Valószínűségszámítás

Földtudomány szak, 2015/2016. tanév őszi félév

Backhausz Ágnes (ELTE TTK Valószínűségelméleti és Statisztika Tanszék)[1]

Tartalomjegyzék

[1]Kérdések, módosítási javaslatok, javítanivalók esetén: `agnes@cs.elte.hu`

1. Valószínűségi mező

1.1. definíció (Kolmogorov-féle valószínűségi mező). *Az* $(\Omega, \mathcal{A}, \mathbb{P})$ *hármast Kolmogorov-féle valószínűségi mezőnek nevezzük, ha*

- Ω *nem üres halmaz;*
- $\mathcal{A}$ *az* Ω *részhalmazaiból álló halmaz (azaz minden* $A \in \mathcal{A}$*-ra* $A \subseteq \Omega$*) úgy, hogy*
  - *(i)* $\Omega \in \mathcal{A}$*;*
  - *(ii) ha* $A_1, A_2, \ldots \in \mathcal{A}$*, akkor* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ *(azaz megszámlálható sok* $\mathcal{A}$*-beli elem uniója is* $\mathcal{A}$*-beli);*
  - *(iii) ha* $A \in \mathcal{A}$*, akkor* $\Omega \setminus A \in \mathcal{A}$ *(azaz* $\mathcal{A}$*-beli halmazok komplementere is* $\mathcal{A}$*-beli.*
- $\mathbb{P} : \mathcal{A} \to [0, 1]$ *függvény, melyre*
  - *(i)* $\mathbb{P}(\Omega) = 1$*;*
  - *(ii) ha* $A_1, A_2, \ldots \in \mathcal{A}$ *és minden* $1 \leq i < j$*-re* $A_i \cap A_j = \emptyset$*, akkor*

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

1.2. definíció. *A Kolmogorov-féle valószínűségi mezővel kapcsolatos elnevezések:*

- Ω*: eseménytér vagy elemi események halmaza.*
- Ω *elemei* ($\omega \in \Omega$)*: elemi események.*
- $\mathcal{A}$*: események halmaza (vagy események* σ*-algebrája).*
- $\mathcal{A}$ *elemei* ($A \in \mathcal{A}$)*: események.*

- $\mathbb{P}$: *valószínűség (probability).*
- Ω *esemény neve: biztos esemény.*
- $\emptyset$ *(üres halmaz) esemény neve: lehetetlen esemény.*
- $A \in \mathcal{A}$ *és* $B \in \mathcal{A}$ *kizáró események, ha* $A \cap B = \emptyset$, *azaz nincs olyan* $\omega \in A$, *melyre* $\omega \in B$ *is teljesül.*

1.3. megjegyzés. *A valószínűség* (*ii*) *tulajdonságát (megszámlálható sok kizáró esemény uniójának valószínűsége a valószínűségük összege)* σ*-additivitásnak hívják.*

1.4. definíció (Diszkrét valószínűségi mező). *Az* $(\Omega, \mathcal{A}, \mathbb{P})$ *valószínűségi mező diszkrét, ha* Ω *véges vagy megszámlálhatóan végtelen.*

1.5. állítás. *Ha* Ω *elemeinek száma* n, *akkor az összes lehetséges esemény száma* 2^n.

Bizonyítás. Az események Ω részhalmazai. Egy n elemű halmaznak 2^n részhalmaza van, ezek közül bármelyik lehet esemény. $\square$

1.1. Példák valószínűségi mezőre

Csapadék

Tegyük fel, hogy holnap 25% valószínűséggel lesz csapadék. Ennek egy modellje Kolmogorov-féle valószínűségi mezővel:

- $\Omega = \{\text{lesz, nem lesz}\}$.
- $\mathcal{A}$ az Ω összes részhalmaza. Ez 2^2 darab:

$$\mathcal{A} = \big\{\emptyset, \{\text{lesz}\}, \{\text{nem lesz}\}, \{\text{lesz, nem lesz}\}\big\}.$$

- $\mathbb{P} : \mathcal{A} \to [0, 1]$ az alábbiak szerint:

$$\mathbb{P}(\emptyset) = 0;\ \mathbb{P}(\text{lesz}) = 0,25;\ \mathbb{P}(\text{nem lesz}) = 0,75;\ \mathbb{P}(\text{lesz, nem lesz}) = 1.$$

Szabályos kockadobás

Egyszer dobunk egy szabályos dobókockával.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A}$ az $\{1, 2, 3, 4, 5, 6\}$ összes részhalmaza. Ez 2^6 darab, vagyis $|\mathcal{A}| = 64$.
- $\mathbb{P}(A) = \frac{|A|}{6}$ minden $A \in \mathcal{A}$-ra. Vagyis például

$$\begin{aligned}\mathbb{P}(5) &= \mathbb{P}(\text{ötöst dobunk}) = 1/6 \qquad (A = \{5\});\\ \mathbb{P}(1, 2) &= \mathbb{P}(\text{háromnál kisebb számot dobunk}) = 2/6 = 1/3;\\ \mathbb{P}(2, 4, 6) &= \mathbb{P}(\text{páros számot dobunk}) = 3/6 = 1/2.\end{aligned}$$

  Vagyis: annak valószínűsége, hogy páros számot dobunk, 1/2.

Három szabályos érmedobás

Háromszor dobunk egy szabályos pénzérmével. A szabályosság (szimmetria) miatt minden lehetséges dobássorozat egyformán valószínű.

- $\Omega = \{FFF, FFI, FIF, FII, IFF, IFI, IIF, III\}$. Vagyis $|\Omega| = 8$.
- $\mathcal{A}$ az Ω összes részhalmaza. Ez 2^8 darab.
- $\mathbb{P}(A) = \frac{|A|}{8}$ minden $A \in \mathcal{A}$-ra. Vagyis például

$$\begin{aligned}\mathbb{P}(FII) &= \mathbb{P}(\text{az első dobás fej, a másik kettő írás}) = 1/8;\\ \mathbb{P}(FFF, FFI, FIF, FII) &= \mathbb{P}(\text{az első dobás fej}) = 4/8 = 1/2;\\ \mathbb{P}(FFI, FIF, IFF) &= \mathbb{P}(\text{pontosan két fejet dobunk}) = 3/8.\end{aligned}$$

  Vagyis: annak valószínűsége, hogy pontosan két fejet dobunk, 3/8.

Két szabályos kockadobás

Kétszer dobunk egy szabályos dobókockával (minden számolás ugyanaz lenne, ha két szabályos dobókockával dobnánk).

- $\Omega = \{11, 12, 13, \ldots, 65, 66\}$, ahol például 13 azt jelenti, hogy az első dobás 1, a második 3 (ez különbözik 31-től). $|\Omega| = 6 \cdot 6 = 36$, vagyis 36 darab elemi esemény (dobássorozat) van: mindkét dobás hatféle lehet.

- $\mathcal{A}$: az Ω összes részhalmaza. $|\mathcal{A}| = 2^{36}$.
- $\mathbb{P}(A) = \frac{|A|}{36}$ minden $A \in \mathcal{A}$-ra. Például:

$$A = \{16, 25, 34, 43, 52, 61\} = \{\text{az összeg 7}\}.$$

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

Tehát $\mathbb{P}(A) = \mathbb{P}(\text{az összeg 7}) = 6/36 = 1/6$. Másik példa:

$$B = \{46, 55, 56, 64, 65, 66\} = \{\text{az összeg legalább 10}\}.$$

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**

Tehát $\mathbb{P}(B) = \mathbb{P}(\text{az összeg legalább 10}) = 6/36 = 1/6$.

Itt A és B kizáró események, a metszetük üres, nincs olyan dobássorozat, ami A-ba és B-be is beletartozna.

Fagyos napok száma decemberben

Ha a decemberi fagyos napok (amikor fagypont alá megy a hőmérséklet) számát figyeljük, akkor az összes lehetőség: $0, 1, \ldots, 31$. Vagyis lehet $\Omega = \{0, 1, \ldots, 31\}$. Ilyenkor $\mathcal{A}$ lehet az Ω összes részhalmaza.

Viszont az egyes lehetőségek valószínűségét nem tudjuk könnyen megmondani. Viszont a $0, 1, \ldots, 31$ nem mind egyformán valószínűek: például az 5 minden bizonnyal valószínűbb, mint a 31.

Az összes eddigi példában diszkrét valószínűségi mező szerepelt, hiszen Ω minden esetben véges. A következő példa már nem diszkrét.

Csapadékmennyiség

Tekintsük a holnapi csapadékmennyiséget milliméterben mérve, de kerekítés nélkül.

- $\Omega = [0, 300]$ (ha feltesszük, hogy 300 mm-nél sosem esik több).
- $\mathcal{A}$: Az események halmazának pontos leírásához mélyebb matematikai felépítés és fogalmak lennének szükségesek, de az alábbi példákban szereplő halmazok jól érthetők és $\mathcal{A}$-beliek.
- $\mathbb{P}$: ez a modelltől függ (amit például a korábbi megfigyelések alapján készíthetünk el).

  $0 \in \Omega$, $\{0\} \in \mathcal{A}$, és $\mathbb{P}(0) = \mathbb{P}(\text{holnap nem lesz csapadék})$.

  $[0, 5) \in \mathcal{A}$, és $\mathbb{P}([0, 5)) = \mathbb{P}(\text{holnap 5 mm-nél kevesebb csapadék lesz})$.

  $[10, 300) \in \mathcal{A}$, és $\mathbb{P}([10, 300]) = \mathbb{P}(\text{legalább 10 mm csapadék lesz})$.

1.6. megjegyzés. *Tekintsük a napi csapadékmennyiséget egész szeptemberen keresztül. Ekkor* $\Omega = [0, 300]^{30}$ *lehetne jó választás: mind a* 30 *naphoz egy* 0 *és* 300 *közötti érték tartozik, összesen egy* 30 *tagú véletlen sorozatot kapunk.*

1.2. Műveletek és valószínűségek

Legyen $(\Omega, \mathcal{A}, \mathbb{P})$ Kolmogorov-féle valószínűségi mező.

1.7. definíció. *Legyenek* $A, B \in \mathcal{A}$ *események.*

- *Unió:* $A \cup B = \{\omega \in \Omega : \omega \in A \text{ vagy } \omega \in B\}$.
- *Metszet:* $A \cap B = \{\omega \in \Omega : \omega \in A \text{ és } \omega \in B\}$.
- *Komplementer/ellentett esemény:* $\overline{A} = \{\omega \in \Omega : \omega \notin A\}$.
- *Különbség:* $B \setminus A = \{\omega \in \Omega : \omega \in B \text{ és } \omega \notin A\}$.
- *A maga után vonja B-t, ha minden* $\omega \in A$*-ra* $\omega \in B$ *is teljesül. Jelölés:* $A \subseteq B$.

1.8. állítás. *Ha* $A \in \mathcal{A}$ *esemény, akkor* $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

Bizonyítás. A és $\overline{A}$ kizáró események, azaz $A \cap \overline{A} = \emptyset$. Így a valószínűség (ii) tulajdonsága alapján (1.1. definíció) alapján $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(\Omega) = 1$, a valószínűség (i) tulajdonságát is felhasználva. □

1.9. állítás. *Tetszőleges $A, B \in \mathcal{A}$ eseményekre* $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$.

Bizonyítás. A definícióból következik, hogy $A \cap B$ és $B \setminus A$ kizáró események (az egyikban csak A-beli, a másikban csak A-n kívüli elemi események vannak, így nem lehet olyan, amit mindkettő tartalmaz). Ezért a valószínűség (ii) tulajdonsága alapján (1.1. definíció)

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

□

1.10. állítás. *Ha az $A, B \in \mathcal{A}$ eseményekre $A \subseteq B$, akkor* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Bizonyítás. Az $A \subseteq B$ feltétel miatt $A \cap B = A$. Az előző állítást és a valószínűség definícióját használjuk, miszerint egy esemény valószínűsége nem lehet negatív:

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

□

1.3. Szitaformulák

1.11. állítás (Szitaformula két eseményre). *Tetszőleges $A, B \in \mathcal{A}$ eseményekre teljesül, hogy*

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

Bizonyítás. Az A és $B \setminus A$ események kizáróak. Használjuk a valószínűség (ii) tulajdonságát (1.1. definíció), majd az 1.9 állítást:

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

□

1.12. állítás (Szitaformula három eseményre). *Tetszőleges $A, B, C \in \mathcal{A}$ eseményekre teljesül, hogy*

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

1.13. állítás (Szitaformula, Poincaré-formula). *Legyenek $A_1, \ldots, A_k$ tetszőleges események. Ekkor az események uniójának valószínűségét a következőképpen számíthatjuk ki:*

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{i=1}^{k} A_i\Big) &= \sum_{i=1}^{k} \mathbb{P}(A_i) - \sum_{1 \le i_1 < i_2 \le k} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \\ &\quad + \sum_{1 \le i_1 < i_2 < i_3 \le k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \\ &\quad - \ldots + (-1)^{k+1} \mathbb{P}(A_1 \cap \ldots \cap A_k) = \\ &= \sum_{t=1}^{k} (-1)^{t+1} \sum_{1 \le i_1 < i_2 < \ldots < i_t \le k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{ik}).\end{aligned}$$

1.14. megjegyzés. *Az utolsó részben a belső összegnek $\binom{k}{t}$ tagja van, hiszen ennyiféleképpen tudunk kiválasztani az $1, 2, \ldots, k$ számok közül k különbözőt, ha a sorrend nem számít (és itt nem számít, hiszen nagyság szerint lesznek rendezve).*

1.4. Példák a szitaformulára

Két esemény

Tegyük fel, hogy a holnapi napon $0,3$ valószínűséggel esik az eső, $0,2$ valószínűséggel esik a hó, $0,1$ valószínűséggel pedig eső is esik és hó is hullik. Annak valószínűségét, hogy lesz csapadék eső vagy hó formájában (az is számít, ha mindkettő lesz), a következőképpen számíthatjuk ki:

$$A : \text{eső}; \qquad \mathbb{P}(A) = 0,3; \qquad B : \text{hó}; \qquad \mathbb{P}(B) = 0,2;$$
$$\mathbb{P}(A \cap B) = \mathbb{P}(\text{eső és hó is lesz}) = 0,1;$$
$$\mathbb{P}(\text{eső vagy hó}) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0,3 + 0,2 - 0,1 = 0,4.$$

Három esemény

Tegyük fel, hogy a holnapi napon Ajkán $0,1$, Békéscsabán $0,3$, Csornán pedig $0,4$ valószínűséggel esik az eső. Annak a valószínűsége, hogy Ajkán és Békéscsabán is esik, $0,06$, annak, hogy Békéscsabán és Csornán is, $0,25$, annak pedig, hogy Ajkán és Csornán is, $0,08$. Végül annak valószínűsége, hogy mindhárom helyen esik az eső, $0,02$. Kérdés: mennyi annak valószínűsége, hogy holnap a három város közül legalább az egyik helyen esik az eső.

