

Valószínűségszámítás előadás, 3. hét, 2020. szeptember 23.
 Szitaformula, feltételes valószínűség, függetlenség

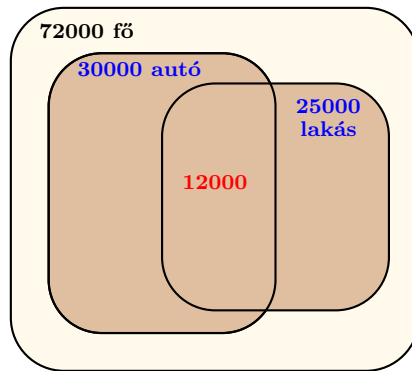
1. Szitaformula

Emlékeztetőül: a valószínűség egyik tulajdonsága az volt, hogy ha $A_1, A_2, \dots \in \mathcal{A}$ páronként kizáró események (azaz $A_i \cap A_j = \emptyset$ tetszőleges $i \neq j$ esetén), akkor

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Kérdés, hogyan számolhatunk, ha nem tételezzük fel, hogy az események páronként kizáróak.

Példa. Egy 72000 fős városban harmincezer embernek van autója, huszonötezernek lakása, tizenkétezernek mindkettő. Mennyi a valószínűsége, hogy egy véletlenszerűen választott lakos az autó és a lakás közül legalább az egyikkel rendelkezik? Feltesszük, hogy minden lakost azonos valószínűséggel választunk.



Legyen A az az esemény, hogy egy véletlenszerűen választott lakosnak van autója, L az, hogy van lakása. Ekkor így számolhatunk az ábra alapján:

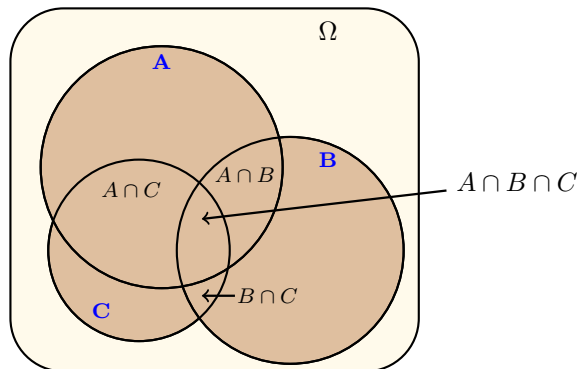
$$\mathbb{P}(A \cup L) = \mathbb{P}(A) + \mathbb{P}(L) - \mathbb{P}(A \cap L) = \frac{30000}{72000} + \frac{25000}{72000} - \frac{12000}{72000} = \frac{43000}{72000} = 59\%.$$

Egy másik lehetséges indoklás ugyanerre, ami általánosan is működik:

$$\mathbb{P}(A \cup L) = \mathbb{P}(A \cup (L \setminus A)) = \mathbb{P}(A) + \mathbb{P}(L) - \mathbb{P}(A \cap L) = 59\%,$$

hiszen A és $L \setminus A$ kizáró események, valamint $L \setminus A$ és $L \cap A$ is kizáróak, így $\mathbb{P}(L \setminus A) + \mathbb{P}(L \cap A) = \mathbb{P}(L)$.

Ha azt kérdeznénk, hogy három esemény esetén mennyi a valószínűsége, hogy legalább az egyik bekövetkezik, hasonlóan számolhatunk:



Ekkor az események valószínűségeit összeadva kétszer számoltuk a metszetekben lévő elemi eseményeket, ha azonban ezeknek levonjuk a valószínűségét, a hármas metszetben lévő elemeket egyszer sem vettük figyelembe, ennek valószínűségét még hozzá kell adnunk:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Ezek alapján megfogalmazhatjuk az alábbi állítást.

