Valószínűségszámítás feladatok

Kevei Péter

2020. december 1.

Tartalomjegyzék

1. Alapfogalmak

1.1. Események

1.1.1. Egy szabályos érmét kétszer feldobunk. Adjuk meg a kísérlet egy matematikai modelljét! Adjuk meg azt az eseményt, hogy (i) dobunk fejet; (ii) két fejet dobunk!

Oldjuk meg a feladatot, ha az érme cinkelt, és a fejdobás valószínűsége p!

Megoldás. Kétszer dobjuk fel az érmét, ezért van egy első és egy második dobás. Tehát a lehetséges kimenetelek

$$\Omega = \{(F, F), (F, I), (I, F), (I, I)\}.$$

Figyeljünk a jelölésre: (F, I) azt jelenti, hogy az első dobás fej, a második írás, azaz ez egy vektor, aminek van első és van második komponense. A $\{\}$-zárójel halmazt jelöl, amiben az elemeknek nincsen sorrendje. Ez egy fontos formalizmus.

A lehetséges kimenetelek száma 4. Ekkor az események halmazát választhatjuk a hatványhalmaznak, azaz az Ω összes részhalmaza esemény. Ezek

$$\begin{aligned}\mathcal{A} = 2^\Omega = \Big\{&\emptyset, \{(F, F)\}, \{(F, I)\}, \{(I, F)\}, \{(I, I)\},\\ &\{(F, F), (F, I)\}, \{(F, F), (I, F)\}, \{(F, F), (I, I)\},\\ &\{(F, I), (I, F)\}, \{(F, I), (I, I)\}, \{(I, F), (I, I)\},\\ &\{(F, F), (F, I), (I, F)\}, \{(F, F), (F, I), (I, I)\},\\ &\{(F, F), (I, F), (I, I)\}, \{(F, I), (I, F), (I, I)\},\\ &\{(F, F), (F, I), (I, F), (I, I)\}\Big\}\end{aligned}$$

Egy n elemű halmaznak 2^n részhalmaza van, azaz $|2^\Omega| = 2^{|\Omega|}$ (milyen praktikus jelölés), azaz esetünkben $|\mathcal{A}| = 2^4 = 16$. Ez még kis Ω esetén is elég nagy, úgyhogy többször nem írom ezeket ki.

Az az esemény, hogy dobunk fejet azt jelenti, hogy vagy az első, vagy (nem kizáró vagy!) a második dobás fej. Azaz

$$A = \{\text{dobunk fejet}\} = \{(F, F), (F, I), (I, F)\}.$$

Azaz a megfelelő halmaznak 3 eleme van, ez egy összetett esemény.

Az az esemény, hogy két fejet dobunk az egyetlen (F, F) kimenetelt tartalmazza, azaz

$$B = \{\text{két fejet dobunk}\} = \{(F, F)\},$$

ez egy elemű, tehát elemi esemény.

Vegyük észre, hogy eddig nem kellett az, hogy az érme cinkelt-e vagy szabályos. Csak a lehetséges kimenetelek kellettek, azaz a fontos, hogy hányszor dobjuk fel az érmét. Szabályos érme esetén a fej és az írás valószínűsége egyaránt 1/2, ahonnan látjuk, hogy mind a négy lehetséges kimenetel egyformán valószínű. Tehát klasszikus valószínűségi mezőnk van, így $\mathbf{P} : \mathcal{A} \to [0, 1]$,

$$\mathbf{P}(A) = \frac{|A|}{4}.$$

Cinkelt érme esetén p a fej valószínűsége, $1 - p$ az írásé. Ekkor, a (F, F) valószínűsége p^2, hiszen mindkétszer fejet kaptunk, azaz

$$\mathbf{P}(\{(F, F)\}) = p^2.$$

Kicsit körülményes a formalizmus, de ilyen. Valószínűsége eseménynek van. Tehát annak az (F, F) elemi eseménynek a valószínűségét vizsgáljuk. Mivel a valószínűség additív halmazfüggvény, elég az elemi események valószínűségét megadni. Ezek

$$\mathbf{P}(\{(F, I)\}) = p \cdot (1-p), \quad \mathbf{P}(\{(I, F)\}) = p \cdot (1-p), \quad \mathbf{P}(\{(I, I)\}) = (1-p)^2.$$

□

1.1.2. Adjuk meg a lottóhúzást leíró valószínűségi mezőt! Mennyi annak a valószínűsége, hogy pont 3 találatunk lesz? Mennyi a valószínűsége, hogy lesz találatunk?

Megoldás. A klasszikus ötöslottón az $1, 2, \ldots, 90$ számok közül húznak ki véletlenszerűen ötöt. Azaz a kísérlet lehetséges kimenetelei számötösök. Mivel a számok kihúzási sorrendje nem számít, ezért növekvő sorrendbe rendezzük az öt számot. Tehát

$$\Omega = \{(a_1, a_2, \ldots, a_5) : 1 \leq a_1 < a_2 < \ldots < a_5 \leq 90\}.$$

Egy 90 elemű halmaz 5 elemű részhalmazainak száma

$$|\Omega| = \binom{90}{5} = \frac{90 \cdot 89 \cdot \ldots \cdot 86}{5!} = 43\,949\,268.$$

Véges alaphalmaz esetén az események halmazát választhatjuk a hatványhalmaznak, azaz

$$\mathcal{A} = 2^{\Omega} = \{\emptyset, \{(1, 2, 3, 4, 5)\}, (1, 2, 3, 4, 6), \ldots, \Omega\}.$$

Mivel bármely számötöst ugyanakkora valószínűséggel húznak ki, ez egy *klasszikus valószínűségi mező*, azaz $\mathbf{P} : \mathcal{A} \to [0,1]$,