Az 1.12. állítás alapján így számolhatunk, ha A az az esemény, hogy Ajkán esik, B az, hogy Békéscsabán, C pedig az, hogy Csornán lesz eső:

$$\begin{aligned}\mathbb{P}(\text{legalább az egyik helyen esik}) &= \mathbb{P}(A\cup B\cup C) = \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)- \\ &- \mathbb{P}(A\cap B)-\mathbb{P}(B\cap C)-\mathbb{P}(C\cap A)+\mathbb{P}(A\cap B\cap C) \\ &= 0,1+0,3+0,4-0,06-0,25-0,08+0,02=0,43=43\%.\end{aligned}$$

2. Klasszikus valószínűségi mező

2.1. definíció (Klasszikus valószínűségi mező). *Az $(\Omega,\mathcal{A},\mathbb{P})$ Kolmogorov-féle valószínűségi mező klasszikus valószínűségi mező, ha*

- *Ω véges sok elemből áll: $\Omega=\{\omega_1,\dots,\omega_n\}$.*
- *$\mathcal{A}$ az Ω összes részhalmazából áll.*
- *$\mathbb{P}(\omega)=\frac{1}{n}$ minden $\omega\in\Omega$-ra, azaz az elemi események valószínűsége egyenlő.*

Az 1.5 állítás alapján ilyenkor $|\mathcal{A}|=2^n$, hiszen az n elemű Ω halmaznak 2^n részhalmaza van.

2.2. állítás. *Ha $(\Omega,\mathcal{A},\mathbb{P})$ klasszikus valószínűségi mező és $A\in\mathcal{A}$ esemény, akkor*

$$\mathbb{P}(A)=\frac{A\text{ elemeinek száma}}{\Omega\text{ elemeinek száma}}=\frac{|A|}{|\Omega|}.$$

Bizonyítás. Legyen továbbra is $\Omega=\{\omega_1,\dots,\omega_n\}$, és $A=\{\omega_{i_1},\dots,\omega_{i_k}\}\in\mathcal{A}$. Mivel az elemi események egyszerre nem következhetnek be (kizáróak), a valószínűség (*ii*) tulajdonsága alapján (1.1. definíció):

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\{\omega_{i_1}\} \cup \{\omega_{i_2}\} \cup \ldots \cup \{\omega_{i_k}\}) = \\ &= \mathbb{P}(\omega_{i_1}) + \mathbb{P}(\omega_{i_2}) + \ldots + \mathbb{P}(\omega_{i_k}) = \\ &= \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{k}{n} = \frac{|\mathcal{A}|}{|\Omega|},\end{aligned}$$

hiszen az a feltevés, hogy minden ω elemi esemény valószínűsége $1/n$. □

2.1. Példák klasszikus valószínűségi mezőre

Az 1.1. szakaszból a szabályos kockadobás, három szabályos érmedobás, két szabályos kockadobás. Ezekben az esetekben az összes lehetséges dobássorozat egyformán valószínű (6, 8, illetve 36 lehetőség a három esetben).

A csapadék nem klasszikus valószínűségi mező: Ω kételemű volt, tehát véges, de az elemi események (lesz, nem lesz eső), nem egyformán valószínűek (az egyik 25%, a másik 75% valószínűséggel következik be az ottani feltevés szerint. A fagyos napok száma sem klasszikus valószínűségi mező: például az 5 és a 31 nem egyformán valószínűek. A csapadékmennyiség sem klasszikus valószínűségi mező, hiszen $\Omega = [0, 300]$ intervallum, azaz végtelen sok elemet tartalmaz.

2.2. Mintavételezés

Egyetlen mintavétel

A gyártási folyamat végén egy dobozban N darab számozott alkatrész került. Közülük M darab selejtes $(0 \geq M \geq N)$, a maradék $N - M$ jó. Kihúzunk egy alkatrészt találomra, úgy, hogy mindegyik egyenlő valószínűséggel választjuk. Ekkor

$$\mathbb{P}(\text{selejtes alkatrészt húzunk}) = M/N.$$

Az ehhez tartozó klasszikus valószínűségi mező a 2.1. definíció értelmében:

- $\Omega = \{a_1, \ldots, a_N\}$, az N számozott alkatrésznek megfelelően.
- $\mathcal{A}$: az Ω összes részhalmaza, 2^N darab (1.5. állítás).
- $\mathbb{P}(a_i) = 1/N$ minden $i = 1, 2, \ldots, N$-re.

Legyen A az az esemény, amely a selejtes a_i-ből áll. Ekkor $|A| = M$, és a 2.2. állítás alapján $\mathbb{P}(A) = M/N$ valóban.

Visszatevés nélküli mintavételezés

2.3. állítás. *Egy dobozban N darab számozott alkatrész van. Közülük M darab selejtes ($0 \leq M \leq N$), a maradék $N-M$ jó. Kihúzunk véletlenszerűen n darabot egyszerre, úgy, hogy minden n nagyságú csoport azonos valószínűséggel szerepel. Ekkor annak valószínűsége, hogy pontosan k darab selejtes alkatrészt húzunk:*

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

Bizonyítás. Mivel n-szer húzunk és csak M selejtes van, a húzott selejtesek száma nem lehet nagyobb sem n-nél, sem M-nél.

Az elemi események a következők: melyik n darab (különböző) alkatrészt húzzuk ki. Mivel N elemből n különbözőt

$$\binom{N}{n} = \frac{N(N-1)\ldots(N-n+1)}{n(n-1)\ldots 1} = \frac{N!}{n!(N-n)!}$$

módon lehet kiválasztani (mivel egyszerre húzunk, sorrend nincs), az elemi események száma $\binom{N}{n}$. A feltételezés szerint minden ilyen csoportot egyenlő valószínűséggel húzunk. Így a 2.2. állítás alapján a kérdéses valószínűség

$$\frac{|A|}{\binom{N}{n}},$$

ahol A az olyan n-es csoportokból áll, amik pontosan k darab selejtes alkatrészt tartalmaznak.

Tehát a pontosan k selejtes alkatrészt tartalmazó csoportok számát kell meghatároznunk. Az M darab selejtes alkatrészből $\binom{M}{k}$-féleképpen választhatjuk ki azt a k különbözőt, mely a csoportba kerül. Az $N-M$ jó alkatrészből $\binom{N-M}{n-k}$-féleképpen választhatjuk ki azt az $n-k$ különbözőt, mely a csoportba kerül. Bárhogyan választottuk a selejteseket, bárhogyan választhatjuk a jókat, vagyis mind az $\binom{M}{k}$ lehetőséget $\binom{N-M}{n-k}$-féleképpen tudjuk a jókkal kiegészíteni. Ezért a pontosan k alkatrészt tartalmazó csoportok száma:

$$|A| = \binom{M}{k} \cdot \binom{N-M}{n-k}.$$

Mivel már beláttuk, hogy annak valószínűsége, hogy pontosan k darab selejtes alkatrészt húzunk, $|A|/\binom{N}{n}$, az állítás bizonyítása kész.

2.4. megjegyzés. *A fenti megfogalmazásban egyszerre választottuk ki az n alkatrészt. Ugyanez az állítás érvényes, ha az alkatrészeket egymás után választjuk ki, és a már kiválasztottakat nem tesszük vissza: a végén ugyanúgy n darab különböző alkatrészt látunk egyszerre, és, ha elfelejtjük a sorrendet, minden csoport azonos valószínűséggel jelenik meg.*

Példa visszatevés nélküli mintavételre: lottósorsolás

Az ötöslottón 90 számból húznak ki öt különbözőt, úgy, hogy minden ötös csoportnak ugyanannyi a valószínűsége (a számokat egymás után húzzák, és a kihúzottat nem teszik vissza). A szelvényen öt számra tippelhetünk. Annak valószínűsége, hogy pontosan három számot találunk el:

$$\mathbb{P}(\text{három találat}) = \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \approx 0,000812.$$

Itt $N = 90$ az összes szám, ami szerepel, a szelvényen szereplő $M = 5$ szám selejtes, majd $n = 5$-öt húznak visszatevés nélküli mintavétellel. A kérdés pedig az, hogy mennyi annak valószínűsége, hogy $k = 3$ találatunk lesz (azaz pontosan $k = 3$ selejtest húzunk).

Összesen $\binom{90}{5}$ lehetőség van a húzásra, $\binom{5}{3}$-féleképpen lehet kiválasztani a szelvényen szereplő számokból, hogy melyik lesz kihúzva, és $\binom{85}{2}$-féle lehetőség van arra, hogy melyik legyen az a két szám, amit nem találtunk el. Továbbra is, ezeket a lehetőségeket tetszőlegesen párosíthatjuk, bármelyik bármelyik másikkal szerepelhet.

Visszatevéses mintavételezés

2.5. állítás. *Egy dobozban N darab számozott alkatrész van. Közülük M darab selejtes ($0 \le M \le N$), a maradék $N-M$ jó. Kihúzunk véletlenszerűen n darabot úgy, hogy az épp kihúzottat mindig azonnal visszatesszük, és minden húzásnál mind az N alkatrészt azonos valószínűséggel húzzuk ki. Ekkor annak valószínűsége, hogy pontosan k darab selejtes alkatrészt húzunk:*

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

Bizonyítás. Mivel n-szer húzunk, és bármelyik húzás lehet selejtes és jó is, a kihúzott selejtesek száma 0 és n között van.

Az elemi események (Ω elemei) a következők: melyik húzásnál melyik alkatrészt húzzuk. Például: $a_3a_2a_5a_2$ (feltéve, hogy $N \geq 5$).

A lehetséges sorozatok száma: N^n, mert mind a n húzás N-féle lehet, és (a visszatevés miatt) bármit is húzunk az első néhány húzás során, a következő húzásban újra az összes alkatrész sorra kerülhet. A szimmetria miatt ezek a lehetséges sorozatok egyformán valószínűek, tehát itt is használhatjuk a 2.2. állítást.

Kérdés, hogy hány olyan sorozat van, ami pontosan k selejtes alkatrészt tartalmaz. A kihúzott selejtes alkatrészek az n húzás során k különböző helyen szerepelnek. Azt kiválasztani tehát, hogy melyik az a k húzás, ahol selejtest húzunk, $\binom{n}{k}$-féleképpen lehet. Ha ez megvan (bárhogyan is választottuk ki a helyeket), a selejtesnek kijelölt k különböző hely mindegyikére M-féle alkatrészt választhatunk, ez összesen M^k lehetőség. Ha ez is megvan, a maradék $n-k$ hely mindegyikére $N-M$-féleképpen választhatjuk ki, hogy oda melyik jó alkatrész kerüljön. Ez tehát $(N-M)^{n-k}$ lehetőség. Mivel bármelyik választást bármelyikkel összetéve olyan sorozatot kapunk, ami pontosan k darab selejtes alkatrészt tartalmaz, az ilyen sorozatok száma $\binom{n}{k}M^k(N-M)^{n-k}$.

A 2.2. állítás alapján az olyan sorozatok számát, melyek pontosan k darab selejtes alkatrészt tartalmaznak, kell elosztanunk az összes sorozat számával, vagyis N^n-nel, így kapjuk az állításban szereplő kifejezést. □

Példa visszatevéses mintavételezésre

Egy fiókban 12 kesztyű van, amiből 8 balkezes, a többi jobbkezes. Négyszer húznak véletlenszerűen, mindig visszatéve az épp kihúzott kesztyűt, és mindig az összes kesztyűt azonos valószínűséggel választják. Annak valószínűsége, hogy pontosan kétszer húznak balkezest:

$$\mathbb{P}(\text{két balkezes húzás}) = \binom{4}{2}\frac{8^2 4^2}{12^4} = 6\frac{64 \cdot 16}{20736} \approx 0,2963.$$

3. Feltételes valószínűség

Kérdés. Valakiről annyit tudunk, hogy három gyermeke van. Mennyi a valószínűsége, hogy a középső gyermeke fiú? És ha elárulja, hogy pontosan egy fia van, mennyi annak valószínűsége, hogy a középső gyermeke fiú?

Kérdés. Mennyi annak valószínűsége, hogy egy véletlenszerűen kiválasztott embernek magas a vérnyomása? Mennyi annak valószínűsége, hogy magas a vérnyomása, ha megmérjük a testtömegét, és megtudjuk róla, hogy túlsúlyos (de nem elhízott)?

3.1. 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:*

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

Az első esetben: A: a középső gyerek fiú; B: pontosan egy fiú van a három gyerek között. Ekkor $A \cap B$ azt jelenti, hogy a középső gyerek fiú, és pontosan egy fiú van, vagyis a másik kettő lány. Mindebből, az egyszerűség kedvéért feltételezve, hogy a fiúk és lányok aránya népességben megegyezik, és a testvérek neme között nincs összefüggés (vagyis mind a nyolc lehetőség: FFF, FFL, FLF, ..., LLL egyformán valószínű):

$$\mathbb{P}(A) = \mathbb{P}(\text{a középső fiú}) = \frac{1}{2}.$$

Ugyanakkor a feltételes valószínűséget így számolhatjuk ki:

$$\mathbb{P}(\text{a középső fiú}|\text{egy fiú van}) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(LFL)}{\mathbb{P}(FLL, LFL, LLF)} = \frac{\frac{1}{8}}{\frac{3}{8}} = 1/3.$$

A második esetben (nagyjából valós adatokkal): a magyar lakosság 25 %-ának van magas vérnyomása (A esemény). A magyar lakosság 30%-a túlsúlyos (de nem elhízott), ez a B esemény. Azok aránya, akik túlsúlyosak (de nem elhízottak) és magasvérnyomás-betegségben is szenvednek, 15 %. Tehát:

$$\mathbb{P}(\text{magas vérnyomás}) = 0,25.$$

$$\mathbb{P}(\text{magas vérnyomás}|\text{túlsúly}) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0,15}{0,3} = 0,5.$$

Példa feltételes valószínűségre

Holnap $0,4$ valószínűséggel nem lesz csapadék, $0,15$ valószínűséggel lesz eső és hó is, $0,25$ valószínűséggel csak eső, $0,2$ valószínűséggel pedig csak hó. Feltéve, hogy lesz eső, mennyi a valószínűsége, hogy havazni is fog?

A esemény: havazni fog. B esemény: lesz holnap eső. Ekkor $\mathbb{P}(B) = 0,15 + 0,25 = 0,4$. Másrészt $A \cap B$ azt jelenti, hogy eső és hó is lesz. Így:

$$\mathbb{P}(\text{havazni fog}|\text{lesz eső}) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0,15}{0,4} = 0,375.$$

3.2. megjegyzés. $0 \leq \mathbb{P}(A|B) \leq 1$ *teljesül minden* $A, B \in \mathcal{A}$ *eseményre, ha* $\mathbb{P}(B) > 0$. *Sőt, ha* B*-t rögzítjük,* $\mathbb{P}(A|B)$ *rendelkezik a valószínűség mindkét, 1.1. definícióban szereplő tulajdonságával.*

3.1. Teljes valószínűség tétele

3.3. definíció (Teljes eseményrendszer). *A* $B_1, B_2, \ldots \in \mathcal{A}$ *(véges vagy megszámlálható sok) esemény együttesét teljes eseményrendszernek nevezzük, ha*

(i) $\bigcup_{i=1}^{\infty} B_i = \Omega$*;*

(ii) $B_i \cap B_j = \emptyset$ *teljesül minden* $1 \leq i < j$*-re;*

(iii) $\mathbb{P}(B_i) > 0$ *minden* $i = 1, 2, \ldots$*-re.*

Példa.