1.1. Állítás. (a) **Szitaformula két eseményre.** *Annak valószínűsége, hogy A és B események közül legalább az egyik bekövetkezik:*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(b) **Szitaformula három eseményre.** *Annak valószínűsége, hogy A, B és C események közül legalább az egyik bekövetkezik:*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

(c) **Szitaformula általánosan:** *az A_1, \dots, A_n események uniójának (vagyis annak, hogy legalább az egyik bekövetkezik) a valószínűsége:*

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \\ &\quad - \dots - \mathbb{P}(A_{n-1} \cap A_n) + \mathbb{P}(A_1 \cap A_2 \cap A_3) + \dots \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}). \end{aligned}$$

Emlékeztetőül: a valószínűség egyik tulajdonsága az volt, hogy ha A_1, A_2, \dots páronként kizáró események, vagyis semelyik kettő nem következhet be, akkor az uniójuk valószínűsége a valószínűségeik összege. Ezt tekinthetjük a szitaformula speciális esetének is, hiszen ilyenkor a képletben minden olyan tag, ahol legalább két esemény metszete szerepel, 0 lesz.

Példa. Egy kislíú Kinder-figurákat gyűjt. Tízféle Kinder-figura van, a tojásokban a többtől függetlenül mindegyik azonos valószínűséggel található. Mennyi a valószínűsége, hogy 20 Kinder-tojás felbontása után mind a tíz figurából lesz legalább egy példánya?

Legyen A_i az az esemény, hogy az i . figura nincs meg a kislíúnak ($i = 1, \dots, 10$). Ekkor annak a valószínűsége, hogy mind a 10 figurából van neki legalább egy: $1 - \mathbb{P}(A_1 \cup \dots \cup A_{10})$.

Kezdjük el felírni a szitaformulát. Az első tag:

$$\mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_{10}) = 10 \cdot \frac{9^{20}}{10^{20}},$$

hiszen A_j azt jelenti, hogy a j . figura nem fordulhat elő, a többi 9 közül bármelyik bármelyik tojásban lehet, az összes, egyformán valószínű eset száma pedig 10^{20} (gondolhatunk a visszatéves mintavételre).

Hasonlóképpen a következő tag, amit le kell vonnunk:

$$\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \dots + \mathbb{P}(A_9 \cap A_{10}) = \binom{10}{2} \cdot \frac{8^{20}}{10^{20}},$$

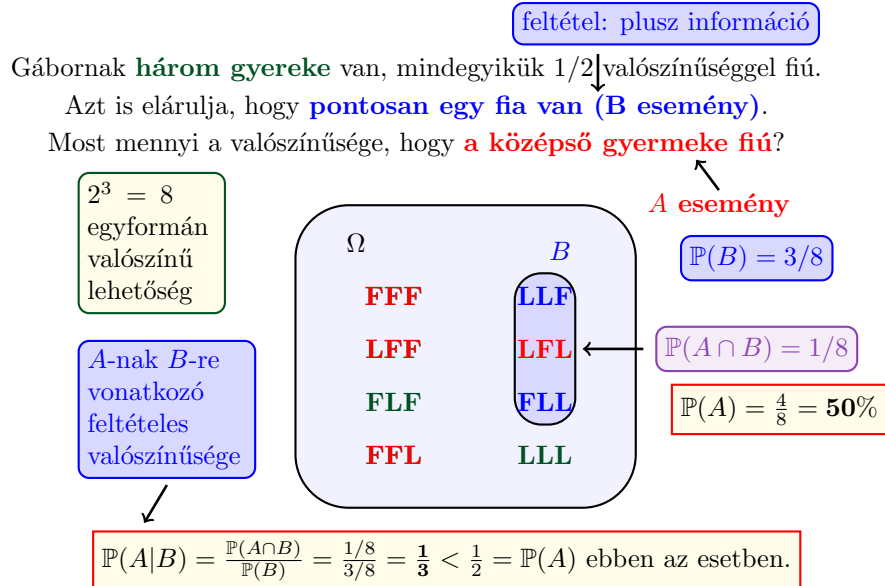
hiszen a tagok száma megegyezik az $1, 2, \dots, 10$ közül kiválasztható párok számával, $A_i \cap A_j$ pedig azt jelenti, hogy az i . és j . figura nem fordulhat elő, így mindegyik tojásban 8-féle figura lehet. Ezt a gondolatmenetet folytatva a szitaformula alapján azt kapjuk, hogy