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|}.$$

Egy szelvénnyel játszunk, a számaink: $1 \leq a_1 < a_2 < a_3 < a_4 < a_5 \leq 90$. (Játszhatunk az $1,2,3,4,5$ számokkal is.) Jelölje A azt az eseményt, hogy pontosan 3 találatunk lesz. Ez úgy lehetséges, hogy az $a_1, \ldots, a_5$ számok közül választunk 3-at, amit $\binom{5}{3}$-féleképp tehetünk meg, és a nem megjelölt 85 szám közül ($\{1,\ldots,90\} \setminus \{a_1,\ldots,a_5\}$ halmazból) kiválasztunk 2 számot. Ez

$$\binom{5}{3} \cdot \binom{85}{2}$$

lehetőség, azaz ennyi a kedvező esetek száma. Tehát a valószínűség

$$\mathbf{P}(A) = \frac{\binom{5}{3} \cdot \binom{85}{2}}{\binom{90}{5}}.$$

Jelölje B azt az eseményt, hogy lesz találatunk. Ez lehet úgy, hogy pontosan 1, pontosan 2, …, pontosan 5 találatunk van. Ehelyett azt számoljuk, hogy nem lesz találatunk, mert ez „egyféleképpen lehet” úgy, hogy a 85 nem megjelölt számból húznak ki 5-öt. Ez

$$\binom{85}{5}$$

lehetőség. Tehát a valószínűség

$$\mathbf{P}(B) = 1 - \mathbf{P}(B^c) = 1 - \frac{\binom{85}{5}}{\binom{90}{5}}.$$

$\square$

1.1.3. Fejezzük ki az A, B, C halmazokkal az alábbi eseményeket!

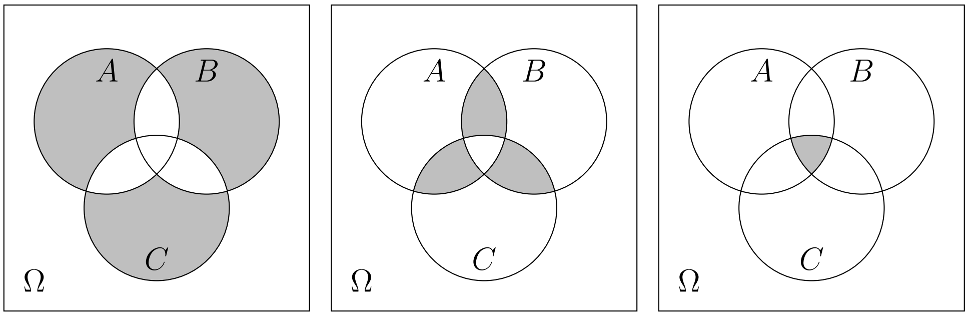
(a) Az A, B, C események közül pontosan $k \in \{1,2,3\}$ következik be.

(b) Az A, B, C események közül legalább k következik be.

(c) Az A, B, C események közül legfeljebb k következik be.

Megoldás. Csak az (a) részt csináljuk meg. Ha pontosan egy esemény következik be, akkor valamelyik bekövetkezik *és* a másik kettő nem, azaz ez az esemény

$$\{\text{pontosan 1 következik be}\} = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C).$$



1. ábra. $k = 1, 2, 3$

Hasonlóan

$$\{\text{pontosan 2 következik be}\} = (A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C),$$

és ha mindhárom bekövetkezik az a legegyszerűbb, hiszen

$$\{\text{pontosan 3 következik be}\} = A \cap B \cap C.$$

$\square$

1.1.4. Három kockával dobva mennyi a valószínűsége, hogy a dobott számok összege 4? Adjuk meg a kísérletet leíró valószínűségi mezőt, ha 3 különböző kockával dobunk, ill. ha 3 egyformával!

Megoldás. Ha három különböző kockával dobunk, akkor világos, hogy a lehetséges kimenetelek halmaza

$$\Omega = \{(i, j, k) : 1 \leq i, j, k \leq 6\}.$$

Egy $\omega = (i, j, k) \in \Omega$ esetén i jelöli az első kockán dobott számot, j a másodikon dobott számot, k pedig a harmadikon dobott számot. Világos, hogy $|\Omega| = 6^3$. Az is világos, hogy minden kimenetel egyformán valószínű, ezért egy klasszikus valószínűségi mezőnk van.

A dobott számok összege úgy lehet 4, ha két 1-est és egy 2-est dobunk, azaz

$$A = \{\text{az összeg 4}\} = \{(1, 1, 2), (1, 2, 1), (2, 1, 1)\}.$$

Ez egy összetett esemény, aminek a valószínűsége

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{6^3} = \frac{1}{72}.$$

Ha a kockák látszólag egyformák, akkor mondhatjuk, hogy a lehetséges kimenetelek halmaza

$$\widetilde{\Omega} = \{(i, j, k) : 1 \leq i \leq j \leq k \leq 6\}.$$

Itt az az $\widetilde{A}$ esemény, hogy a dobott számok összege 4 egy elemi esemény, hiszen az egyetlen lehetséges kimenetel $(1, 1, 2)$.

Itt már az elemszámot is nehezebb meghatározni, de nem ez a fő baj. Vegyük észre, hogy az $(1, 1, 1)$ lehetséges kimenetel azt jelenti, hogy mindhárom kockával 1-est dobtunk. Ugyanakkor az $(1, 1, 2)$ kimenetel azt jelenti (ebben a modellben nincs $(1, 2, 1)$), hogy két kockával 1-est dobtunk, a harmadikkal pedig 2-est. Ekkor persze az $(1, 1, 2)$ kimenetel valószínűsége nagyobb, egész pontosan háromszorosa az $(1, 1, 1)$ kimenetel valószínűségének. Tehát a valószínűségi mező nem klasszikus. Persze a keresett valószínűség nem változik.