B_1: holnap nem lesz csapadék;

B_2: holnap lesz csapadék, de nem több 5 mm-nél;

B_3: a holnapi csapadékmennyiség több mint 5 mm, de kevesebb mint 10 mm;

B_4: a holnapi csapadékmennyiség meghaladja a 10 mm-t.

Ekkor B_1, B_2, B_3, B_4 teljes eseményrendszer (közülük pontosan az egyik következik be).

3.4. tétel (Teljes valószínűség tétele). *Legyen* $A \in \mathcal{A}$ *tetszőleges esemény,* $B_1, B_2, \ldots$ *pedig teljes eseményrendszer. Ekkor*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Bizonyítás.

1. lépés: Felhasználva, hogy a 3.3. definíció (i) része szerint $\bigcup_{i=1}^{\infty} B_i = \Omega$:

$$\mathbb{P}(A) = \mathbb{P}\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty}(A \cap B_i)\right).$$

2. lépés: Az $A \cap B_i$, $i = 1, 2, \ldots$ események páronként kizáróak (vagyis semelyik kettő nem következhet be egyszerre). Ugyanis

$$(A \cap B_i) \cap (A \cap B_j) \subseteq B_i \cap B_j = \emptyset \qquad (1 \le i < j),$$

hiszen a 3.3. definíció (ii) része szerint a B_i események is páronként kizáróak.

3. lépés: A valószínűség (ii) tulajdonsága (az 1.1. definíció) megszámlálható sok páronként kizáró esemény uniójának valószínűsége a valószínűségek összege, így az 1. lépést folytatva, majd felhasználva a feltételes valószínűség definícióját (a 3.1. definíció):

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{\infty}(A \cap B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \qquad \square$$

Példa. Hanna sátorozni megy Badacsonyba. Ha van csapadék, de nem több 5 mm-nél, akkor 15% valószínűséggel ázik be a sátra. Ha több mint 5 mm, de kevesebb mint 10 mm csapadék lesz, akkor 35% valószínűséggel. Végül, ha több mint 10 mm csapadék lesz, akkor 60% valószínűséggel ázik be a sátor. Az előrejelzés szerint 10% valószínűséggel lesz csapadék, de nem több 5 mm-nél, 30% valószínűséggel a holnapi csapadékmennyiség több mint 5 mm, de kevesebb mint 10 mm, és 20% annak valószínűsége, hogy a csapadékmennyiség holnap meghaladja a 10 mm-t. Mennyi annak valószínűsége, hogy holnap beázik Hanna sátra?

Az előző példában már láttuk, hogy az ott felsorolt B_1, B_2, B_3, B_4 események teljes eseményrendszert alkotnak. Legyen az A esemény az, hogy Hannának beázik a sátra. Ezekre alkalmazzuk a teljes valószínűség tételét:

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

3.2. Bayes-tétel

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

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Bizonyítás. Használjuk a feltételes valószínűség definícióját (3.1. definíció), mindkétszer:

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

Mivel A esemény, $B_1, B_2, \ldots$ pedig teljes eseményrendszer, ezután használhatjuk a teljes valószínűség tételét (3.4. tétel) a $\mathbb{P}(A)$ valószínűség kifejezésére, így megkapjuk az állítást. □

Példa. Másnap Hanna a beázott sátorról küld képeket. Mennyi annak valószínűsége, hogy Badacsonyban több mint 10 mm eső esett ezen a napon?

A kérdés a B_4 esemény (több mint 10 mm eső) valószínűsége, feltéve, hogy az A esemény (a sátor beázik) bekövetkezett, azaz $\mathbb{P}(B_4|A)$. Már láttuk, hogy $\mathbb{P}(A) > 0$, B_1, B_2, B_3, B_4 pedig továbbra is teljes eseményrendszer, így alkalmazhatjuk Bayes tételét:

$$\begin{aligned}\mathbb{P}(B_4|A) &= \frac{\mathbb{P}(A|B_4)\mathbb{P}(B_4)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \ldots + \mathbb{P}(A|B_4)\mathbb{P}(B_4)} = \\ &= \frac{0,6 \cdot 0,2}{0 \cdot 0,3 + 0,15 \cdot 0,1 + 0,35 \cdot 0,3 + 0,6 \cdot 0,2} = \frac{0,12}{0,24} = 0,5.\end{aligned}$$

3.6. megjegyzés. *Szorzási szabály: ha $A_1, A_2, \ldots, A_n$ események, és teljesül, hogy $\mathbb{P}(A_1 \cap \ldots \cap A_n) > 0$, akkor*

$$\mathbb{P}(A_1 \cap \ldots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

4. Események függetlensége

4.1. definíció (Két esemény függetlensége). *Az $A, B \in \mathcal{A}$ események függetlenek, ha*

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

Példa. Annak valószínűsége, hogy holnap Budapesten esik az eső, $0,2$. Annak valószínűsége, hogy holnap Torontóban esik az eső, $0,1$. Annak valószínűsége, hogy holnap Budapesten és Torontóban is esik az eső, $0,02$. Ekkor az a két esemény, hogy Budapesten (B), illetve Torontóban (T) esik az eső, *független*, hiszen

$$\mathbb{P}(B \cap T) = \mathbb{P}(\text{mindkét helyen esik az eső}) = 0,02 = 0,2 \cdot 0,1 = \mathbb{P}(B)\mathbb{P}(T).$$

Példa. Annak valószínűsége, hogy holnap Budapesten esik az eső, $0,2$. Annak valószínűsége, hogy holnap Budaörsön esik az eső, $0,15$. Annak valószínűsége, hogy holnap Budapesten és Budaörsön is esik az eső, $0,12$. Ekkor az a két esemény, hogy Budapesten (B), illetve Budaörsön (C) esik az eső, *nem független*, hiszen

$$\mathbb{P}(B \cap C) = 0,12 \neq 0,03 = 0,2 \cdot 0,15 = \mathbb{P}(B)\mathbb{P}(C).$$

Észrevehetjük azt is, hogy ilyenkor

$$\mathbb{P}(B|C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{0,12}{0,15} = 0,8 \neq 0,2 = \mathbb{P}(B).$$

Vagyis C-t feltételezve B nagyobb valószínűséggel következik be, mint feltételezés nélkül (a budaörsi eső esetén Budapesten nagyobb valószínűséggel esik, mint a budapesti eső feltétel nélküli valószínűsége).

4.2. megjegyzés. *Ha $A, B \in \mathcal{A}$ események, és $\mathbb{P}(B) > 0$: A és B pontosan akkor függetlenek, ha $\mathbb{P}(A|B) = \mathbb{P}(A)$.*

4.3. definíció (Függetlenség). *Az $A_1, A_2, \ldots, A_n \in \mathcal{A}$ véges sok esemény (teljesen) függetlenek, ha minden $1 \leq k \leq n$-re és $1 \leq i_1 < i_2 < \ldots < i_k \leq n$-re*

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

Az $A_1, A_2, \ldots \in \mathcal{A}$ megszámlálhatóan végtelen sok esemény (teljesen) független, ha minden $k \in \mathbb{N}$-re és $1 \leq i_1 < i_2 < \ldots < i_k$-ra

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

4.4. definíció (Páronkénti függetlenség). *Az $A_1, A_2, \ldots \in \mathcal{A}$ események páronként függetlenek, ha minden $1 \leq i < j$ esetén A_i és A_j függetlenek, azaz*

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Példa. Szabályos dobókockával kétszer dobunk. Legyen

A: az első dobás páros.

B: a második dobás páros.

C: a két dobás összege páros.

Ekkor $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$, továbbá $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/4$. Ugyanakkor, ha az első és a második dobás páros, akkor a két dobás összege is biztosan páros, így

$$\mathbb{P}(A \cap B \cap C) = 1/4 \neq 1/2 \cdot 1/2 \cdot 1/2 = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C).$$

Mindebből az látható, hogy az A, B, C események páronként függetlenek, de nem teljesen függetlenek (amit az is mutat, hogy ha kettőről elárulják, hogy bekövetkeztek-e, a harmadikról már biztosan el tudjuk dönteni, hogy az bekövetkezett-e).

4.5. állítás (Komplementer függetlensége). *Ha az $A_1, A_2, \ldots, A_n \in \mathcal{A}$ események függetlenek, akkor az $\overline{A_1}, A_2, \ldots, A_n \in \mathcal{A}$ események is függetlenek.*

4.6. állítás. *Ha az $A, B \in \mathcal{A}$ események függetlenek, akkor*

- *A és $\overline{B}$ függetlenek;*
- *$\overline{A}$ és B függetlenek;*
- *$\overline{A}$ és $\overline{B}$ függetlenek.*

4.7. állítás. *Legyenek $A, B \in \mathcal{A}$ események úgy, hogy $\mathbb{P}(A) = 0$, vagy $\mathbb{P}(A) = 1$. Ekkor A és B függetlenek.*

5. Valószínűségi változók és eloszlásuk

5.1. definíció (Valószínűségi változó). *Egy $X : \Omega \to \mathbb{R}$ függvény valószínűségi változó, ha tetszőleges $a < b$ valós számokra*

$$\{\omega \in \Omega : a < X(\omega) \leq b\} \in \mathcal{A}.$$

Példa. Valakinek három gyereke születik. Legyen X a fiúk száma. Ekkor

$$\Omega = \{FFF, FFL, FLF, FLL, LFF, LFL, LLF, LLL\};$$

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

5.1. Diszkrét valószínűségi változók eloszlása

5.2. definíció (Diszkrét valószínűségi változó). *Az $X : \Omega \to \mathbb{R}$ valószínűségi változó diszkrét, ha véges sok vagy megszámlálható sok értéket vesz fel.*

5.3. definíció (Eloszlás). *Legyenek az X valószínűségi változó lehetséges értékei $x_1, x_2, \ldots \in \mathbb{R}$, és legyen*

$$p_i = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\}) = \mathbb{P}(X = x_i) \qquad (i = 1, 2, \ldots).$$

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

Ezentúl a definícióban szereplő rövidebb, $\mathbb{P}(X = x_i)$ jelölést fogjuk használni.

Vegyük észre, hogy $0 \leq p_i \leq 1$ teljesül minden $i = 1, 2, \ldots$ esetén, és $\sum_{i=1}^{\infty} p_i = 1$. Ugyanis $\{X = x_i\}$ teljes eseményrendszer (pontosan egy következik be közülük), tehát az uniójuk Ω, így pedig az 1.1. definíció $\mathbb{P}$-re vonatkozó részéből következik az állítás.

5.4. definíció (Valószínűségeloszlás). *Azt mondjuk, hogy a $p_1, p_2, \ldots$ sorozat valószínűségeloszlás, ha*

- *$p_i \geq 0$ minden $i = 1, 2, \ldots$ esetén;*
- *$\sum_{i=1}^{\infty} p_i = 1$.*

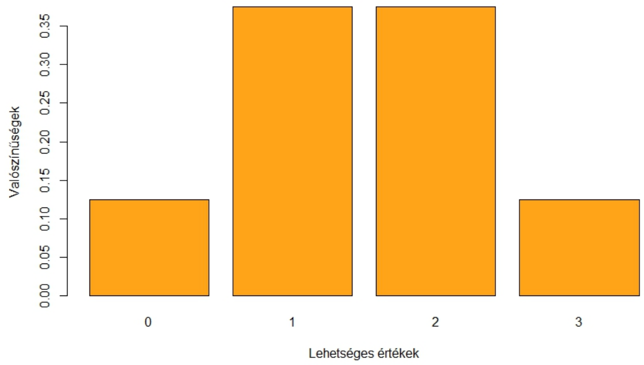
Tehát azt láttuk, hogy ha (x_i, p_i) egy X valószínűségi változó eloszlása, akkor p_i valószínűségeloszlás.

Példa: három gyerek. Az előző példában: háromszor gyerek születik, X a fiúk száma. Ekkor X lehetséges értékei: $0, 1, 2, 3$, vagyis $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$. Tegyük fel, hogy minden gyerek a többiektől függetlenül $1/2$ valószínűséggel fiú. Eszerint mind a nyolc lehetőség valószínűsége $1/8$, hiszen például $\mathbb{P}(FLF) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$. Így, a fenti csoportosításból adódóan:

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

Mindezek alapján X eloszlása az alábbi sorozat:

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



1. ábra.
A fiúk számának lehetséges értékei a három gyerek közül és a hozzájuk tartozó valószínűségek

Ez az eloszlás látható az 1. ábrán.

Példa: szabályos kockadobás. Egyszer dobunk szabályos dobókockával, jelölje Y a dobott számot. Ekkor Y eloszlása (2. ábra):

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

A 3. ábrán 1000 (számítógéppel szimulált) szabályos kockadobás után mind a hat lehetséges értékről feltüntettük a relatív gyakoriságát: az előfordulások számának és az összes dobásnak a hányadosát.

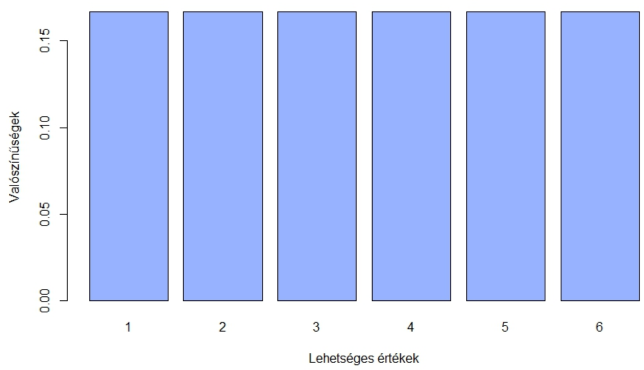
Példa: jégeső. Tegyük fel, hogy júliusban $0,7$ valószínűséggel nincs jégeső, $0,2$ valószínűséggel egyszer fordul elő, $0,1$ valószínűséggel pedig kétszer. Legyen Z a júliusi jégesők száma. Ekkor Z eloszlása:

$$(0, 7/10), \qquad (1, 2/10), \qquad (2, 1/10).$$

5.2. Diszkrét valószínűségi változó várható értéke

5.5. definíció (Várható érték, diszkrét eset). *Legyen $X : \Omega \to \mathbb{R}$ olyan diszkrét valószínűségi változó, melynek eloszlása $(x_1, p_1), (x_2, p_2), \ldots$. Ekkor X várható értéke:*

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



2. ábra.
A szabályos kockadobás lehetséges értékei és a hozzájuk tartozó valószínűségek

A definícióból is látszik, hogy a várható érték csak a valószínűségi változó eloszlásától függ.