$$\mathbb{P}(A_1 \cup \dots \cup A_{10}) = 10 \cdot \frac{9^{20}}{10^{20}} - \binom{10}{2} \cdot \frac{8^{20}}{10^{20}} + \binom{10}{3} \cdot \frac{7^{20}}{10^{20}} - \dots = \sum_{k=1}^{10} (-1)^{k+1} \binom{10}{k} \frac{(10-k)^{20}}{10^{20}} = 78,53\%.$$

Ez tehát annak valószínűsége, hogy legalább az egyik fajta figura hiányzik, és így a keresett valószínűség: $100 - 78,53 = 21,47\%$.

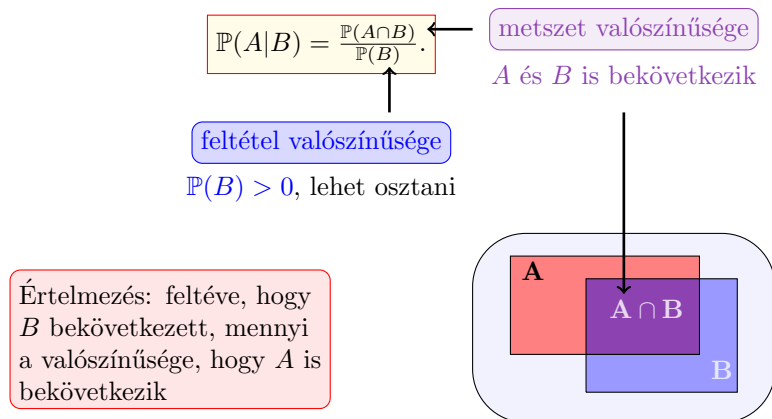
2. Feltételes valószínűség

A feltételes valószínűség fogalmának segítségével választ kaphatunk arra, hogy egy plusz információ hogyan fejeződik ki a valószínűségek meghatározásakor. Tekintsük először az alábbi példát, utána pedig ugyanezt általánosabban.



Általánosabban: legyenek $A, B \in \mathcal{A}$ események úgy, hogy $\mathbb{P}(B) > 0$.

Ekkor az A eseménynek B -re vonatkozó **feltételes valószínűsége**:



Ez alapján írhatjuk fel a feltételes valószínűség definícióját.

2.1. Definíció (Feltételes valószínűség). Legyenek $A, B \in \mathcal{A}$ események, és tegyük fel, hogy $\mathbb{P}(B) > 0$. Az A esemény B -re vonatkozó feltételes valószínűsége:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Példa. Egy 72000 fős városban harmincezer embernek van autója, huszonötezernek lakása, tizenkét ezernek mindkettő. Egy véletlenszerűen választott lakosról tudjuk, hogy van autója. Erre vonatkozóan

mennyi a feltételes valószínűsége, hogy az illetőnek van lakása?

$$\mathbb{P}(L|A) = \frac{\mathbb{P}(L \cap A)}{\mathbb{P}(A)} = \frac{12000/72000}{30000/72000} = \frac{12000}{30000} = 40\% > \mathbb{P}(L) = \frac{25000}{72000} = 34,7\%.$$

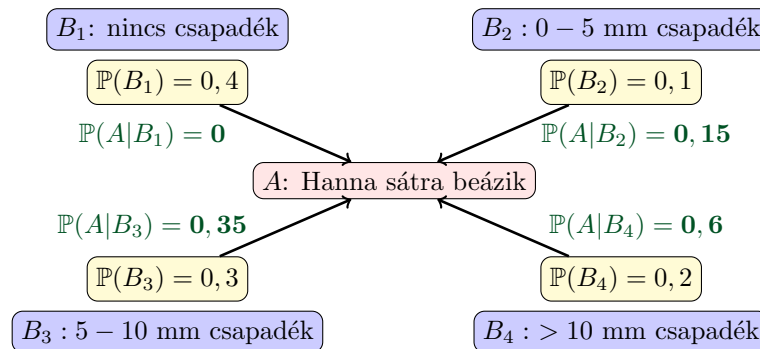
2.1. Teljes valószínűség tétele, Bayes-tétel