A tanulság: *Mindig különböztessük meg a kockákat!*. $\square$

1.1.5. Egy szabályos dobókockával egyszer dobunk. Adjuk meg a kísérletet leíró valószínűségi mezőt! Adjuk meg azt az eseményt és annak valószínűségét, hogy (i) páros számot dobunk; (ii) prímet dobunk; (iii) hatost dobunk!

1.1.6. Egy szabályos dobókockával kétszer dobunk. Adjuk meg a kísérletet leíró valószínűségi mezőt! Adjuk meg azt az eseményt és annak valószínűségét, hogy (i) két hatost dobunk; (ii) dobunk hatost; (iii) mindkétszer páratlan számot dobunk!

1.1.7. Egy szabályos érmét tízszer feldobunk. Adjuk meg a kísérlet egy matematikai modelljét! Adjuk meg azt az eseményt és annak valószínűségét, hogy (i) nem dobunk fejet; (ii) az első dobás fej; (iii) pontosan 5 fejet dobunk!

Oldjuk meg a feladatot, ha az érme cinkelt, és a fejdobás valószínűsége p!

1.1.8. Legyenek $A_1, \ldots, A_n$ tetszőleges események. Mit jelent az $A_1 \circ \ldots \circ A_n$ esemény?

A $\circ$ a szimmetrikus differenciát jelöli, azaz $A \circ B = (A - B) \cup (B - A)$.

([2] 1.1.7)

1.1.9. Igazoljuk az alábbi formulák helyességét!

(a) $A \circ B = (A \cup B) - A \cap B$;

(b) $A \cap (B - C) = A \cap B - A \cap C$;

(c) $A - (A - (B - C)) = A \cap B \cap C^c$;

(d) $A \cup B = A \circ B \circ (A \cap B)$.

1.1.10. Bizonyítsuk be, hogy tetszőleges $A_1, A_2, \ldots, A_n$ események esetén

$$\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots + \mathbf{P}(A_n) - (n-1).$$

([2] 1.2.2)

1.1.11. Bizonyítsuk be, hogy tetszőleges A, B, C eseményekre

(a) $\mathbf{P}(A \circ C) \leq \mathbf{P}(A \circ B) + \mathbf{P}(B \circ C)$;

(b) ha $\mathbf{P}(A \circ B) = 0$ akkor $\mathbf{P}(A) = \mathbf{P}(B)$;

(c) $|\mathbf{P}(A \cap B) - \mathbf{P}(A \cap C)| \leq \mathbf{P}(B \circ C)$.

([2] 1.2.3 és 1.2.4)

1.1.12. Bizonyítsuk be, hogy tetszőleges A és B eseményre

$$|\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| \leq \frac{1}{4}.$$

Mikor van egyenlőség? ([2] 1.2.5)

1.1.13. Igazoljuk, hogy tetszőleges A és B eseményre

$$\mathbf{P}^2(A \cap B) + \mathbf{P}^2(A \cap B^c) + \mathbf{P}^2(A^c \cap B) + \mathbf{P}^2(A^c \cap B^c) \geq \frac{1}{4}.$$

Mikor van egyenlőség?

1.1.14. Igazoljuk, hogy tetszőleges $A_1, \ldots, A_n$ eseményekre

$$-\left(1 - \frac{1}{n}\right)^n \leq \mathbf{P}(A_1 \cap \ldots \cap A_n) - \mathbf{P}(A_1) \ldots \mathbf{P}(A_n) \leq \left(1 - \frac{1}{n}\right) n^{-1/(n-1)}.$$

(Vígh Viktor, Polygon)

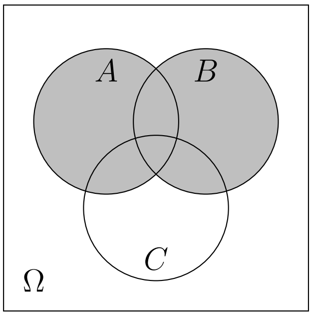
1.2. Szita formula

1.2.1. Egy pénzügyi befektető cég három cégbe fekteti pénzét, melyek rendre $0,19$, $0,25$, illetve $0,28$ valószínűséggel mennek csődbe az elkövetkező öt évben. Annak a valószínűsége, hogy az első és a második cég is csődbe megy $0,05$, hogy az első és a harmadik is csődbe megy $0,1$, míg hogy a második és a harmadik is becsődöl annak is $0,1$. Annak az esélye, hogy mindhárom cég becsődöl 2%. Mennyi a valószínűsége, hogy

(a) az első vagy a második cég csődbe megy?

(b) egyik cég sem megy csődbe?

(Szűcs Gábor feladata)



2. ábra. Az első vagy a második csődbe megy.

Megoldás. Ez egy egyszerű szita-formulás feladat. Jelölje A, B, illetve C azt az eseményt, hogy az első, második, illetve harmadik cég csődbe megy az első öt évben. A feladat megadja a egyes, kettes és hármas metszetek valószínűségét.

(a) Az, hogy az első vagy a második cég csődbe megy a $A \cup B$ esemény. Ennek a valószínűsége

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) = 0,19 + 0,25 - 0,05 = 0,39.$$

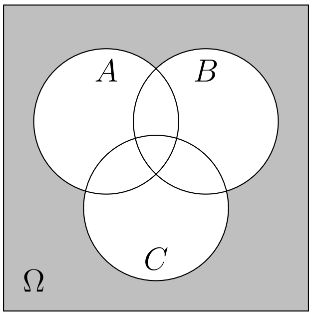
(b) Az, hogy egyik cég sem megy csődbe, az a $(A \cup B \cup C)^c$ esemény. A szita formula szerint

$$\begin{aligned}\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) &+ \mathbf{P}(B) + \mathbf{P}(C) \\ &- \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) \\ &+ \mathbf{P}(A \cap B \cap C) \\ &= 0,19 + 0,25 + 0,28 - 0,05 - 0,1 - 0,1 + 0,02 \\ &= 0,45.\end{aligned}$$

□

1.2.2. A Faluvégi Kurta Kocsma előtt 5 bicikli áll. Záróra előtt egymás után jön ki az 5 tulajdonos, és mindegyikük véletlenszerűen választ egy kerékpárt. Mennyi a valószínűsége, hogy senki sem a saját biciklijén jutott haza?