Példák várható értékre

Fiúk száma. Továbbra is három gyerek születik, egymástól függetlenül $1/2 - 1/2$ valószínűséggel fiúk, illetve lányok. X a fiúk száma. A fenti számolás alapján

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

Kockadobás. Y a dobott számot jelöli. Ekkor Y várható értéke:

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

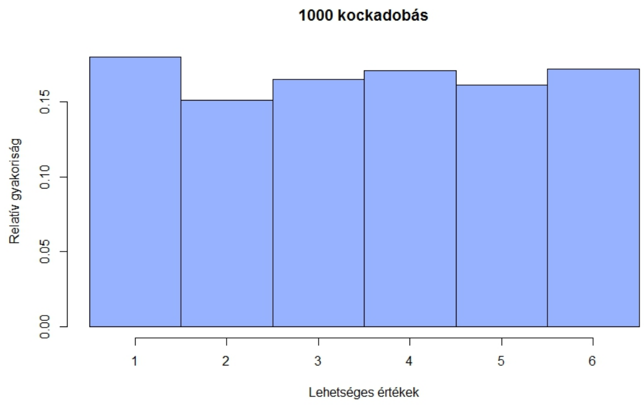
Jégeső. A fenti példában a júliusi jégeső számának várható értéke:

$$\mathbb{E}(Z) = 0 \cdot 0,7 + 1 \cdot 0,2 + 2 \cdot 0,1 = 0,4.$$

5.3. Diszkrét valószínűségi változó szórása

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

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



3. ábra. 1000 szabályos kockadobásból készített hisztogram

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

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

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

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

Példák szórásnégyzetre

Fiúk száma. X a fiúk száma a három gyerek között. Már láttuk, hogy X lehetséges értékei: $0, 1, 2, 3$, és $\mathbb{E}(X) = 3/2$. A definíció alapján tehát ezt kell kiszámítanunk:

$$D^2(X) = \mathbb{E}\big((X - 3/2)^2\big).$$

Az $X - 3/2$ valószínűségi változó eloszlása (minden lehetséges érték mellett feltüntetjük a valószínűségét):

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

Ebből látszik, hogy az $(X - 3/2)^2$ valószínűségi változó lehetséges értékei csak $9/4$, illetve $1/4$. Az első esetben: $(X - 3/2)^2 = 9/4$ pontosan akkor teljesül, ha $X = 0$ vagy $X = 3$. Ennek valószínűsége $1/8 + 1/8 = 1/4$. Tehát $(X - 3/2)^2$ eloszlása:

$$(1/4, 3/4); \qquad (9/4, 1/4).$$

Az 5.5. definíció alapján:

$$D^2(X) = \mathbb{E}\big((X-3/2)^2\big) = \frac{1}{4}\cdot\frac{3}{4} + \frac{9}{4}\cdot\frac{1}{4} = \frac{12}{16} = \frac{3}{4}.$$

Másik megoldás az 5.8. állítás alapján. Az X valószínűségi változó lehetséges értékei $0, 1, 2, 3$ voltak. Így az X^2 valószínűségi változó lehetséges értékei $0, 1, 4, 9$. Eloszlása pedig:

$$(0, 1/8); \qquad (1, 3/8); \qquad (4, 3/8); \qquad (9, 1/8).$$

Így az 5.5. definíció szerint

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

Ebből és a korábbi számolásból

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

Végül pedig a fiúk számának szórása:

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

Kockadobás. Most már csak az 5.8. állítás alapján számolunk. Az Y^2 valószínűségi változó lehetséges értékei: $1, 4, 9, 16, 25, 36$, mindegyiknek $1/6$ a valószínűsége. Tehát

$$\mathbb{E}(Y^2) = 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6} = \frac{91}{6}.$$

Tehát, mivel már láttuk, hogy $\mathbb{E}(Y) = 3,5 = \frac{7}{2}$:

$$D^2(Y) = \mathbb{E}(Y^2) - \big[\mathbb{E}(Y)\big]^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12} = 2,9167,$$

amiből a kockadobás értékének szórása ennek négyzetgyöke:

$$D(Y) = 1,7078.$$

Jégeső. A fenti példában a júliusi jégesők száma (amit Z-vel jelöltünk) 0 volt $0,7$ valószínűséggel, 1 volt $0,2$ valószínűséggel, és 2 volt $0,1$ valószínűséggel. Azt is láttuk, hogy $\mathbb{E}(Z) = 0,4$. Tehát most

$$\mathbb{E}(Z^2) = 0^2\cdot 0,7 + 1^2\cdot 0,2 + 2^2\cdot 0,1 = 0,6.$$

Tehát a jégesők számának szórása:

$$D(Z) = \sqrt{\mathbb{E}(Z^2) - \big[\mathbb{E}(Z)\big]^2} = \sqrt{0,6 - 0,4^2} = \sqrt{0,44} = 0,6633.$$

6. Nevezetes diszkrét eloszlások

6.1. Binomiális eloszlás

Legyen n pozitív egész, $0 < p < 1$ szám. Tegyük fel, hogy n-szer ismétlünk egy kísérletet, ami minden egyes alkalommal a többitől függetlenül p valószínűséggel sikerül. Legyen az X valószínűségi változó a sikeres kísérletek száma. Ekkor X lehetséges értékei: $0, 1, \ldots, n$. Annak valószínűsége, hogy a kísérlet pontosan k-szor sikerül: $\binom{n}{k} p^k (1-p)^{n-k}$, ha $0 \leq k \leq n$. Ennek a valószínűségi változónak az eloszlását binomiális eloszlásnak nevezzük (jelölése: $X \sim \mathrm{Bin}(n, p)$), az alábbi definíció szerint.

6.1. definíció (Binomiális eloszlás n renddel és p paraméterrel). *A binomiális eloszlás n renddel és p paraméterrel a (k, p_k) sorozat, ahol*

$$k = 0, 1, 2, \ldots, n; \qquad p_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

A binomiális tétel szerint $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1$, így ez valóban valószínűségeloszlás.

6.2. állítás (A binomiális eloszlás tulajdonságai). *Legyen X binomiális eloszlású valószínűségi változó n renddel és p paraméterrel.*

(a) Az $X \sim \mathrm{Bin}(n, p)$ binomiális eloszlású valószínűségi változó várható értéke:

$$\mathbb{E}(X) = np.$$

(b) Az $X \sim \mathrm{Bin}(n, p)$ binomiális eloszlású valószínűségi változó szórása:

$$D(X) = \sqrt{np(1-p)}.$$

(c) Az a k érték, melyre $p_k = \mathbb{P}(X = k)$ maximális (vagyis X módusza): $[(n+1)p]$, ahol $[\cdot]$ az egész részt jelöli (azaz a legnagyobb olyan egész szám, mely $(n+1)p$-nél kisebb). Ha $(n+1)p$ egész, akkor az eggyel kisebb k is maximális értéket ad.

Példa: fiúk száma

A korábbi példában: három gyermek mindegyike a többitől függetlenül 1/2 valószínűséggel fiú. X jelölte a fiúk számát. Ekkor X binomiális eloszlású $n = 3$ renddel és $p = 0.5$ paraméterrel, ahogy láttuk is:

$$\mathbb{P}(\text{pontosan } k \text{ fiú születik}) = \mathbb{P}(X = k) = \binom{3}{k} 0,5^k 0,5^{3-k}.$$

Például

$$\mathbb{P}(\text{pontosan 2 fiú születik}) = \mathbb{P}(X = 2) = \binom{3}{2} 0,5^2 0,5^1 = 3 \cdot \left(\frac{1}{2}\right)^3 = 3/8.$$

X várható értéke és szórása (ahogy láttuk is):

$$\mathbb{E}(X) = 3 \cdot 1/2 = 3/2; \qquad D(X) = \sqrt{3 \cdot 1/2 \cdot 1/2} = \sqrt{\frac{3}{4}} \approx 0,866.$$

Példa: visszatevéses mintavételezés

A 2.2. szakaszban leírt példában N alkatrész közül M volt selejtes. n-szer húzunk, úgy, hogy az éppen kiválasztott alkatrészt a húzás után visszatesszük. Minden alkalommal minden alkatrészt azonos valószínűséggel húzunk. Ilyenkor az n húzás során kapott selejtes alkatrészek száma binomiális eloszlású n renddel és $p = M/N$ paraméterrel.

Példa: mérőműszerek

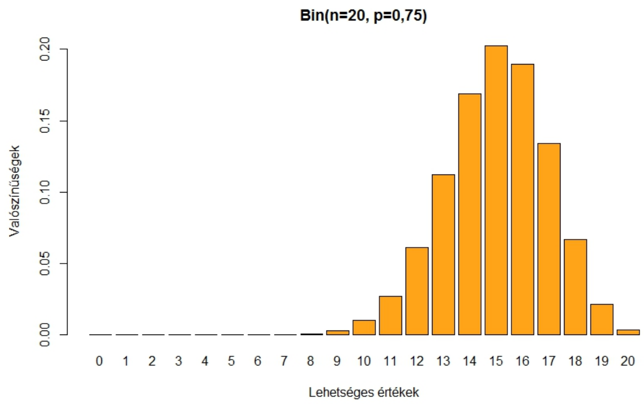
Az ország területén 20 időjárási mérőműszert helyeztek el. Tegyük fel, hogy egy adott napon mindegyik a többitől függetlenül 0,75 valószínűséggel küld adatot a központba (különben hiba történik). Jelölje Y azt, hogy egy adott napon hány műszertől érkezik adat a központba. Ekkor Y binomiális eloszlású $n = 20$ renddel és $p = 0,75$ paraméterrel. Vagyis

$$\mathbb{P}(\text{pontosan } k \text{ adat érkezik}) = \mathbb{P}(Y = k) = \binom{20}{k} 0,75^k \cdot 0,25^{20-k}.$$

Például

$$\mathbb{P}(\text{pontosan 9 adat érkezik}) = \mathbb{P}(Y = k) = \binom{20}{9} 0,75^9 \cdot 0,25^{11} \approx 0,003;$$

$$\mathbb{P}(\text{pontosan 16 adat érkezik}) = \mathbb{P}(Y = k) = \binom{20}{16} 0,75^{16} \cdot 0,25^4 \approx 0,1897.$$



4. ábra.
Mérőműszerektől érkező adatok számának eloszlása: binomiális eloszlás, $n = 20$, $p = 0,75$.

A beérkező adatok számának várható értéke és szórása:

$$\mathbb{E}(Y) = n \cdot p = 20 \cdot 0,75 = 15;$$
$$D(Y) = \sqrt{np(1-p)} = \sqrt{20 \cdot 0,75 \cdot 0,25} = 1,936.$$

Itt a legnagyobb valószínűséget is le tudjuk olvasni: $(n+1)p = 21 \cdot 0,75 = 15,75$. Tehát $[(n+1)p] = 15$. Vagyis $k = 15$-re kapjuk a legnagyobb valószínűséget, ahogy az ábrán is látható.

6.2. Hipergeometrikus eloszlás

A visszatevéses mintavételnél (2.2. szakasz) a kihúzott selejtes alkatrészek számát hipergeometrikus eloszlásúnak fogjuk nevezni, az alábbi definíció értelmében.

6.3. definíció (Hipergeometrikus eloszlás). *Legyenek N, M, n pozitív egészek úgy, hogy $1 \leq n \leq M \leq N$. A hipergeometrikus eloszlás N, M és n paraméterekkel a (k, p_k) sorozat, ahol*

$$k = 0, 1, \ldots, n; \qquad p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

6.4. állítás (A hipergeometrikus eloszlás tulajdonságai). *Legyen az X valószínűségi változó hipergeometrikus eloszlású N, M és n paraméterekkel.*

(a) Az X valószínűségi változó várható értéke:

$$\mathbb{E}(X) = n \cdot \frac{M}{N}.$$

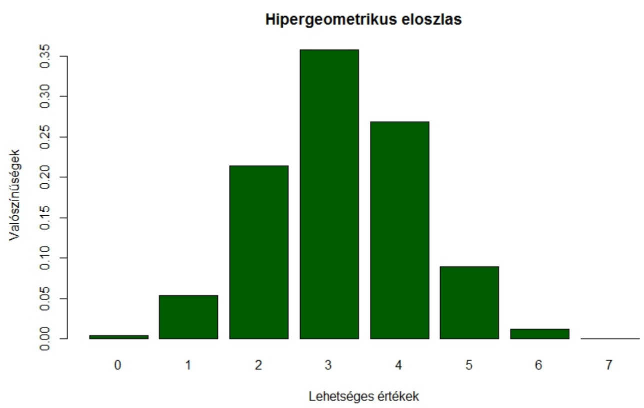
(b) Az X valószínűségi változó szórása:

$$D(X) = \sqrt{n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}}.$$

Példa: kézilabdacsapat

Egy húszfős osztályban kilenc 180 cm-nél magasabb gyerek van. Tornaórán véletlenszerűen kiválasztanak hét gyereket, akik egy csapatba kerülnek a kézilabdameccsen. Legyen X a 180 cm-nél magasabb gyerekek száma a kézilabdacsapatban. Ekkor X hipergeometrikus eloszlású $N = 20, M = 9, n = 7$ paraméterekkel. Tehát a lehetséges értékei $k = 0, 1, \dots, 7$, és ilyenkor

$$\mathbb{P}(\text{pontosan } k \text{ magas gyerek}) = \mathbb{P}(X = k) = \frac{\binom{9}{k}\binom{11}{7-k}}{\binom{20}{7}}.$$



5. ábra.
Magas gyerekek számának eloszlása: hipergeometrikus eloszlás,
$N = 20,\ M = 9, n = 7$.

Például

$$\mathbb{P}(\text{pontosan 2 magas gyerek}) = \mathbb{P}(X = k) = \frac{\binom{9}{2}\binom{11}{5}}{\binom{20}{7}} = 0,2145;$$

$$\mathbb{P}(\text{pontosan 4 magas gyerek}) = \mathbb{P}(X = k) = \frac{\binom{9}{4}\binom{11}{3}}{\binom{20}{7}} = 0,2682.$$

A csapatba kerülő 180 cm-nél magasabb gyerekek számának várható értéke és szórása:

$$\mathbb{E}(X) = 7 \cdot \frac{9}{20} = 3,15;$$

$$D(X) = \sqrt{7 \cdot \frac{9}{20} \cdot \left(1 - \frac{9}{20}\right) \cdot \frac{2}{8}} = 0,6581.$$

6.3. Geometriai eloszlás

Független kísérleteket végzünk, melyek mindegyike p valószínűséggel sikerül. Azt kérdezzük, hogy hányadik kísérlet volt az első sikeres. Ez a mennyiség geometriai eloszlású, az alábbi definíció alapján.

6.5. definíció (Geometrai eloszlás/Pascal-eloszlás). *Legyen $p \in (0,1)$ rögzített szám. Azt mondjuk, hogy a (k, p_k) sorozat geometriai eloszlású p paraméterrel, ha lehetséges értékei*

$$k = 1, 2, \ldots,$$

a hozzájuk tartozó valószínűségek pedig

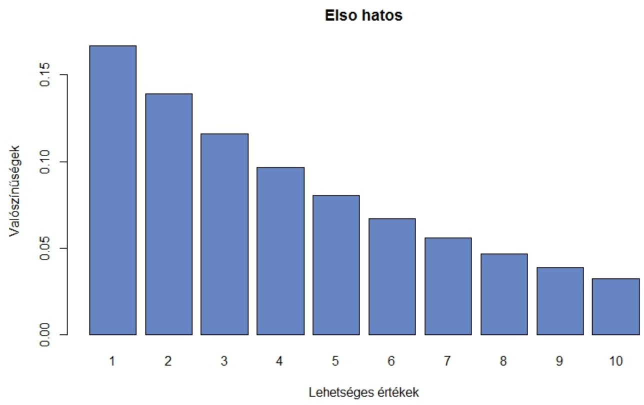
$$p_k = (1-p)^{k-1} p.$$

Ez valóban eloszlás:

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{s=0}^{\infty} (1-p)^s = p \cdot \frac{1}{1-(1-p)} = 1,$$

a geometrai sor összege alapján.