Példa. Hanna sátorozni megy Sopronba. Az alábbi táblázat mutatja, hogy adott mennyiségű csapadék esetén mennyi valószínűséggel ázik be a sátra, illetve az előrejelzés szerint mennyi az adott csapadékmennyiség valószínűsége.

csapadék (mm)	0	0 – 5	5 – 10	10-nél több
beázás valószínűsége	0	15%	35%	60%
előrejelzés	40%	10%	30%	20%

Mennyi a valószínűsége, hogy **Hanna sátra beázik** aznap este?

Hanna másnap reggel a **beázott sátorról** küld képeket. Erre feltélesen mennyi annak valószínűsége, hogy Sopronban **több mint 10 mm eső** esett ezen a napon?



Azt látjuk, hogy a kérdéses esemény felbontható négy különböző, egymást kizáró esemény uniójára, amiknek a valószínűségét viszont ki tudjuk számítani:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}((A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4)) = \\ &= \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) + \mathbb{P}(A \cap B_4) = \\ &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) + \mathbb{P}(A|B_4)\mathbb{P}(B_4) = \\ &= 0 \cdot 0,4 + 0,15 \cdot 0,1 + 0,35 \cdot 0,3 + 0,6 \cdot 0,2 = 24\%. \end{aligned}$$

A másik kérdésünk a következő volt: Hanna a **beázott sátorról** küld képeket. Erre feltélesen mennyi annak valószínűsége, hogy Sopronban **több mint 10 mm eső** esett ezen a napon? Azaz mennyi $\mathbb{P}(B_4|A)$?

csapadék (mm)	0	0 – 5	5 – 10	10-nél több
beázás valószínűsége	0	15%	35%	60%
előrejelzés	40%	10%	30%	20%

$$\begin{aligned} \mathbb{P}(B_4|A) &= \frac{\mathbb{P}(A \cap B_4)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_4)\mathbb{P}(B_4)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \dots + \mathbb{P}(A|B_4)\mathbb{P}(B_4)} = \\ &= \frac{0,6 \cdot 0,2}{0 \cdot 0,4 + 0,15 \cdot 0,1 + 0,35 \cdot 0,3 + 0,6 \cdot 0,2} = \frac{0,12}{0,24} = 50\%. \end{aligned}$$

Vagyis feltéve, hogy beázott a sátor, 50% valószínűséggel volt 10 mm-nél több csapadék. Ez több, mint 20%: feltéve, hogy beázott a sátor, valószínűbb a sok csapadék, mint az új információ nélkül.

A fenti gondolatmenetet az alábbi módon általánosíthatjuk.

2.2. Definíció (Teljes eseményrendszer). A $B_1, B_2, \dots \in \mathcal{A}$ (véges vagy megszámlálható sok) esemény együttesét teljes eseményrendszernek nevezzük, ha

- (i) $\bigcup_{i=1}^{\infty} B_i = \Omega$, azaz a B_i -k közül legalább az egyik bekövetkezik, minden elemi esemény szerepel valamelyik eseményben;
- (ii) $B_i \cap B_j = \emptyset$ teljesül minden $1 \leq i < j$ -re, azaz páronként kizáróak, semelyik kettő nem következhet be egyszerre, semelyik elemi esemény nem szerepel egyszerre két eseményben is;
- (iii) $\mathbb{P}(B_i) > 0$ minden $i = 1, 2, \dots$ -re, azaz mindegyiknek pozitív a valószínűsége.