1.2.3. A Jonas Brothers nevű együttes újra összeáll és koncertet adnak. A PepsiCo cég a következő ötlettel áll elő: a kólásüvegeik kupakjában elrejtik a banda egy-egy tagjának a nevét és azok között, akik összegyűjtik mindhárom nevet kisorsolnak egy VIP belépőt. Kevin neve a kupakok felén szerepel, Joe-val a kupakok egyharmadában találkozhatunk és Nick a legritkább, neve átlagosan minden hatodik kupakban szerepel. Mennyi a valószínűsége,



3. ábra. Egyik sem megy csődbe.

hogy 5 kólát vásárolva sikerül kigyűjtenünk a három testvért? (Segítség: a kupakokra gondoljunk úgy mintha egy zsákból húznánk egy nevet, melyben Kevin háromszor, Joe kétszer, Nick pedig egyszer szerepel.)

1.2.4. Két szabályos kockát r-szer feldobunk. Legyen p_r annak a valószínűsége, hogy az $(1,1),(2,2),\ \ldots,(6,6)$ mindegyike legalább egyszer előfordul. Határozzuk meg p_r-et! ([2] 1.2.13)

1.2.5. Sorban elhelyezett n dobozba találomra berakunk N golyót úgy, hogy az összes elhelyezés egyformán valószínű. Mennyi a valószínűsége, hogy az első k doboz egyike sem üres? ([2] 1.2.14)

1.2.6. Egy urnában k-féle színű golyó van, mindegyik színűből ugyanannyi darab. Egyenként húzunk a golyókból úgy, hogy minden húzás után visszatesszük a kihúzott golyót, és minden húzásnál bármelyik golyó ugyanolyan valószínűséggel kerülhet kihúzásra.

(a) Mennyi annak a q_n valószínűsége, hogy legalább n húzás kellett ahhoz, hogy minden szín előforduljon?

(b) Mennyi annak a p_n valószínűsége, hogy n húzás során minden szín előfordult, és ez az n-edik húzásnál következik be először (vagyis az első $(n-1)$ húzás során csak $(k-1)$ szín fordult elő) ?

([2] 1.2.15)

Megoldás. Ez egy kicsit nehezebb szita-formulás feladat.

(a) Az, hogy legalább n húzás kell, hogy minden szín előforduljon pontosan azt jelenti, hogy $n-1$ húzás után még nem volt minden szín. Jelölje $A_{i,n-1} = A_i k$ azt az eseményt, hogy az első $n-1$ húzás során nem volt i színű golyó. Ekkor az az esemény, hogy $n-1$ húzás után nem volt minden szín, pontosan azt jelenti, hogy az A_1, A_2, …, A_k események közül legalább egy bekövetkezett, azaz

$$C_n := \{\text{legalább } n \text{ húzás kell}\} = \cup_{i=1}^{k} A_i.$$

Az A_i események nem kizáróak, ezért az unió valószínűségét szita-formulával határozhatjuk meg. Eszerint

$$\begin{aligned}\mathbf{P}(A_1 \cup \ldots \cup A_k) &= \mathbf{P}(A_1) + \ldots + \mathbf{P}(A_k) \\ &\quad - [\mathbf{P}(A_1 \cap A_2) + \ldots + \mathbf{P}(A_{k-1} \cap A_k)] \\ &\quad \ldots \\ &\quad \pm \mathbf{P}(A_1 \cap \ldots \cap A_k) \\ &= \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \mathbf{P}(A_1 \cap \ldots \cap A_j),\end{aligned}$$

ahol az utolsó egyenlőségnél felhasználtuk, hogy a j-es metszetek valószínűségei megegyeznek (hát persze, ugyanakkora valószínűséggel nem volt sem piros sem kék, mint sárga meg zöld. A metszetek valószínűségeit könnyű meghatározni. Valóban

$$\mathbf{P}(A_1) = \frac{(k-1)^{n-1}}{k^{n-1}},$$

hiszen az összes eset k^{n-1}, mert az $n-1$ húzás során mindig k-féle golyót kaphatunk, és a kedvező esetek száma meg $(k-1)^{n-1}$, hiszen 1-es színű golyót nem húztam, bármi más lehetett. Hasonlóan,

$$\mathbf{P}(A_1 \cap A_2) = \frac{(k-2)^{n-1}}{k^{n-1}},$$

hiszen ekkor már sem 1-es sem 2-es színű golyót nem húzhattam. Általánosan

$$\mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_j) = \frac{(k-j)^{n-1}}{k^{n-1}}.$$

Ezt visszahelyettesítve

$$\begin{aligned}\mathbf{P}(\cup_{j=1}^{k} A_j) &= \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \mathbf{P}(A_1 \cap \ldots \cap A_j) \\ &= \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \frac{(k-j)^{n-1}}{k^{n-1}}.\end{aligned}$$

(b) Legyen D_n az az esemény, hogy pontosan n húzás kellett. Könnyű látni, hogy

$$D_n = C_n \backslash C_{n+1},$$

hiszen ha pontosan n kellett, akkor legalább n kellett, de nem kellett $n+1$. Nyilván $C_n \supset C_{n+1}$, hiszen ha legalább $n+1$ húzás kellett, akkor legalább n, ezért

$$\mathbf{P}(D_n) = \mathbf{P}(C_n) - \mathbf{P}(C_{n+1}).$$

□

1.2.7. Egy kockát addig dobunk, amíg mind a 6 szám elő nem fordul. Legyen p_n annak a valószínűsége, hogy ez először az n-edik dobás után következik be. Határozzuk meg p_n-et! ([2] 1.2.16)

1.2.8. Tekintsük n elem véletlen permutációját. Mekkora a valószínűsége, hogy nem lesz fixpont? Mekkora a valószínűsége, hogy pontosan k fixpont lesz?

