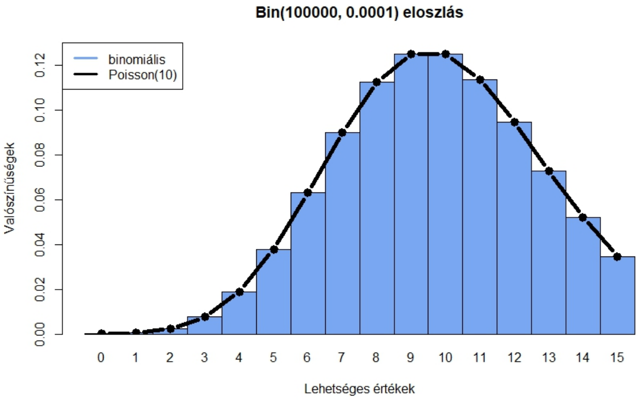
$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

$$= \frac{100000 \cdot 99999 \cdot \ldots (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx$$

$$\approx \frac{100000^k \cdot 0,0001^k}{k!} \left(1 - \frac{10}{100000}\right)^{100000} \approx \frac{10^k}{k!} e^{-10},$$

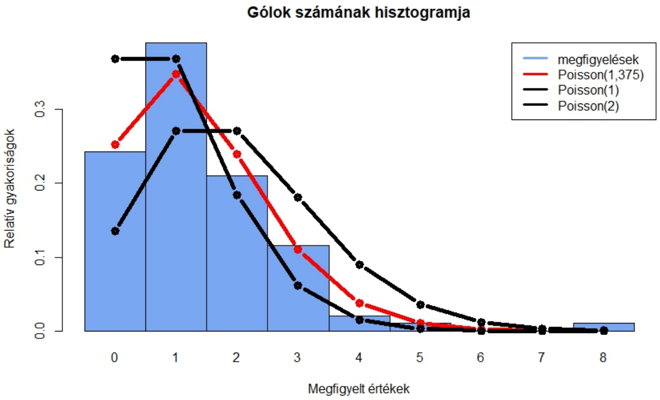
felhasználva, hogy $\lim_{n\to\infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$ tetszőleges $x > 0$-ra.

A binomiális eloszlás közelítése



Binomiális eloszlás $n = 100000$ renddel és $p = 0,0001$ paraméterrel (vízszintes tengely: k, oszlopok magassága: $\mathbb{P}(X = k)$).
Feketével a $\frac{10^k}{k!}e^{-10}$ függvény (ez lesz a Poisson(10)-eloszlás).

Poisson-eloszlás



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

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt**:
  - lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
  - 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

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt**:
  - lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
  - 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:

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt**:
  - lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
  - 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.

Poisson-eloszlás: definíció

- ha n független kísérletből mindegyik p valószínűséggel sikeres, ahol n „nagy” és p „kicsi”: $\lambda = np$ a sikeres kísérletek számának várható értéke;
- annak valószínűsége, hogy pontosan k kísérlet sikeres, $\frac{\lambda^k}{k!}e^{-\lambda}$-val közelíthető (a számolás és az ábra alapján);
- ez a gyakran használt Poisson-eloszlás **ritkán bekövetkező események számának** modellezésére.

Definíció

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

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

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

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

Poisson-eloszlás: példa

Az X valószínűségi változó λ **paraméterű Poisson-eloszlású**, ha lehetséges értékei:

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

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

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

Példa. Tegyük fel, hogy egy városban az egy nap alatt bekövetkező autóbalesetek száma Poisson-eloszlású, és **várható értéke** $3,61$. Ekkor az egy nap alatt bekövetkező autóbalesetek számának **szórása**:

$$D(X) = \sqrt{3,61} = 1,9.$$

Annak valószínűsége, hogy pontosan 5 baleset lesz:

$$\mathbb{P}(X = 5) = \frac{3,61^5}{5!} e^{-5} = 14\%.$$

Binomális és Poisson-eloszlás az R-ben

- ha X binomiális eloszlású n renddel és p paraméterrel, akkor
$$\mathbb{P}(X = k): \qquad \mathtt{dbinom}(\mathtt{k}, \mathtt{size} = \mathtt{n}, \mathtt{prob} = \mathtt{p})$$
és
$$\mathbb{P}(X \leq k): \qquad \mathtt{pbinom}(\mathtt{k}, \mathtt{size} = \mathtt{n}, \mathtt{prob} = \mathtt{p}).$$
- Továbbá az
```
minta<-rbinom(r, size=n, prob=p)
```
eredményeképpen a „minta” vektorba r darab *független* adott binomiális eloszlású valószínűségi változó kerül.
- ha X Poisson-eloszlású λ paraméterrel, akkor
$$\mathbb{P}(X = k): \qquad \mathtt{dpois}(\mathtt{k}, \mathtt{lambda} = \lambda)$$
és
$$\mathbb{P}(X \leq k): \qquad \mathtt{ppois}(\mathtt{k}, \mathtt{lambda} = \lambda).$$
- Továbbá az
```
minta<-rpois(r, lambda=λ)
```
eredményeképpen a „minta” vektorba r darab *független* λ paraméterű eloszlású valószínűségi változó kerül.

Példa R-ben

```
> minta=rbinom(100, size=100000, prob=0.0001)
> hist(minta)
> dbinom(8, size=100000, prob=0.0001)
[1] 0.1126013
> minta=rpois(100, lambda=3.61)
> hist(minta)
> mean(minta)
[1] 3.91
> dpois(5, lambda=3.61)
[1] 0.1382139
```