2.1. Tétel (Teljes valószínűség tétele). Legyen $A \in \mathcal{A}$ tetszőleges esemény, B_1, B_2, \dots pedig teljes eseményrendszer. Ekkor

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) + \dots = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Bizonyítás. A bizonyítás ugyanúgy történhet, mint a példa esetében.

$$\mathbb{P}(A) \stackrel{(a)}{=} \mathbb{P}((A \cap B_1) \cup (A \cap B_2) \cup \dots) \stackrel{(b)}{=} \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \dots \stackrel{(c)}{=} \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \dots$$

Itt használjuk, hogy (a) a (B_j) teljes eseményrendszer, így az unió valóban A lesz; (b) a valószínűségek az additivásra vonatkozó tulajdonságát: megszámlálható sok, páronként kizáró esemény esetén az unió valószínűsége a valószínűségek összege, és itt, mivel B_j -k kizáróak, az A -val vett metszeteik is azok, hiszen

$$(A \cap B_i) \cap (A \cap B_j) \subseteq B_i \cap B_j = \emptyset;$$

illetve (c) a feltételes valószínűség definícióját. □

2.2. Tétel (Bayes-tétel). Legyen $A \in \mathcal{A}$ olyan esemény, melyre $\mathbb{P}(A) > 0$, B_1, B_2, \dots pedig teljes eseményrendszer. Ekkor minden $k = 1, 2, \dots$ -re teljesül, hogy

$$\begin{aligned} \mathbb{P}(B_k|A) &= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) + \dots} = \\ &= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}. \end{aligned}$$

Bizonyítás. Az állítás könnyen következik az előző tételből, hiszen

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)}$$

a feltételes valószínűség definíciója alapján, a nevezőben pedig használhatjuk a 2.1. tételt (az ottani feltételek most is érvényesek). □

Példa. Tegyük fel, hogy egy város lakosságának 36%-a rendelkezik diplomával, 42%-a érettségivel (mint legmagasabb végzettség), a többiek egyikkel sem. A diplomások 23%-a, az érettségizettek 21%-a, az érettségivel nem rendelkezők 10%-ának van saját autója.

- Mennyi a valószínűsége, hogy egy véletlenszerűen választott lakos saját autóval rendelkezik?
- A város egyik lakosa elárulja, hogy van saját autója. Ez alapján mennyi a feltételes valószínűsége, hogy az illető diplomával rendelkezik?

Tekintsük az alábbi eseményeket: A : autó; B_1 : diploma; B_2 : érettségi; B_3 : nem érettségizett. Itt B_1, B_2, B_3 teljes eseményrendszert alkotnak, közülük pontosan az egyik következik be. Így alkalmazhatjuk a teljes valószínűség tételét:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) = \\ &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) = \\ &= 0,23 \cdot 0,36 + 0,21 \cdot 0,42 + 0,1 \cdot 0,22 = 19,3\%. \end{aligned}$$

Vagyis egy véletlenszerűen választott lakos 19,3% valószínűséggel rendelkezik saját autóval.

A második kérdéshez szintén a B_1, B_2, B_3 teljes eseményrendszerre és az A eseményre a Bayes-tételt alkalmazhatjuk:

$$\begin{aligned}\mathbb{P}(B_1|A) &= \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3)} = \\ &= \frac{0,23 \cdot 0,36}{0,23 \cdot 0,36 + 0,21 \cdot 0,42 + 0,1 \cdot 0,22} = 42,9\%.\end{aligned}$$

Vagyis egy autóval rendelkező lakosnak 42,9% valószínűséggel van diplomája. Ez nagyobb a diplomások arányánál (36%), hiszen a diplomások között nagyobb az autósok aránya, mint a teljes népességben.

3. Események függetlensége

Tegyük fel, hogy egy városban

- összesen 100000 ember él;
- 15000 embernek **van saját autója** (A esemény):

$$\mathbb{P}(A) = \frac{15000}{100000} = \mathbf{0,15}.$$

- 25000-nek **több a jövedelme az átlagosnál** (B esemény):

$$\mathbb{P}(B) = \frac{25000}{100000} = \mathbf{0,25}.$$

- 10000 ember van, akinek **több a jövedelme az átlagosnál és saját autóval rendelkezik**:

$$\mathbb{P}(A \cap B) = \frac{10000}{100000} = \mathbf{0,1}.$$

Független-e A és B , vagyis az, hogy egy véletlenszerűen választott lakos jövedelme több az átlagosnál, és saját autóval rendelkezik?