1.3. Kombinatorikus valószínűség

1.3.1. Hét törpe közül Hófehérke leültet ötöt egy kör alakú asztalhoz. Tegyük fel, hogy az összes lehetséges elrendezés egyformán valószínű. Mennyi a valószínűsége, hogy Morgó és Kuka nem kerül egymás mellé?

Megoldás. Minden lehetséges választás egyformán valószínű, ezért klasszikus a valószínűségi mező. Tehát kedvező (vagy kedvezőtlen) és összes esetet kell számolni. Először mindig az összes esetet számoljuk.

Hófehérke $\binom{7}{5}$-féleképp választhat ki 7 törpe közül 5-öt, akik leülnek. Ők 5!-féleképpen ülhetnek le, hiszen az első székre 5, a másodikra 4, $\dots$, az ötödikre 1 törpe ülhet. Figyeljünk, most megkülönböztettük a székeket, nem csak a szomszédság számít! Tehát az összes esetek száma

$$\binom{7}{5} \cdot 5!.$$

Mindig meg kell gondolni, hogy a kedvező, vagy a kedvezőtlen eseteket könnyebb összeszámolni. Most talán a kedvezőtlent. Ekkor mindketten az 5 leültetett törpe között vannak, és a maradék 3 törpét Hófehérke $\binom{5}{3}$-féleképp választhatja ki. Az ültetésnél Morgó 5 helyre ülhet, Kuka pedig 2-re, hiszen Morgó mellé kell ültetni. A többi 3 törpét 3!-féleképp lehet leültetni. Összesen

$$\binom{5}{3} \cdot 5 \cdot 2 \cdot 3!$$

a rossz esetek száma. A keresett valószínűség

$$\begin{aligned}
&\mathbf{P}(\text{Morgó és Kuka nem ül egymás mellé})\\
&= 1 - \mathbf{P}(\text{Morgó és Kuka egymás mellé ül})\\
&= 1 - \frac{\binom{5}{3} \cdot 5 \cdot 2 \cdot 3!}{\binom{7}{5} \cdot 5!}\\
&= 1 - \frac{5}{21} = \frac{16}{21}.
\end{aligned}$$

□

1.3.2. Tíz pár cipőből véletlenül kiválasztunk négy darabot. Mekkora a valószínűsége, hogy nem lesz egy pár sem?

Megoldás. Tíz pár cipő az 20 db cipő, ezért az összes esetek száma

$$\binom{20}{4} = \frac{20 \cdot 19 \cdot 18 \cdot 17}{4!} = 4845.$$

Számoljuk a kedvező eseteket. Mivel nincs pár, összesen 4 pár cipőből választok ki egy-egy cipőt, ezt $\binom{10}{4}$-féleképpen tehetem meg, majd mindegyik párból a bal vagy jobb cipőt választhatom 2^4-féleképpen. Összesen

$$\binom{10}{4} \cdot 2^4 = 3360.$$

lehetőség. Tehát a keresett valószínűség

$$\mathbf{P}(\text{nincs pár}) = \frac{\binom{10}{4} \cdot 2^4}{\binom{20}{4}} = \frac{224}{323}.$$

□

1.3.3. Egy sakktáblán találomra elhelyezünk 8 bástyát. Mi a valószínűsége, hogy egyik sem üti a másikat?

1.3.4. Egy hallgató 40 tétel közül 20-at úgy megtanul, hogy abból jelesre tud vizsgázni, a másik 20-ból csak jóra. A vizsgatétel kiválasztásakor 40 tétel közül húz 2 tételt, majd ebből választ egyet és ebből felel. Mennyi a valószínűsége, hogy jelesre vizsgázik? ([2])

1.3.5. Egy dobókockával n-szer dobunk. Mi a valószínűsége, hogy a dobott számok összege osztható

(a) 6-tal?

(b) 5-tel? (IMC, 1999/2/2)

1.3.6. Egy urnában csak piros, zöld és kék golyók vannak. A piros golyók száma 18. Egy golyó kihúzása esetén annak a valószínűsége, hogy nem piros golyót húzunk 1/15-del kisebb, mint azé, hogy zöld vagy piros golyót húzunk. Annak a valószínűsége viszont, hogy kék vagy piros golyót húzunk 11/10-szer nagyobb, mint annak a valószínűsége, hogy zöld vagy piros golyót húzunk. Hány zöld és hány kék golyó van az urnában? (2008-as érettségi feladat)

1.3.7. Egy vendéglőben az egyik asztalnál 9 vendég ül. Négyen kólát, hárman sört rendeltek, ketten pedig ásványvizet rendeltek. A kissé feledékeny pincér emlékszik, hogy miből mennyit rendeltek, de azt már elfelejtette, hogy ki mit kért. Ezért véletlenszerűen osztja ki az italokat. Mekkora a valószínűsége, hogy mindenki azt kapja, amit rendelt?