6.6. állítás (A geometriai eloszlás tulajdonságai). *Legyen Y geometriai eloszlású valószínűségi változó p paraméterrel.*



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

(a) Ekkor Y várható értéke és szórása:

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

(b) A $\mathbb{P}(Y = k)$ valószínűség $k = 1$-re maximális.

Példa: első hatos.

Szabályos dobókockával dobunk addig, amíg hatost nem kapunk. Jelölje Y, hogy ez hányadik dobásnál sikerült, vagyis hányadik dobásnál jött ki az első hatos. A dobások egymástól függetlenek. Ekkor Y geometriai eloszlású $p = 1/6$ paraméterrel, lehetséges értékei $k = 1, 2, \ldots$, és

$$\mathbb{P}(Y = k) = p_k = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

Például

$$\mathbb{P}(Y = 4) = \mathbb{P}(\text{a negyedik dobásra jött ki az első hatos}) = \left(\frac{5}{6}\right)^3 \frac{1}{6}.$$

Valamint Y várható értéke és szórása:

$$\mathbb{E}(Y) = 6; \qquad D(Y) = \sqrt{\frac{5/6}{1/36}} = \sqrt{30} = 5,477.$$

6.4. Negatív binomiális eloszlás.

Független kísérleteket végzünk, melyek mindegyike p valószínűséggel sikerül. Azt kérdezzük, hogy hányadik kísérlet volt az r. sikeres. Ez a mennyiség negatív binomiális eloszlású, az alábbi definíció alapján.

6.7. definíció (Negatív binomiális eloszlás). *Legyen $p \in (0,1)$ rögzített szám. Azt mondjuk, hogy a (k, p_k) sorozat geometriai eloszlású r renddel és p paraméterrel, ha lehetséges értékei*

$$k = r, r+1, r+2, \ldots,$$

a hozzájuk tartozó valószínűségek pedig

$$p_k = \binom{r-1}{k-1}(1-p)^{k-r}p^r.$$

6.8. megjegyzés. *Az $r = 1$ rendű, p paraméterű negatív binomiális eloszlás megegyezik a p paraméterű geometriai eloszlással.*

6.9. állítás (A negatív binomiális eloszlás tulajdonságai). *Legyen Z negatív binomiális eloszlású valószínűségi változó p paraméterrel. Ekkor*

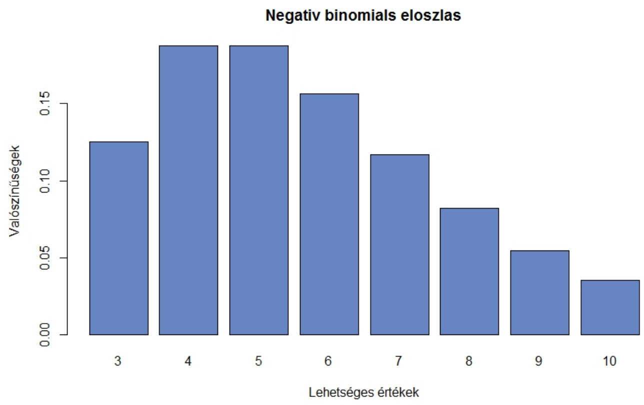
$$\mathbb{E}(Z) = \frac{r}{p}; \qquad D(Z) = \sqrt{r\frac{1-p}{p^2}}.$$

Példa: harmadik defekt. Bálint minden biciklizéskor a többitől függetlenül $1/2$ valószínűséggel kap defektet. A harmadik defekt után lecseréli a biciklibelsőt. Jelölje Z, hogy hányadik használatkor kell belsőt cserélnie. Ekkor Z lehetséges értékei $k = 3, 4, \ldots$ (az első két alkalommal biztosan nem kell cserélnie), eloszlása pedig negatív binomiális $r = 3$ renddel és $p = 1/2$ paraméterrel. Továbbá

$$\mathbb{E}(Z) = 3 \cdot 1/(1/2) = 6; \qquad D(Z) = \sqrt{3 \cdot \frac{0,5}{0,5^2}} = 1,225.$$

6.5. Poisson-eloszlás

Az alábbi eloszlás jól használható, ha kis valószínűséggel sikeres kísérletből végzünk sokat, és a sikeres kísérletek számát akarjuk egyszerűen modellezni. Például: a balesetek száma egy biztosítónál, ahol sok ügyfél mindegyike kis valószínűséggel okoz balesetet egy év alatt.



7. ábra.
A harmadik defekt eloszlása: negatív binomiális eloszlás, $r = 3$, $p = 1/2$, $k = 10$-ig.

6.10. definíció (Poisson-eloszlás). *Legyen $s > 0$. Az s paraméterű Poisson-eloszlás lehetséges értékei $k = 0, 1, 2, \ldots$, a hozzájuk tartozó valószínűségek pedig:*

$$p_k = \frac{s^k}{k!} e^{-s} \qquad (k = 0, 1, \ldots).$$

6.11. állítás (A Poisson-eloszlás tulajdonságai). *Legyen X Poisson-eloszlású valószínűségi változó s paraméterrel.*

(a) *Ekkor X várható értéke, szórása és szórásnégyzete:*

$$\mathbb{E}(X) = s; \qquad D(X) = \sqrt{s}; \qquad D^2(X) = s.$$

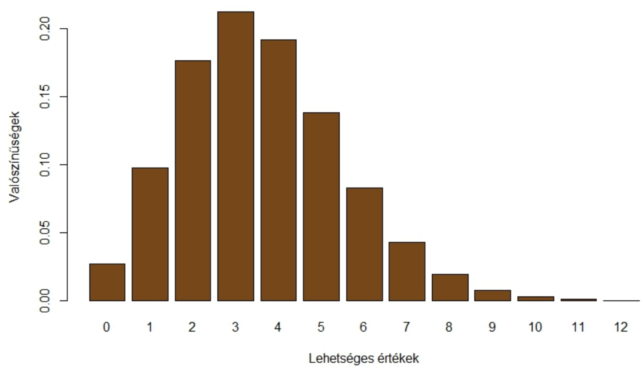
(b) *A $\mathbb{P}(X = k)$ valószínűség $k = [s]$ esetén maximális. Ha s egész, az eggyel kisebb k is a legnagyobb értéket adja.*

Példa. Jelölje X azt, hogy egy nyár alatt Budapesten hányszor esik jégeső. Tegyük fel, hogy a jégesők száma Poisson-eloszlású, $s = 3,61$ paraméterrel. Ekkor

$$\mathbb{P}(X = k) = \frac{3,61^k}{k!} e^{-3,61} \qquad (k = 0, 1, 2, \ldots).$$

Például

$$\mathbb{P}(X = 3) = \mathbb{P}(\text{pontosan háromszor esik jégeső}) = \frac{3,61^3}{6} \cdot e^{-3,61} = 0,2121.$$



8. ábra.
A nyári jégesők számának eloszlása: Poisson-eloszlás, $s = 3,61$, $k = 12$-ig.

Továbbá a jégesők számának várható értéke és szórása:

$$\mathbb{E}(X) = s = 3,61; \qquad D(X) = \sqrt{s} = \sqrt{3,61} = 1,9.$$

Poisson-eloszlás közelítése binomiálissal

A jégesők számát leíró Poisson-eloszlást összehasonlítjuk egy binomiális eloszlással. Ez utóbbi paraméterét úgy választjuk, hogy a két eloszlás várható értéke megegyezzen, vagyis $s = np$ teljesüljön.

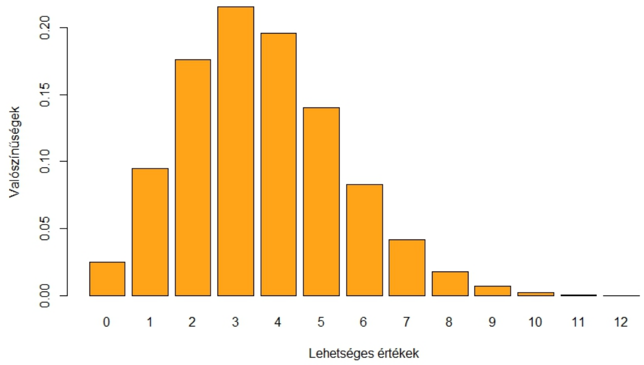
Egy évben 92 nyári nap van. Ha feltennénk, hogy minden nap a többitől függetlenül $p = 0,0392$ valószínűséggel esik jégeső, akkor ebben a modellben a jégesők száma (jelöljük Y-nal) binomiális eloszlású $n = 92$ renddel és $p = 0,0392$ paraméterrel. Ennek az eloszlásnak várható értéke és szórása:

$$\mathbb{E}(Y) = np = 92 \cdot 0,0392 = 3,61; \qquad D(X) = \sqrt{92 \cdot 0,0392 \cdot 0,9608} = 1,861.$$

Sőt például $k = 3$-ra

$$\mathbb{P}(Y = 3) = \binom{92}{3} 0,0392^3 \cdot 0,9608^{89} = 0,2153.$$

A 8. és a 9. ábrák összehasonlítása is mutatja, hogy a Poisson-eloszlás $s = 3,61$ paraméterrel és a binomiális eloszlás $n = 92$ renddel és $p = 0,0392$ paraméterrel közel vannak egymáshoz. Itt tehát $s = np$ teljesült. Táblázatban, $k = 7$-ig (tehát X a Poisson-eloszlás, és Y a binomiális):



9. ábra.
A nyári jégesők számának közelítése: binomiális eloszlás, $n = 92$, $p = 0,0392$, $k = 12$-ig.

k	0	1	2	3	4	5	6	7
$\mathbb{P}(X = k)$	0,027	0,098	0,176	0,212	0,191	0,138	0,083	0,042
$\mathbb{P}(Y = k)$	0,025	0,094	0,176	0,215	0,195	0,14	0,083	0,043

Általánosabban is igaz, hogy nagy n esetén az n rendű és p paraméterű binomiális eloszlás közel van az $s = np$ paraméterű Poisson-eloszláshoz, abban az értelemben, hogy minden rögzített k-ra a k-hoz tartozó valószínűségek közel vannak. Erről szól az alábbi állítás.

6.12. állítás (A Poisson– és binomiális eloszlás kapcsolata). *Legyen s pozitív szám, és $p_n = s/n$ minden $n = 1, 2, \dots$ egészre. Legyen X Poisson-eloszlású valószínűségi változó s paraméterrel, Y_n pedig binomiális eloszlású valószínűségi változó n renddel és p_n paraméterrel. Ekkor tetszőleges $k = 0, 1, 2, \dots$ esetén*

$$\lim_{n\to\infty} \mathbb{P}(Y_n = k) = \mathbb{P}(X = k),$$

azaz

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{s^k}{k!} e^{-s} \qquad (k = 0, 1, 2, \dots).$$

7. Eloszlásfüggvény és sűrűségfüggvény

7.1. definíció (Eloszlásfüggvény). *Legyen $X : \Omega \to \mathbb{R}$ valószínűségi változó. Ekkor X eloszlásfüggvénye az alábbi $F : \mathbb{R} \to [0,1]$ függvény:*

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \qquad \text{minden } t \in \mathbb{R} \text{ valós számra.}$$

7.2. állítás. *Legyenek $a, b \in \mathbb{R}$ tetszőleges valós számok. Ekkor*

$$\mathbb{P}(a < X \leq b) = F(b) - F(a).$$

Bizonyítás. Ez azonnal adódik az $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$ egyenlőségből (amihez lásd az 1.10. állítás bizonyítását: $A = \{X \leq a\} \subseteq \{X \leq b\} = B$) és az eloszlásfüggvény definíciójából. □

Példa. Szabályos kockával dobunk. A dobott számot jelölje X. Legyen F az X eloszlásfüggvénye. Ekkor

$$F(0) = \mathbb{P}(X \leq 0) = 0; \qquad F(1) = \mathbb{P}(X \leq 1) = 1/6;$$

$$F(\pi) = \mathbb{P}(X \leq \pi) = \mathbb{P}(X \leq 3) = 1/2; \qquad F(6) = \mathbb{P}(X \leq 6) = 1.$$

7.3. tétel (Az eloszlásfüggvény tulajdonságai). *Legyen X valószínűségi változó, F pedig az eloszlásfüggvénye. Ekkor*

(i) *F monoton növő: $a < b$ esetén $F(a) \leq F(b)$.*

(ii) $\lim_{t \to -\infty} F(t) = 0; \qquad \lim_{t \to \infty} F(t) = 1.$

(iii) *F jobbról folytonos, azaz minden $t \in \mathbb{R}$ valós számra $\lim_{s \to t-} F(s) = F(t)$.*

7.4. megjegyzés. *Ha a $G : \mathbb{R} \to [0,1]$ függvény rendelkezik a tételben szereplő $(i) - (iii)$ tulajdonságokkal, akkor van olyan valószínűségi változó, melynek G az eloszlásfüggvénye.*

7.5. definíció. *Azt mondjuk, hogy az X valószínűségi változó folytonos, ha eloszlásfüggvénye folytonos.*

Egy valószínűségi változó pontosan akkor folytonos, ha $\mathbb{P}(X = t) = 0$ teljesül minden t számra. Így például a kockadobás nem folytonos: $\mathbb{P}(X = 1) = 1/6 \neq 0$. Általánosabban, nincs olyan valószínűségi változó, mely egyszerre diszkrét és folytonos is lenne. Olyan viszont van, ami sem nem diszkrét, sem nem folytonos (például a napi csapadékmennyiség, ami pozitív valószínűséggel nulla, viszont nem megszámlálhatóan végtelen az értékkészlete).

7.6. definíció (Abszolút folytonosság és sűrűségfüggvény). *Az X valószínűségi változó abszolút folytonos, ha van olyan $f : \mathbb{R} \to \mathbb{R}$ függvény, melyre*

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} f(s)ds$$

teljesül minden $t \in \mathbb{R}$ számra. Ilyenkor az f függvényt az X valószínűségi változó sűrűségfüggvényének nevezzük.