Hasonlítsuk össze **az autóval rendelkezők arányát a teljes lakosságban**, illetve **az átlagosnál nagyobb jövedelműek között**. Ha független a két esemény, akkor a két arány megegyezik.

Az autóval rendelkezők aránya a teljes lakosságban: $\mathbb{P}(A) = \frac{15000}{100000} = \mathbf{0,15}$	≠	Az autóval rendelkezők aránya az átlagosnál nagyobb jövedelműek között: $\frac{10000}{25000} = \mathbf{0,4}$
---------------------------------------------------------------------------------------------------------------	---	--------------------------------------------------------------------------------------------------------------

Akkor egyezett volna meg a két arány, ha $\mathbb{P}(A|B) = \mathbb{P}(A)$, azaz

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Leftrightarrow \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B).$$

3.1. Definíció (Események függetlensége). Az $A, B \in \mathcal{A}$ események **függetlenek**, ha

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

vagyis **a metszet valószínűsége a valószínűségek szorzata**.

Az $A_1, A_2, \dots \in \mathcal{A}$ események **függetlenek**, ha tetszőleges $k \geq 1$ és $1 \leq i_1 < i_2 < \dots < i_k \leq n$ számokra

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

Azért van szükség rá, hogy a fenti tulajdonságot minden i_1, i_2, \dots, i_k -ra előírjuk, mert például lehetséges, hogy az A_1, A_2, \dots, A_n események közül bármelyik kettő független, de a fenti definíció értelmében nem függetlenek.

Független eseményekre lehet példa (legalábbis közelítőleg):

- holnap Budapesten esik az eső, és holnap New Yorkban esik az eső (térben távoli események);
- holnap Budapesten esik az eső, és egy év múlva Budapesten esik az eső (időben távoli események);
- visszatevéses mintavételnél: az elsőként húzott golyó fekete; a másodikként húzott golyó fekete (hiszen mindig visszateszük a kihúzott golyót, és mindegyik közül azonos valószínűséggel húzunk)

Nem független eseményekre lehet példa (legalábbis közelítőleg):

- holnap Budapesten esik az eső, és holnap Budaörsön esik az eső (térben közeli események);
- holnap Budapesten esik az eső, és holnapután múlva Budapesten esik az eső (időben közeli események);
- visszatevés nélküli mintavételnél: az elsőként húzott golyó fekete; a másodikként húzott golyó fekete (hiszen az első húzás után már kevesebb fekete golyó maradt)

Két szabályos dobókockával dobunk.

(a) Független-e az alábbi két esemény: **az első dobás hatos**; **a második dobás hatos**?

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

A: az első dobás hatos; **B**: a második dobás hatos; **$A \cap B$** : mindkét dobás hatos

$$\frac{1}{36} = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{1}{6} \cdot \frac{1}{6}.$$

Az A és B események **függetlenek**.

(b) Független-e az alábbi két esemény: **az első dobás hatos**; **az összeg 10**?

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

A: az első dobás hatos; **C**: az összeg 10; **$A \cap C$** : az első dobás hatos, a második négyes

$$\frac{1}{36} = \mathbb{P}(A \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(C) = \frac{1}{6} \cdot \frac{1}{12}.$$

Az A és C események **nem függetlenek**.

(c) Független-e az alábbi két esemény: **az első dobás hatos**; **az összeg 7**?

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

A: az első dobás hatos; **D**: az összeg 7; **A ∩ D**: az első dobás hatos, a második egyes

$$\frac{1}{36} = \mathbb{P}(A \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(D) = \frac{1}{6} \cdot \frac{1}{6}.$$

Az A és D események **függetlenek**.