Megoldás. A vendégeknek az italokat 9!-féleképpen oszthatja ki a pincér. Ekkor megkülönböztettük a kólákat! A kedvező leosztások száma

$$4! \cdot 3! \cdot 2,$$

hiszen a kólákat 4!-féleképpen oszthatja ki jó, a söröket 3!-féleképpen, a vizeket pedig 2-féleképpen. Tehát a keresett valószínűség

$$\mathbf{P}(\text{ mindenki azt kapja amit rendelt }) = \frac{4! \cdot 3! \cdot 2}{9!}.$$

2. megoldás. Nem kell megkülönböztetnünk az azonos italokat. Ekkor az összes esetek száma

$$\frac{9!}{4! \cdot 3! \cdot 2},$$

hiszen a kólákat 4!-féleképpen, a söröket 3!-féleképp, a vizeket 2-féleképp permutálhatjuk. Ha az összes eseteket így számoljuk, akkor kedvező eset csak 1 van, amikor mindenki azt kapja amit kért. Tehát a keresett valószínűség

$$\mathbf{P}(\text{ mindenki azt kapja amit rendelt }) = \frac{4! \cdot 3! \cdot 2}{9!},$$

ami persze ugyanaz, mint az előbb. Többféleképpen számolhatjuk az eseteket, de arra kell nagyon ügyelni, hogy ugyanúgy számoljuk az összes esetet, mint a kedvezőt. □

1.3.8. 2008-ban a Bajnokok Ligájában 4 angol (MU, Arsenal, Chelsea, Liverpool), egy olasz (Roma), egy spanyol (Barcelona), egy német (Schalke) és egy török (Fenerbahce) csapat jutott a 8 közé. Sorsolással határozzák meg a negyeddöntők párosítását. Határozzuk meg annak a valószínűségét, hogy a negyeddöntőben

(a) a Roma elkerüli a MU-t!

(b) az angol csapatok elkerülik egymást!

1.3.9. A Bajnokok Ligájában 2017-ben három spanyol csapat jutott a 8 közé: az Atlético Madrid, a Barcelona és a Real Madrid. Sorsolással határozták meg a negyeddöntők párosítását (itt már nincs kiemelés, és azonos nemzet csapatai is összekerülhetnek). Mennyi volt a sorsolás előtt a valószínűsége annak, hogy a negyeddöntőben

(a) Barcelona – Real Madrid párharc lesz?

(b) lesz spanyol párharc?

Megoldás. (a) Sokféleképpen számolhatjuk az eseteket. Arra kell mindig figyelni, hogy a kedvezőeket ugyanúgy számoljuk össze, mint az összeset. A feladat szempontjából lényegtelen, hogy ki kezd otthon, és az is, hogy hanyadiknak húzták ki az adott párt. Tehát az összes eset

$$\frac{\binom{8}{2} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{2}}{4!},$$

hiszen először kiveszem az első párt a 8 csapatból, aztán a maradék 6 csapatból a második párt, majd a 4-ből a harmadik párt, végül a negyedik párt. De azt mondtuk, hogy mindegy a párok sorrendje, ezért leosztom az egészet az összes lehetséges sorrenddel, ami 4!.

Ekkor a kedvező esetek összeszámolásánál rögzítem a Barcelona – Real Madrid párt. A maradék 6 csapatból választom a második párt, majd a maradék 4-ből a harmadikat, végül a negyediket. Most csak 3 párt választottam, ezért 3! a sorrendek száma. Összegezve a kedvező esetek száma:

$$\frac{\binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{2}}{3!}.$$

A kérdéses valószínűség

$$\mathbf{P}(\text{Barcelona} - \text{RM}) = \frac{\text{kedvező}}{\text{összes}} = \frac{4}{\binom{8}{2}} = \frac{1}{7}.$$

2. megoldás. Miután kijött az 1/7, gyanús, hogy egyszerűbben is lehetett volna:

– Andrés, most akkor a Madriddal játszunk?
– Nem, Leo. Hét csapat közül sorsolnak egyet nekünk. Bármelyiket egyforma eséllyel kapjuk, tehát 1/7 a valószínűsége, hogy a Madriddal játszunk.

(b) A változatosság kedvéért, most máshogy számoljuk az összes esetet. Mindent figyelünk, hogy kit hanyadiknak húztak, ki kezd otthon. Ekkor az összes esetek száma 8!, hiszen az elsőre kihúzott csapat 8-féle lehet, …. Ekkor a kedvező eseteket is eszerint kell számolni. Mivel 3 spanyol csapat van, ezért ha van spanyol párharc, az csak úgy lehet, hogy két spanyol csapat egymás ellen játszik, a harmadik meg mással. Ezt 3-féleképpen tehetjük meg, a két csapat 8-féleképpen játszhat egymás ellen (4 pár, ki kezd otthon). A maradék 6 csapat pedig 6!-féleképpen rendezhetem el. Összesen

$$3 \cdot 8 \cdot 6!.$$

Tehát a keresett valószínűség

$$\mathbf{P}(\text{spanyol párharc}) = \frac{3 \cdot 8 \cdot 6!}{8!} = \frac{3}{7}.$$

□

1.3.10. Máté nagymamája meggylevest készít a vasárnapi ebédhez. Összesen 5 szem szegfűszeget tesz a levesbe. A levest 4 egyforma adagra osztják. Mennyi a valószínűsége, hogy Máté levesében nem lesz szegfűszeg? Mennyi a valószínűsége, hogy legalább 3 szegfűszeg lesz a levesében?

1.3.11. Egy unatkozó gyakorlatvezető dolgozatiratás során arra lett figyelmes, hogy a csoportjában az összes lány egy sorban ül. A csoportban 10 hallgató van, közülük 3 lány. A teremben 4 sor van és minden sorban 4 hely, és feltesszük, hogy mindenki véletlenszerűen választ helyet, azaz minden leülési konfiguráció egyforma valószínűségű. Mennyi a kérdéses esemény valószínűsége?

1.3.12. Egy halastóban M aranyhal és K ezüsthal van. Egy horgász addig fogja ki egyesével a halakat, amíg már csak egyszínű hal marad a tóban (tehát vagy csupa aranyhal, vagy csupa ezüsthal). Mennyi a valószínűsége, hogy a Gyuri nevű ezüsthal megússza a horgászkalandot?

Megoldás. Ez egy kicsit nehezebb feladat. Könnyen belefuthatunk egy olyan összeszámlálási feladatba, amit nehéz megoldani.