7.7. állítás. *Legyen az X abszolút folytonos valószínűségi változó, melynek sűrűségfüggvénye f. Ekkor tetszőleges $a < b$ számokra teljesül, hogy*

$$\mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f(s)ds.$$

Példa: lépcsős sűrűségfüggvény. Tegyük fel, hogy a holnapi csapadékmennyiség sűrűségfüggvénye az alábbi:

$$f(s) = \begin{cases} 0,2, & \text{ha } s \in [0,1]; \\ 0,4, & \text{ha } s \in (1,3]; \\ 0 & \text{ha } s < 0 \text{ vagy } s > 2. \end{cases}$$

Jelölje a holnapi csapadékmennyiséget X. Ekkor annak valószínűsége, hogy holnap legfeljebb $0,5$ mm csapadék lesz:

$$\mathbb{P}(0 \leq X \leq 0,5) = \int_{0}^{0,5} f(s)ds = \int_{0}^{0,5} 0,2ds = 0,5 \cdot 0,2 = 0,1.$$

Annak valószínűsége, hogy a holnapi csapadékmennyiség 0,5 mm és 2 mm között lesz:

$$\mathbb{P}(0,5 \leq X \leq 2) = \int_{0,5}^{2} f(s)ds = \int_{0,5}^{1} 0,2ds + \int_{1}^{2} 0,4ds = 0,5 \cdot 0,2 + 1 \cdot 0,4 = 0,5.$$

7.8. állítás (Az eloszlásfüggvény és sűrűségfüggvény kapcsolata). *Legyen X abszolút folytonos valószínűségi változó, melynek F az eloszlásfüggvénye.*

(a) *Ha f az X sűrűségfüggvénye, akkor minden $t \in \mathbb{R}$ számra*

$$F(t) = \int_{-\infty}^{t} f(s)ds.$$

(b) *Az $f(t) = F'(t)$ függvény (azokra a t-kre, ahol F differenciálható) az X sűrűségfüggvénye.*

7.9. állítás (A sűrűségfüggvény tulajdonságai). *Legyen $f : \mathbb{R} \to \mathbb{R}$ az X valószínűségi változó sűrűségfüggvénye. Ekkor*

(i) *$f(s) \geq 0$ teljesül "majdnem minden" $s \in \mathbb{R}$-re (például véges vagy megszámlálható sok kivétel lehetséges).*

(ii) *$\int_{-\infty}^{\infty} f(s)ds = 1$.*

7.10. megjegyzés. *Ha a $g : \mathbb{R} \to \mathbb{R}$ függvényre teljesül a fenti (i) és (ii) tulajdonság, akkor van olyan X valószínűségi változó, melynek g a sűrűségfüggvénye.*

7.1. Abszolút folytonos valószínűségi változók várható értéke, szórása és momentumai

7.11. definíció (Várható érték, abszolút folytonos eset). *Legyen X abszolút folytonos valószínűségi változó, melynek sűrűségfüggvénye f. Ekkor X várható értéke:*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} s \cdot f(s)ds,$$

ha ez az integrál létezik és véges.

7.12. definíció (Szórásnégyzet és szórás). *Tegyük fel, hogy az X valószínűségi változó abszolút folytonos, és sűrűségfüggvénye f. Ekkor X szórásnégyzete:*

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

szórása pedig

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

ha ezek a várható értékek léteznek.

7.13. állítás (A szórásnégyzet kiszámítása). *A szórásnégyzetet a következőképpen számíthatjuk ki abszolút folytonos X valószínűségi változó esetén:*

$$D^2(X) = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2 = \int_{-\infty}^{\infty} s^2 f(s)ds - \left[\int_{-\infty}^{\infty} s \cdot f(s)ds\right]^2,$$

ahol f az X sűrűségfüggvénye.

8. Nevezetes abszolút folytonos eloszlások

8.1. Egyenletes eloszlás

8.1. definíció (Egyenletes eloszlás). *Legyenek $a < b$ valós számok. Azt mondjuk, hogy az X valószínűségi változó egyenletes eloszlású az $[a, b]$ intervallumon, ha sűrűségfüggvénye*

$$f(s) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq s \leq b; \\ 0, & \text{különben.} \end{cases}$$

Vegyük észre, hogy $f(s) \geq 0$ minden s-re, és $\int_{-\infty}^{\infty} f(s)ds = \int_a^b \frac{1}{b-a}ds = 1$, így f valóban lehet sűrűségfüggvény, a 7.9. állítás és az azt követő megjegyzés alapján.

8.2. állítás (Az egyenletes eloszlás tulajdonságai). *Legyen az X valószínűségi változó egyenletes eloszlású az $[a, b]$ intervallumon. Ekkor a következők teljesülnek.*

(i) X eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0, & \text{ha } t \leq a; \\ \frac{t-a}{b-a}, & \text{ha } a < t < b; \\ 1, & \text{ha } t \geq b. \end{cases}$$

(ii) Ha $a \leq c \leq d \leq b$, akkor

$$\mathbb{P}(c \leq X \leq d) = \int_c^d f(s)ds = \int_c^d \frac{1}{b-a}ds = \frac{d-c}{b-a}.$$

(iii) Az X valószínűségi változó várható értéke és szórása:

$$\mathbb{E}(X) = \frac{a+b}{2}; \qquad D(X) = \frac{b-a}{\sqrt{12}}.$$

Vegyük észre, hogy a várható érték a megadott intervallum közepe, a szórás pedig az intervallum hosszával arányos – hosszabb intervallum esetén nagyobb a szórás.

Példa. Csomagot várunk, a futár 10 és 12 óra között érkezik. Feltesszük, hogy érkezésének időpontja egyenletes eloszlású a $[10, 12]$ intervallumon. Ekkor az előző állítás alapján az alábbiak igazak.

- Annak valószínűsége, hogy 10 és 11 óra között érkezik: $(11-10)/(12-10) = 1/2$.
- Annak valószínűsége, hogy 10:15 és 10:30 között érkezik, $1/8 = 0,125$.
- Érkezési időpontjának várható értéke: $(10+12)/2 = 11$ óra.
- Érkezési időpontjának szórása: $(12-10)/\sqrt{12} = 1/\sqrt{3} = 0,5774$.

8.2. Normális eloszlás

8.3. definíció (Normális eloszlás). *Legyen m valós, σ pedig pozitív szám. Azt mondjuk, hogy az Y valószínűségi változó normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, ha sűrűségfüggvénye*

$$f(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(s-m)^2}{2\sigma^2} \right) \qquad (s \in \mathbb{R}).$$

Jelölése: $Y \sim N(m, \sigma^2)$.

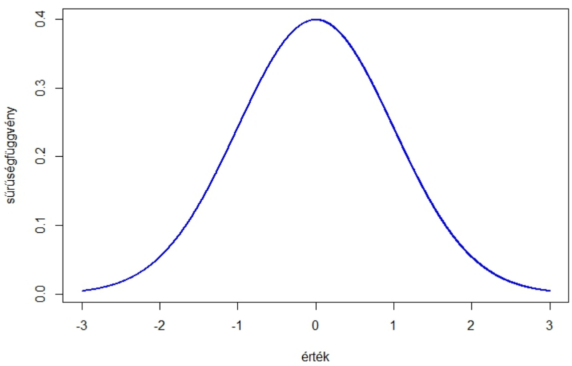
Ez az f valóban sűrűségfüggvény, és hogy az így megadott Y-ra $\mathbb{E}(Y) = m$ és $D^2(Y) = \sigma^2$ (ezeket nem bizonyítjuk).

8.4. definíció (Standard normális eloszlás). *Azt mondjuk, hogy a Z valószínűségi változó standard normális eloszlású, ha normális eloszlású $m = 0$ várható értékkel és $\sigma^2 = 1$ szórásnégyzettel, azaz $Z \sim N(0, 1)$, és sűrűségfüggvénye:*

$$f(s) = \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \qquad (s \in \mathbb{R}).$$

A standard normális eloszlás eloszlásfüggvényét Φ-vel jelöljük, vagyis

$$\Phi(t) = \mathbb{P}(Y \le t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds.$$



10. ábra. A standard normális eloszlás sűrűségfüggvénye ($m = 0, \sigma = 1$)

8.5. állítás (A normális eloszlás tulajdonságai). *Ha az Y valószínűségi változó normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, valamint $a < b$ valós számok, akkor*

(i)

$$\begin{aligned}\mathbb{P}(a < Y < b) = \mathbb{P}(a \le Y \le b) = &\\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right) ds = \\ &= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).\end{aligned}$$

(ii) $\mathbb{P}(Y < b) = \mathbb{P}(Y \le b) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{b} \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right) ds = \Phi\left(\frac{b-m}{\sigma}\right).$

(iii) $\mathbb{P}(a < Y) = \mathbb{P}(a \le Y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{a}^{\infty} \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right) ds = 1 - \Phi\left(\frac{a-m}{\sigma}\right).$

(iv) *Y várható értéke m, szórása σ.*

(v) *Legyenek u, v valós számok. Ekkor $uY+v$ is normális eloszlású valószínűségi változó, $um + v$ várható értékkel és $u^2\sigma^2$ szórásnégyzettel.*

Példa. Tegyük fel, hogy a holnapi csúcshőmérséklet (Celsius-fokban mérve) normális eloszlású $m = 2$ várható értékkel és $\sigma^2 = 9$ szórásnégyzettel. Jelöljük ezt a valószínűségi változót Y-nal. Ekkor annak valószínűsége, hogy a holnapi csúcshőmérséklet 0 és 4 $°C$ között lesz:

$$\begin{aligned}
\mathbb{P}(0 < Y < 4) = \mathbb{P}(Y < 4) - \mathbb{P}(Y < 0) &= \Phi\left(\frac{4-2}{3}\right) - \Phi\left(-\frac{2}{3}\right) = \\
&= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{2}{3}\right) = \Phi\left(\frac{2}{3}\right) - \left(1 - \Phi\left(\frac{2}{3}\right)\right) \\
&= 2\Phi\left(\frac{2}{3}\right) - 1 = 2 \cdot 0,7486 - 1 = 0,4972,
\end{aligned}$$

felhasználva a normális eloszlás táblázatát (vagy bármilyen statisztikai szoftvert) és a

$$\Phi(-x) = 1 - \Phi(x)$$

tulajdonságot (mely a standard normális eloszlás sűrűségfüggvényének 0-ra vonatkozó szimmetriájából adódik).

8.3. Exponenciális eloszlás

8.6. definíció (Exponenciális eloszlás). *Legyen $b > 0$ pozitív szám. Azt mondjuk, hogy az X valószínűségi változó exponenciális eloszlású b paraméterrel, ha sűrűségfüggvénye:*

$$f(s) = \begin{cases} be^{-bs}, & \text{ha } s > 0; \\ 0 & \text{különben.} \end{cases}$$

8.7. állítás (Az exponenciális eloszlás tulajdonságai). *Legyen X exponenciális eloszlású $b > 0$ paraméterrel. Ekkor a következők teljesülnek.*

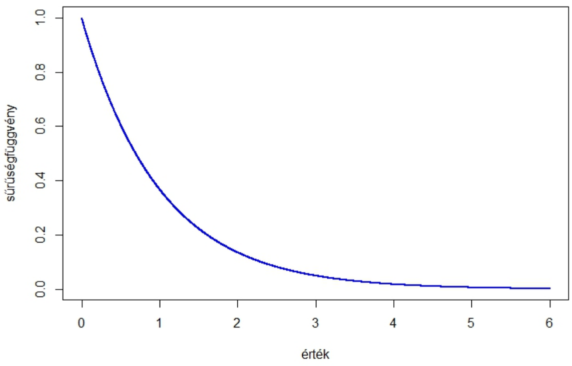
(i) X eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X < t) = \int_{-\infty}^{t} f(s)ds = \begin{cases} 1 - e^{-bs}, & \text{ha } s > 0; \\ 0 & \text{különben.} \end{cases}$$

(ii) X várható értéke: $\mathbb{E}(X) = 1/b$, szórása: $D(X) = 1/b$.

(iii) **Örökifjú tulajdonság.** *Legyenek s, t pozitív számok. Ekkor*

$$\mathbb{P}(X \geq s + t | X \geq s) = \mathbb{P}(X \geq t).$$



11. ábra. A $b = 1$ paraméterű exponenciális eloszlás sűrűségfüggvénye

Példa. Tegyük fel, hogy egy radioaktív részecske élettartamának elolszása (másodpercben mérve) 10 paraméterű exponenciális eloszlású valószínűségi változó. Jelöljük ezt X-szel. Ekkor annak valószínűsége, hogy a részecske több mint 1 másodpercig életben marad:

$$\mathbb{P}(X > 1) = 1 - F(1) = 1 - (1 - e^{-10\cdot 1}) = e^{-10} = 0,0000454.$$

Annak valószínűsége, hogy a részecske élettartama $0,1$ és $0,2$ másodperc közé esik:

$$\begin{aligned}\mathbb{P}(0,1 \leq X \leq 0,2) = F(0,2) - F(0,1) &= 1 - e^{-10\cdot 0,2} - (1 - e^{-10\cdot 0,1}) = \\ &= 1/e - 1/e^2 = 0,2325.\end{aligned}$$

A részecske élettartamának várható értéke: $\mathbb{E}(X) = 1/10 = 0,1$.

9. Valószínűségi változók együttes eloszlása

9.1. definíció (Valószínűségi vektorváltozó). *Az* $\underline{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ *függvény valószínűségi vektorváltozó, ha tetszőleges* $a_i < b_i$ $(i = 1, 2, \ldots, n)$ *valós számokra teljesül, hogy*

$$\{\omega \in \Omega : a_1 < X_1(\omega) \leq b_1, a_2 < X_2(\omega) \leq b_2, \ldots, a_n < X_n(\omega) \leq b_n\} \in \mathcal{A}.$$

Ha $\underline{X}$ *valószínűségi vektorváltozó, akkor az* X_i *valószínűségi változó eloszlását az* $\underline{X}$ i*. peremeloszlásának nevezzük.*

Az $\underline{X}$ *valószínűségi vektorváltozó diszkrét, ha értékkészlete véges vagy megszámlálhatóan végtelen.*

9.1. Együttes eloszlás– és sűrűségfüggvény

9.2. definíció. *Az* $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó együttes eloszlásfüggvénye az* $F : \mathbb{R}^n \to [0, 1]$ *függvény, melyre*

$$F(\underline{t}) = F(t_1, \ldots, t_n) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n), \text{ ha } (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

9.3. definíció. *Az* $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó abszolút folytonos, ha van olyan* $f : \mathbb{R}^n \to \mathbb{R}$ *függvény, melyre*

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ *esetén. Ilyenkor az* f *függvényt az* $\underline{X}$ *együttes sűrűségfüggvényének nevezzük.*

9.2. Valószínűségi változók függetlensége

9.4. definíció. *Azt mondjuk, hogy az* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *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.*

Végtelen hosszú sorozatra a definíció így módosul. Az $X_1, X_2, X_3 \ldots$ *valószínűségi változók függetlenek, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.*

Független valószínűségi változókra példa:

- Két kockadobásnál az elsőként (X_1) és másodikként dobott szám (X_2).
- A holnapi csapadékmennyiség Budapesten és Torontóban.
- Két találomra választott ember testmagassága.

Nem független valószínűségi változókra példa:

- Két kockadobásnál az első szám és a két dobott szám összege.
- A holnapi csapadékmennyiség Budapesten és Budaörsön.

- Két testvér testmagassága.