Képzeljük el, hogy a halakat véletlenszerűen megszámozzuk 1-től $(M + K)$-ig, és a sorszám azt jelöli, hogy a horgász hanyadiknak horgászná ki az adott halat. Nem húzza ki mindet, az utolsó néhány sorszámú hal megússza, de az a véletlentől függ, hogy hányan. Azt kell észrevenni, hogy Gyuri pontosan akkor ússza meg, ha az ő sorszáma nagyobb, mint az utolsó aranyhal sorszáma. Tehát csak arra kell figyelni, hogy az M aranyhalhoz képest Gyuri sorszáma hol van. Tehát Gyuri pontosan akkor ússza meg, ha $M+1$ hal közül ő az utolsó, aminek a valószínűsége

$$\frac{1}{M+1}.$$

□

1.3.13. Egy n házaspárból álló társaság leül egy kör alakú asztalhoz úgy, hogy férfiak és nők felváltva ülnek, egyébként a sorrend véletlen. Mi a valószínűsége, hogy egyetlen feleség sem kerül a férje mellé?

1.3.14. Egy vacsorára n feleség érkezik a férjével és egy barátnőjével. A $3n$ embert egy kör alakú asztalhoz kell leültetni. Minden feleség a barátnője mellett akar ülni. Ha ezen feltételt betartva, de egyébként véletlenszerűen ültetnek, jelölje p_n annak a valószínűségét, hogy semelyik férj nem ül a felesége mellett. Határozzuk meg p_n-et!

1.3.15. Az A, B, C, D, E, F kereskedőcégek mindegyike az öt másik céggel kötött egy-egy üzletet az előző hónapban (bármelyik két cég között pontosan egy üzletkötés jött létre). Az ellenőrző hatóság véletlenszerűen kiválaszt a hat cég előző havi (egymás közötti) üzletkötései közül négyet, és azokat ellenőrzi. Mekkora a valószínűsége, hogy az A vagy a B cég üzletkötései közül is ellenőriznek legalább egyet? (Érettségi 2018)

1.3.16. Egy szabályos kockával 11-szer dobunk. Mennyi a valószínűsége, hogy az egymást követő 1, 2, 3, 4, 5, 6 eredménysorozat nem fordul elő?

1.3.17. Száz alma közül tíz férges. Véletlenül kiválasztva ötöt, mi a valószínűsége, hogy lesz közte férges?

1.3.18. Mekkora a valószínűsége, hogy az ötöslottón kihúzott számok között nem lesznek egymást követők?

Megoldás. Ez egy kicsit nehezebb. Mivel 90 szám közül választunk ki ötöt, így az összes esetek száma

$$\binom{90}{5}.$$

A kedvező eseteket trükkös összeszámolni. A keresett esemény

$$A = \{(i_1, i_2, \ldots, i_5) : 1 \le i_1 < i_2 - 1 < i_3 - 2 < i_4 - 3 < i_5 - 4 \le 86\},$$

hiszen az, hogy nincsenek szomszédosak, pontosan azt jelenti, hogy $i_2 > i_1+1$, $i_3 > i_2 + 1$, …, $i_5 > i_4 + 1$, ami átrendezve ugyanaz ami fönt van. Ekkor majdnem kész is vagyunk, hiszen az $(i_1, i_2-1, \ldots, i_5-4)$ az $1, \ldots, 86$ számok közül kiválasztott számötös, azaz az ilyenek száma

$$\binom{86}{5}.$$

Valójában megadtunk egy bijekciót (egy-egy értelmű leképezést) a nemszomszédos számötösök halmaza és az $1, \ldots, 86$ halmaz számötösei között. Tehát a keresett valószínűség

$$\mathbf{P}(\text{nincs szomszédos}) = \frac{\binom{86}{5}}{\binom{90}{5}}.$$

□

1.3.19. Egy embernek n egyforma kinézetű kulcsa van, melyek közül pontosan egy nyitja az ajtót. Emberünk véletlenül választva sorra próbálja a kulcsokat addig, amíg a jó kulcs elő nem kerül. Valamely $k \in \{1, 2, \ldots, n\}$ esetén mennyi a valószínűsége, hogy a k-adik próbálkozása sikeres, ha

(a) a kipróbált rossz kulcsokat mindig félreteszi?

(b) a kipróbált rossz kulcsokat sose teszi félre?

1.3.20. Egy $2n$ lányból és $2n$ fiúból álló társaságot véletlenszerűen két egyenlő létszámú csoportra osztunk. Legyen p_n annak a valószínűsége, hogy a csoportokon belül is megegyezik a lányok és a fiúk száma.

(a) Határozzuk meg p_n-et!

(b) Számítsuk ki a $\lim_{n\to\infty} p_n\sqrt{n}$ határértéket!

1.3.21. Egy n elemű halmaz összes részhalmaza közül kiválasztunk egyet, majd újra az összes közül kiválasztunk még egyet. Mennyi a valószínűsége, hogy az egyik halmaz tartalmazza a másikat? ([5, 3.11])

1.3.22. Bergengóciában minden ember egymástól függetlenül $p \in (0, 1)$ valószínűséggel hazudik, csak két kivétel van: az Elnök és a Rádióriporter, akik megbízhatók. Az Elnök elhatározza, hogy ismét indul a választáson, és ezt közli egy emberrel, aki továbbadja a hírt, …, végül az n-edik ember elmondja a Rádióriporternek. (Az n ember közt nem szerepel sem a Rádióriporter, sem az Elnök.) Mikor valószínűbb, hogy a Rádióriporter a valódi döntést közvetíti, $n = 19$ vagy $n = 20$ esetén? ([5, 3.40])

1.3.23. Egy $n \times n$-es négyzetrács bal alsó és jobb felső sarkába leteszünk egy-egy pókot. A pókok $2n$ lépésben helyet cserélnek úgy, hogy egymástól függetlenül minden útvonalat ugyanakkora valószínűséggel választanak. A pókok egyszerre indulnak, és minden másodpercben egy lépést tesznek. Mi a valószínűsége, hogy a két pók találkozik?