9.5. állítás. *Tegyük fel, hogy az $X_1, \ldots, X_n$ valószínűségi változók függetlenek, és $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ függvények. Ekkor a $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ valószínűségi változók is függetlenek. Továbbá, ha $h : \mathbb{R}^k \to \mathbb{R}$, akkor a $h(X_1, \ldots, X_k), X_{k+1}, \ldots, X_n$ valószínűségi változók is függetlenek.*

9.6. állítás (Függetlenség és sűrűségfüggvény). *Tegyük fel, hogy az $\underline{X} = (X_1, \ldots, X_n)$ valószínűségi vektorváltozó abszolút folytonos, együttes sűrűségfüggvénye f, továbbá az X_i valószínűségi változó sűrűségfüggvénye f_i minden $i = 1, 2, \ldots, n$ esetén. Ezekkel a jelölésekkel: $X_1, X_2, \ldots, X_n$ pontosan akkor függetlenek, ha*

$$f(t_1, \ldots, t_n) = f_1(t_1) \cdot f_2(t_2) \ldots f_n(t_n)$$

teljesül bármely $t_1, t_2, \ldots, t_n$ valós számokra.

9.7. állítás (Függetlenség egész értékű esetben). *Tegyük fel, hogy az $\underline{X} = (X_1, \ldots, X_n)$ valószínűségi vektorváltozó diszkrét, sőt minden peremeloszlása egész értékű. Az $X_1, X_2, \ldots, X_n$ valószínűségi változók pontosan akkor függetlenek, ha*

$$\mathbb{P}(X_1 = k_1, \ldots, X_n = k_n) = \mathbb{P}(X_1 = k_1) \cdot \mathbb{P}(X_2 = k_2) \ldots \mathbb{P}(X_n = k_n)$$

teljesül bármely $k_1, k_2, \ldots, k_n$ egész számokra.

Példa. Egy szabályos dobókockát feldobunk n-szer, jelölje X_i az i. dobás értékét ($i = 1, \ldots, n$). Ekkor $X_1, X_2, \ldots, X_n$) függetlenek. Például

$$\mathbb{P}(X_1 = 3, X_2 = 4) = \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 = 4) = \frac{1}{36};$$

$$\mathbb{P}(X_1 = 3, X_2 = 4, X_3 = 2) = \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 = 4) \cdot \mathbb{P}(X_3 = 2) = \frac{1}{6^3}.$$

Vegyük érszre, hogy ez összhangban van azzal, hogy az első két dobás eredménye 36, a az első három dobás eredménye $6^3 = 216$-féle dobássorozat lehet.

10. Várható érték, szórás, kovariancia, korreláció

10.1. A várható érték tulajdonságai

10.1. állítás. *Legyenek $X, Y, X_1, X_2, \ldots, X_n$ olyan valószínűségi változók, melyeknek várható értéke létezik vagy az 5.5. definíció, vagy a 7.11. definíció* *értelmében. Ekkor a következők teljesülnek.*

(a) Ha $a \leq X \leq b$ teljesül 1 valószínűséggel valamely a, b valós számokra, akkor $a \leq \mathbb{E}(X) \leq b$.

(b) **Konstanssal szorzás.** *Ha $c \in \mathbb{R}$, akkor*

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

(c) **Összeg várható értéke.** $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

(d) **Szorzat várható értéke független esetben.** *Ha az X és Y valószínűségi változók függetlenek, akkor $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.*

(e) Legyen $g : \mathbb{R}^n \to \mathbb{R}$ függvény. Ekkor, ha $X_1, \ldots, X_n$ diszkrét valószínűségi változók, akkor

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \sum_{(t_1, \ldots, t_n)} g(t_1, \ldots, t_n)\mathbb{P}(X_1 = t_1, \ldots, X_n = t_n),$$

ha a jobb oldal abszolút konvergens, és az összegzés az $X_1, X_2, \ldots, X_n$ lehetséges értékeire történik. Ha $X_1, \ldots, X_n$ abszolút folytonos valószínűségi változók f együttes sűrűségfüggvénnyel, akkor

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(t_1, \ldots, t_n) f(t_1, \ldots, t_n) dt_1 \ldots dt_n,$$

ha a jobb oldal abszolút konvergens.

Gyakran előfordul, hogy függvényként a k. hatványra emelést használjuk.

10.2. definíció (Momentumok). *Legyen X olyan diszkrét vagy abszolút folytonos valószínűségi változó, melyre X^k várható értéke létezik. Ekkor az X valószínűségi változó k. momentuma:*

$$\mathbb{E}(X^k) \qquad (k = 1, 2, \ldots).$$

Az állítás (*e*) része alapján, ha X diszkrét valószínűségi változó, akkor

$$\mathbb{E}(X^k) = \sum_{i=1}^{\infty} u_i^k \mathbb{P}(X = u_i),$$

ahol $u_1, u_2, \ldots$ az X lehetséges értékei. Ha X abszolút folytonos, és sűrűségfüggvénye f, akkor viszont így számolhatjuk a k. momentumot:

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} t^k f(t) dt.$$

10.2. A szórásnégyzet tulajdonságai

10.3. állítás. *Legyenek $X, Y, X_1, X_2, \ldots, X_n$ olyan valószínűségi változók, melyeknek szórása létezik. Ekkor a következők teljesülnek.*

(a) **Nemnegativitás.** $D(X) \geq 0$.

(b) $D^2(X) = 0$ *akkor és csak akkor, ha* $\mathbb{P}(X = c) = 1$ *valamilyen* $c \in \mathbb{R}$ *számra.*

(c) **Konstans hozzáadása** $D^2(X + b) = D^2(X)$ *tetszőleges* $b \in \mathbb{R}$ *számra.*

(d) **Konstanssal való szorzás.** $D^2(a \cdot X) = a^2 D^2(X)$, *és* $D(a \cdot X) = |a| D(X)$ *tetszőleges* $a \in \mathbb{R}$ *számra.*

(e) **Összeg szórásnégyzete független esetben.** *Ha X és Y függetlenek, akkor* $D^2(X+Y) = D^2(X) + D^2(Y)$. *Általánosabban, ha $X_1, X_2, \ldots, X_n$ függetlenek, akkor* $D^2(X_1 + \ldots + X_n) = D^2(X_1) + \ldots + D^2(X_n)$.

10.3. A kovariancia és tulajdonságai

10.4. definíció (kovariancia). *Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor az X és Y kovarianciája:*

$$\operatorname{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\big].$$

10.5. állítás. *Legyenek $X, Y, Z, X_1, \ldots, X_n$ olyan valószínűségi változók, melyek szórása létezik. Ekkor a következők teljesülnek.*

(a) **A kovariancia kiszámítása.** $\operatorname{cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y)$.

(b) **Szimmetria.** $\operatorname{cov}(X, Y) = cov(Y, X)$.

(c) **Konstanssal való kovariancia.** $\operatorname{cov}(X, c) = 0$, *ha* $c \in \mathbb{R}$.

(d) **Kapcsolat a szórásnégyzettel.** $\operatorname{cov}(X, X) = D^2(X)$.

(e) **Linearitás.** *Egyrészt* $\operatorname{cov}(X + Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)$, *másrészt tetszőleges* $c \in \mathbb{R}$ *számra* $\operatorname{cov}(cX, Y) = c \cdot \operatorname{cov}(X, Y)$.

(f) **Függetlenséggel való kapcsolat.** *Ha az X és Y valószínűségi változók függetlenek, akkor* $\operatorname{cov}(X, Y) = 0$.

(g) **Összeg szórásnégyzete.** $D^2(X+Y) = D^2(X) + D^2(Y) + 2\mathrm{cov}(X,Y)$. *Továbbá*

$$D^2\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} D^2(X_i) + 2\sum_{i<j} \mathrm{cov}(X,Y).$$

(h) **Különbség szórásnégyzete** $D^2(X-Y) = D^2(X) + D^2(Y)$.

Példa. Legyen X Poisson-eloszlású valószínűségi változó 2 paraméterrel. Ekkor

$$\mathrm{cov}(X+3, 2\cdot X) \overset{(e)}{=} 2\mathrm{cov}(X+3, X) \overset{(e)}{=} 2\mathrm{cov}(X,X) + 2\mathrm{cov}(3,X) = \overset{(c,d)}{=} 2D^2(X) = 2\cdot 2 = 4,$$

ahol felhasználtuk a Poisson-eloszlás szórásnégyzetére vonatkozó összefüggést (6.11. állítás (a) rész).

10.6. definíció (Korrelálatlanság). *Ha az X, Y valószínűségi változók kovarianciája 0, akkor azt mondjuk, hogy X és Y korrelálatlanok.*

10.7. állítás (Függetlenség és korrelálatlanság). *(i) Ha az X és Y valószínűségi változók függetlenek és szórásuk létezik, akkor korrelálatlanok, azaz* $\mathrm{cov}(X,Y) = 0$.

(ii) Vannak olyan U, V valószínűségi változók, melyek nem függetlenek, de korrelálatlanok, azaz $\mathrm{cov}(U,V) = 0$.

Az állítás (ii) részére példa a következő. Legyen X és Y két szabályos kockadobás, ezek függetlenek. Legyen továbbá $U = X+Y, V = X-Y$. Ekkor

$$\mathrm{cov}(U,V) = \mathrm{cov}(X+Y, X-Y) \overset{(e,d)}{=} D^(X) - \mathrm{cov}(X,Y) + \mathrm{cov}(X,Y) - D^2(X) = \overset{(f)}{=} D^2(X) - D^2(Y) = 0.$$

Ugyanakkor U és V nem függetlenek, például

$$0 = \mathbb{P}(U = 11, V = 0) \neq \mathbb{P}(U = 11)\cdot \mathbb{P}(V = 0) = \frac{3}{36}\cdot\frac{1}{6}.$$

10.4. A korrelációs együttható

10.8. definíció (Korrelációs együttható). *Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y korrelációs együtthatója:*

$$R(X,Y) = \begin{cases} \frac{\mathrm{cov}(X,Y)}{D(X)D(Y)}, & \text{ha } D(X) > 0, D(Y) > 0; \\ 0, & \text{ha } D(X) = 0 \text{ vagy } D(Y) = 0. \end{cases}$$

10.9. állítás (A korrelációs együttható tulajdonságai). *Legyenek X és Y olyan valószínűségi változók, melyek szórása létezik.*

(i) Ekkor teljesül, hogy

$$|R(X,Y)| \leq 1.$$

(ii) Legyen $a > 0$ valós szám, b tetszőleges valós szám. Ekkor

$$R(X, aX+b) = 1 \text{ és } R(X, -aX+b) = -1.$$

(iii) Tegyük fel, hogy $|R(X,Y)| = 1$. Ekkor léteznek olyan a és b valós számok, hogy az $Y = aX+b$ egyenlet 1 valószínűséggel teljesül.

Példa. Kétszer dobunk szabályos dobókockával. Kiszámítjuk az először dobott számnak és a dobott számok összegének korrelációs együtthatóját.

Ehhez legyen X az először dobott szám, Y a másodszor dobott szám. A kérdés X és $X+Y$ korrelációs együtthatója. Felhasználva, hogy X és Y függetlenek, és így egyrészt a kovarianciájuk nulla, másrészt az összegük szórásnégyzete a szórásnégyzeteik összege (10.3. állítás):

$$\begin{aligned} R(X, X+Y) &= \frac{\mathrm{cov}(X, X+Y)}{D(X)D(X+Y)} = \frac{\mathrm{cov}(X,X) + \mathrm{cov}(X,Y)}{D(X)\sqrt{D^2(X) + D^2(Y)}} \\ &= \frac{D^2(X)}{D(X)\cdot\sqrt{2}\cdot D(X)} = \frac{1}{\sqrt{2}} > 0. \end{aligned}$$

11. Egyenlőtlenségek

11.1. állítás (Markov-egyenlőtlenség). *Legyen $t > 0$ tetszőleges pozitív szám, X pedig olyan véges várható értékű valószínűségi változó, mely csak nemnegatív értékeket vesz fel, vagyis melyre $X \geq 0$ teljesül. Ekkor*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

11.2. állítás (Csebisev-egyenlőtlenség). *Legyen X véges szórású valószínűségi változó, $s > 0$ pozitív szám. Ekkor*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq s) \leq \frac{D^2(X)}{s^2}.$$

11.3. következmény. *Legyen X véges szórású valószínűségi változó, $s > 0$ pozitív szám. Ekkor*

$$\mathbb{P}(|X - \mathbb{E}(X)| < s) \geq 1 - \frac{D^2(X)}{s^2}.$$

12. Valószínűségi változók összege

12.1. Konvolúció

12.1. állítás. *Legyenek X és Y olyan független valószínűségi változók, melyek lehetséges értékei egész számok. Ekkor*

$$\mathbb{P}(X + Y = k) = \sum_{i=-\infty}^{\infty} \mathbb{P}(X = i)\mathbb{P}(Y = k - i).$$

12.2. állítás. *Legyenek X és Y egymástól független, abszolút folytonos valószínűségi változók. Legyen az X sűrűségfüggvénye f, az Y-é pedig g. Ekkor az $X + Y$ valószínűségi változó is abszolút folytonos, és sűrűségfüggvénye:*

$$h_{X+Y}(t) = \int_{-\infty}^{\infty} f(s)g(t - s)ds.$$

12.2. Nevezetes eloszlások összege

12.3. állítás (Binomiális valószínűségi változók összege). *Tegyük fel, hogy $X_1, X_2, \ldots, X_n$ független valószínűségi változók, úgy, hogy X_i eloszlása binomiális m_i renddel és p paraméterrel ($i = 1, \ldots, n$). Ekkor $X_1 + \ldots + X_n$ eloszlása binomiális eloszlás $m_1 + m_2 + \ldots + m_n$ renddel és p paraméterrel.*

12.4. állítás (Geometriai eloszlású valószínűségi változók összege). *Tegyük fel, hogy $X_1, X_2, \ldots, X_n$ független valószínűségi változók, úgy, hogy X_i geometriai eloszlású p paraméterrel ($i = 1, \ldots, n$). Ekkor $X_1 + \ldots + X_n$ eloszlása negatív binomiális eloszlás n renddel és p paraméterrel.*

12.5. állítás (Negatív binomiális eloszlás). *Legyenek $X_1, \dots, X_n$ független valószínűségi változók, úgy, hogy X_i negatív binomiális eloszlású m_i renddel és p paraméterrel ($i = 1, \dots, n$). Ekkor $X_1 + \ldots + X_n$ eloszlása negatív binomiális eloszlás $m_1 + m_2 + \ldots + m_n$ renddel és p paraméterrel.*

12.6. állítás (Poisson-eloszlás). *Legyenek $X_1, \dots, X_n$ független valószínűségi változók, úgy, hogy X_i Poisson-eloszlású s_i paraméterrel ($i = 1, \dots, n$). Ekkor*

(a) $X_1 + X_2$ Poisson-eloszlású $s_1 + s_2$ paraméterrel.

(b) $X_1 + X_2 + \ldots + X_n$ Poisson-eloszlású $s_1 + \ldots + s_n$ paraméterrel.

12.7. állítás (Normális eloszlás). *Legyenek $X_1, \dots, X_n$ független valószínűségi változók, úgy, hogy X_i normális eloszlású m_i várható értékkel és σ_i^2 szórásnégyzettel ($i = 1, \dots, n$). Ekkor*

(a) X_1+X_2 normális eloszlású m_1+m_2 várható értékkel és $\sigma_1^2+\sigma_2^2$ szórásnégyzettel.

(b) $X_1 + X_2 + \ldots + X_n$ normális eloszlású $m_1 + \ldots + m_n$ várható értékkel és $\sigma_1^2 + \ldots + \sigma_n^2$ szórásnégyzettel.

12.3. Az átlag várható értéke és szórása

12.8. állítás (Az összeg várható értéke értéke és szórása). *Tegyük fel, hogy $X_1, X_2, \dots, X_n$ független, azonos eloszlású valószínűségi változók (vagyis eloszlásfüggvényük megegyezik). Ekkor*

$$\mathbb{E}(X_1 + \ldots + X_n) = n\mathbb{E}(X_1); \qquad D(X_1 + \ldots + X_n) = \sqrt{n}D(X_1).$$

12.9. állítás (Az átlag várható értéke értéke és szórása). *Tegyük fel, hogy $X_1, X_2, \dots, X_n$ független, azonos eloszlású valószínűségi változók. Ekkor*

$$\mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1); \qquad D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1)}{\sqrt{n}}.$$

13. A nagy számok törvényei

13.1. definíció (Sztochasztikus konvergencia). *Legyen $X_1, X_2, \dots$ valószínűségi változók sorozata. Azt mondjuk, hogy ez a sorozat sztochasztikusan konvergál az Y valószínűségi változóhoz, ha minden $\varepsilon > 0$-ra*

$$\mathbb{P}(|X_n - Y| > \varepsilon) \to 0$$

teljesül $n \to \infty$ esetén.

13.2. tétel (A nagy számok gyenge törvénye). *Legyenek $X_1, X_2, \dots$ olyan valószínűségi változók, melyek függetlenek, és azonos eloszlásúak (vagyis eloszlásfüggvényük megegyezik). Tegyük fel, hogy $D(X_1)$ létezik. Ekkor*

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1)$$

sztochasztikusan $n \to \infty$ esetén.

Ez tehát azt jelenti, hogy ha $\overline{X}_n = (X_1 + X_2 + \ldots + X_n)/n$ jelöli az első n valószínűségi változó átlagát, akkor minden $\varepsilon > 0$ esetén

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty).$$

13.3. definíció (1 valószínűségű konvergencia). *Legyen $X_1, X_2, \dots$ valószínűségi változók sorozata. Azt mondjuk, hogy ez a sorozat 1 valószínűséggel konvergál a Z valószínűségi változóhoz, ha*

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \to Z(\omega)\}) = 1.$$

13.4. megjegyzés. *Ha $X_n \to Z$ teljesül 1 valószínűséggel, akkor $X_n \to Z$ sztochasztikusan is. Ennek a megfordítása viszont nem igaz (van olyan sztochasztikusan konvergens sorozat, mely nem 1 valószínűséggel konvergens).*

13.5. tétel (A nagy számok erős törvénye). *Legyenek $X_1, X_2, \dots$ valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel még, hogy $\mathbb{E}(X_1)$ létezik. Ekkor*

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1)$$

teljesül 1 valószínűséggel $n \to \infty$ esetén.

14. Centrális határeloszlástétel

14.1. definíció (Eloszlásbeli konvergencia). *Legyen $X_1, X_2, \ldots$ valószínűségi változók sorozata, X_i eloszlásfüggvénye F_i ($i = 1, 2, \ldots$ esetén). Legyen továbbá Y valószínűségi változó, melynek eloszlásfüggvénye F. Azt mondjuk, hogy az $(X_n)_{n\in\mathbb{N}}$ sorozat tart Y-hoz eloszlásban, ha*

$$F_n(t) \to F(t) \qquad (n \to \infty)$$

teljesül minden olyan $t \in \mathbb{R}$-re, melyre F folytonos t-ben.

14.2. tétel (Centrális határeloszlástétel). *Legyenek $X_1, X_2, \ldots$ független azonos eloszlású valószínűségi változók, melyeknek szórása létezik. Használjuk a következő jelöléseket: $\mathbb{E}(X_1) = m$ és $D(X_1) = s$. Legyen Y standard normális valószínűségi változó: $Y \sim N(0, 1)$. Ekkor*

$$\frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{s\sqrt{n}} \to Y$$

eloszlásban $n \to \infty$ esetén. Vagyis tetszőleges $a < b$ valós számokra teljesül, hogy

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{s\sqrt{n}} < b\right) = \mathbb{P}(a \le Y < b),$$

azaz (mivel Y standard normális eloszlású):

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{s\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Ez utóbbi összefüggést így is átfogalmazhatjuk:

$$\mathbb{P}(nm + as\sqrt{n} \le X_1 + X_2 + \ldots + X_n < nm + bs\sqrt{n}) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

teljesül $n \to \infty$ esetén.

Példa. Legyenek $X_1, X_2, \ldots$ független $s = 0,5$ paraméterű exponenciális eloszlású valószínűségi változók. Ezek azonos eloszlásúak, véges szórásúak, és

$$\mathbb{E}(X_1) = \frac{1}{s} = 2; \qquad D(X_1) = \frac{1}{s} = 2.$$

A nagy számok erős törvénye szerint

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}X_j = 2$$

1 valószínűséggel.

A centrális határeloszlástétel szerint tetszőleges $a < b$ számokra

$$\lim_{n\to\infty}\mathbb{P}\left(a \le \frac{\sum_{j=1}^{n}X_j - 2n}{2\sqrt{n}} < b\right) = \int_a^b e^{-x^2/2}dx = \Phi(b) - \Phi(a).$$

Átrendezve:

$$\lim_{n\to\infty}\mathbb{P}\left(2a\sqrt{n} \le \sum_{j=1}^{n}X_j - 2n < 2b\sqrt{n}\right) = \int_a^b e^{-x^2/2}dx = \Phi(b) - \Phi(a).$$

$$\lim_{n\to\infty}\mathbb{P}\left(2n + 2a\sqrt{n} \le \sum_{j=1}^{n}X_j < 2n + 2b\sqrt{n}\right) = \int_a^b e^{-x^2/2}dx = \Phi(b) - \Phi(a).$$

$$\lim_{n\to\infty}\mathbb{P}\left(2 + \frac{2a}{\sqrt{n}} \le \frac{\sum_{j=1}^{n}X_j}{n} < 2 + \frac{2b}{\sqrt{n}}\right) = \int_a^b e^{-x^2/2}dx = \Phi(b) - \Phi(a).$$

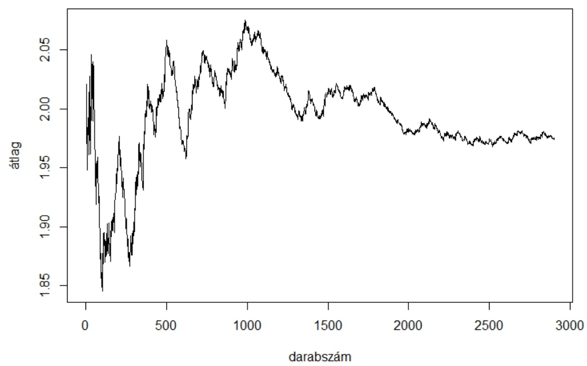
Tehát például $a = -1$ és $b = 1$ választással

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left(2 - \frac{2}{\sqrt{n}} \le \frac{\sum_{j=1}^{n}X_j}{n} < 2 + \frac{2}{\sqrt{n}}\right) &= \int_{-1}^{1} e^{-x^2/2}dx = \Phi(1) - \Phi(-1)\\ &= \Phi(1) - [1 - \Phi(1)] = 2\Phi(1) - 1 = 2\cdot 0,8413 - 1 = 0,6826.\end{aligned}$$

Ugyanakkor például

$$\lim_{n\to\infty}\mathbb{P}\left(1,99993 \le \frac{\sum_{j=1}^{n}X_j}{n} < 2,00008\right) = 1.$$

A 12. ábrán a következő látható. $X_1, X_2, \ldots, X_{3000}$ független exponenciális eloszlású valószínűségi változók $s = 0,5$ paraméterrel, mint előbb. Minden $n = 100, 101, \ldots, 3000$-re kiszámítjuk az $\frac{1}{n}(\sum_{j=1}^{n}X_j)$ átlagot, és ezt ábrázoljuk n függvényében. Az átlag $\mathbb{E}(X_1) = 1/s = 2$-höz konvergál a nagy számok erős törvénye szerint 1 valószínűséggel. Az ábrán az átlag nem megy nagyon közel a kettőhöz, de elég nagy n-re $1,95$ és $2,05$ közé esik. Az átlag szórása $n = 3000$-re $2/\sqrt{3000} = 0,0365$ a 12.9. állítás szerint.



12. ábra.
Az átlag változása a darabszám függvényében független $\exp(0,5)$ eloszlású mintánál

15. További nevezetes abszolút folytonos eloszlások

15.1. definíció (gamma-függvény). *Ha $a > 0$ pozitív szám, legyen*

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Parciális integrálással kiszámítható, hogy $\Gamma(a) = (a-1)\Gamma(a-1)$ minden $a > 1$-re, és így $\Gamma(n) = (n-1)!$, ha n pozitív egész.

15.2. definíció (gamma-eloszlás). *Legyenek a és λ pozitív számok. Azt mondjuk, hogy az X valószínűségi változó gamma-eloszlású a renddel és λ paraméterrel, ha sűrűségfüggvénye*

$$f(t) = \begin{cases} \frac{\lambda^a t^{a-1}}{\Gamma(a)} e^{-\lambda t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

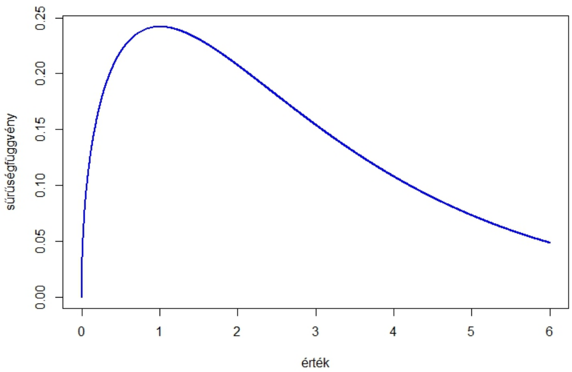
15.3. definíció (χ^2-eloszlás). *Legyenek $X_1, X_2, \ldots, X_q$ független standard normális eloszlású valószínűségi változók. Az*

$$Y = X_1^2 + X_2^2 + \ldots + X_n^2$$

valószínűségi változó eloszlását q szabadsági fokú χ^2-eloszlásnak nevezzük. Ennek sűrűségfüggvénye:

$$f_1(t) = \begin{cases} \frac{t^{q/2-1}}{2^{q/2}\Gamma(q/2)} e^{-t/2}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

A q szabadsági fokú χ-négyzet eloszlás megegyezik az $a = q/2$ rendű és $\lambda = 1/2$ paraméterű Γ-eloszlással.



13. ábra. Az $n = 3$ szabadsági fokú χ^2-eloszlás sűrűségfüggvénye

15.4. definíció (F-eloszlás). *Legyenek m, n pozitív egészek, $X_1, \ldots, X_m$, $Y_1, Y_2, \ldots, Y_n$ pedig független standard normális eloszlású valószínűségi változók. Ekkor az*

$$F = \frac{n(X_1^2 + X_2^2 + \ldots + X_m^2)}{m(Y_1^2 + Y_2^2 + \ldots + Y_n^2)}$$

valószínűségi változó eloszlását m, n paraméterű F-eloszlásnak nevezzük.

15.5. definíció (t-eloszlás). *Legyenek $X_1, X_2, \ldots, X_n$ és Y független standard normális eloszlású valószínűségi változók. Ekkor a*

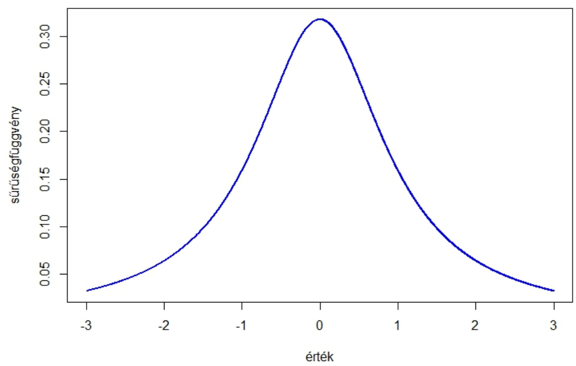
$$Z = \frac{Y}{\sqrt{(X_1^2 + X_2^2 + \ldots + X_n^2)/n}}$$

valószínűségi változó eloszlását n szabadsági fokú t-eloszlásnak (vagy Student-eloszlásnak) nevezzük.

15.6. definíció (Cauchy-eloszlás). *Az $n = 1$ szabadsági fokú t-eloszlást Cauchy-eloszlásnak nevezzük. Vagyis, ha X és Y független standard normális eloszlású valószínűségi változók, akkor X/Y Cauchy-eloszlású. Ennek sűrűségfüggvénye:*

$$f_2(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

A Cauchy-eloszlásnak sem várható értéke, sem szórása nem létezik.

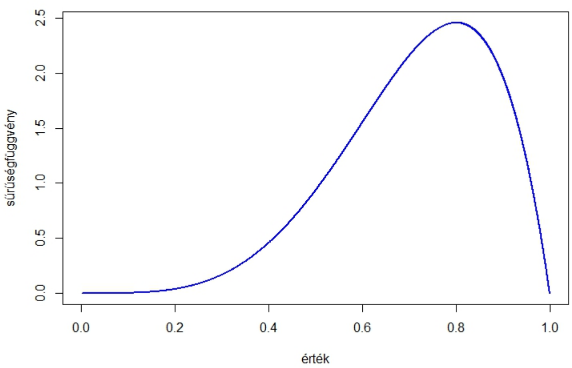


14. ábra. A Cauchy-eloszlás sűrűségfüggvénye

15.7. definíció (beta-eloszlás). *Legyenek $a, b > 1$ számok. Azt mondjuk, hogy az U valószínűségi változó beta-eloszlású a és b paraméterekkel, ha sűrűségfüggvénye*

$$f_3(t) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}, & t \in [0, 1]; \\ 0, & t < 0 \text{ vagy } t > 1. \end{cases}$$

Vegyük észre, hogy a sűrűségfüggvény csak a $[0, 1]$ intervallumon vesz fel pozitív értékeket, vagyis a beta-eloszlásból sorsolt értékek mindig 0 és 1 közé esnek.



15. ábra. A beta-eloszlás sűrűségfüggvénye $a = 5$ és $b = 2$ paraméterekkel

Impresszum

A jegyzet első változata 2015 nyarán készült Budapesten, az Eötvös Loránd Tudományegyetem Természettudományi Karának földtudomány alapszak másodéves valószínűségszámítás előadásához.

Köszönetnyilvánítás

- Csiszár Villőnek, akinek az informatikus szak számára készült valószínűségszámítás jegyzetére támaszkodva épült fel ez az előadás.
- Varga Lászlónak, konzultációért, kiegészítésekért, javaslatokért.

További ajánlott irodalom

(1) Csiszár Villő: Valószínűségszámítás jegyzet. 2009.
`http://www.cs.elte.hu/~villo/esti/valszam.pdf`

(2) Denkinger Géza: Valószínűségszámítás. Nemzeti Tankönyvkiadó, Budapest, 2001.

(3) Denkinger Géza: Valószínűségszámítási gyakorlatok. Nemzeti Tankönyvkiadó, Budapest, 1999.