1.3.24. Mind a 4 pár különböző cipőmet az előszobában levő beépített szekrényben tartom. Az előszobában kiégett a villany, így sötétben keresgélve véletlenszerűen kiveszek 4 cipőt. Mennyi a valószínűsége, hogy pontosan egy összeillő párt veszek ki? Mekkora ugyanez a valószínűség, ha 4 pár egyforma cipőm van?

1.3.25. Az iskolai karácsonyi vásárra készülődve Blanka, Csenge és Dóri feladata az volt, hogy különböző figurákat hajtogassanak színes papírból. Összesen 70 figurát hajtogattak. A figurák kétheted részét Dóri készítette, a maradékot pedig fele-fele arányban Blanka és Csenge. Számítsa ki annak a valószínűségét, hogy a 70 figura közül véletlenszerűen kiválasztott két figurát ugyanaz a lány készítette! (Érettségi 2018)

1.3.26. Egy kertész három juhar-, négy tölgy- és öt nyírfát ültet egy sorba véletlen sorrendben, mindegyik fát egyenlő valószínűséggel választva. Mennyi annak a valószínűsége, hogy nem kerül egymás mellé két nyír?

1.3.27. Egyes vidékeken elterjedt a következő babona: egy lány 6 fűszálat fog a markába úgy, hogy azok a kezéből mindkét irányban kiállnak. Egy másik lány mindkét oldalon páronként összecsomózza a fűszálakat. Ha így egy zárt lánc keletkezik, akkor arra következtetnek, hogy a lány a következő évben férjhez megy. Ha a csomózás teljesen véletlenszerűen történik, mennyi a valószínűsége, hogy zárt láncot kapunk. Mi a helyzet $2n$ fűszál esetén? ([2, 1.3.29])

1.3.28. Mennyi a valószínűsége, hogy a kör kerületén adott $2n$ pontot találomra párokba szedve és mindegyik pár két pontját húrral összekötve a kapott n számú húr közül egyik sem metszi a másikat? (Schweitzer, 1952, 6)

A következő feladatok a Boltzmann–Maxwell-, Bose–Einstein- és a Fermi–Dirac-statisztikákat tárgyalják. Bővebben Feller [4].

1.3.29. A Boltzmann–Maxwell-statisztikánál r golyót úgy helyezünk el n dobozba, hogy mind az n^r elhelyezés egyformán valószínű. Határozzuk meg annak a p_k valószínűségét, hogy pontosan k golyó kerül az első dobozba! Számítsuk ki ezt a határértéket, ha $n \to \infty$ úgy, hogy $r/n \to \lambda$!

1.3.30. A Bose–Einstein-statisztika esetén a feltétel az, hogy a lehetséges $\binom{n+r-1}{r}$ számú elhelyezés egyformán valószínű. (Ez a modell jól használható a nukleonok és páros számú elemi részecskét tartalmazó atomok esetében.) Határozzuk meg annak a q_k valószínűségét, hogy pontosan k golyó kerül az első dobozba! Számítsuk ki ezt a határértéket, ha $n \to \infty$ úgy, hogy $r/n \to \lambda$!

1.3.31. A Fermi–Dirac-statisztika esetén egy dobozba legfeljebb egy golyó kerülhet, és az összes ilyen $\binom{n}{r}$ elhelyezés egyformán valószínű. (Ez a modell alkalmazható elektronokra, neutronokra és protonokra.) Határozzuk meg annak az r_k valószínűségét, hogy pontosan k golyó kerül az első dobozba, $k = 0, 1$, és számítsuk ki ezt a határértéket, ha $n \to \infty$ úgy, hogy $r/n \to \lambda$!

1.4. Geometria valószínűség

1.4.1. A $[0, 1]$ intervallumot felosztjuk két véletlenül rádobott ponttal három részre. Mennyi annak a valószínűsége, hogy

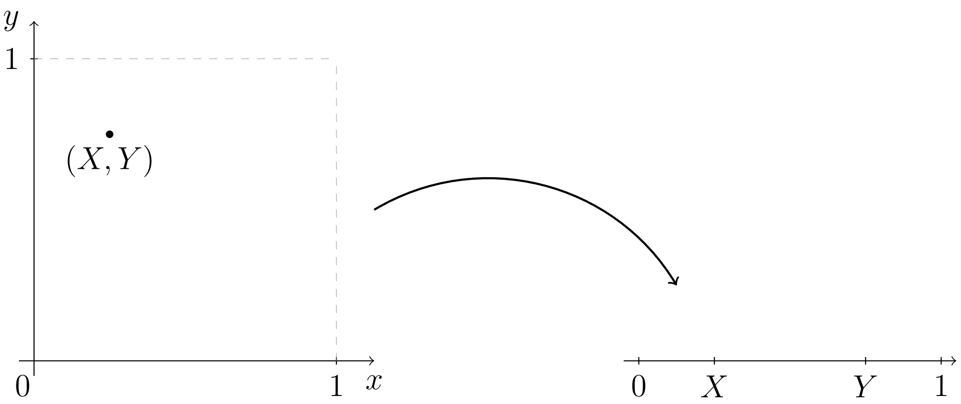
(a) mindhárom szakasz hossza nagyobb, mint $1/4$?

(b) mindhárom szakasz hossza kisebb mint $1/2$?

(c) a szakaszokból háromszög szerkeszthető?

(d) a szakaszokból hegyesszögű háromszög szerkeszthető?

Megoldás. Csak az (a) részt csináljuk meg. Két pontot választunk a $[0, 1]$ intervallumon, jelölje X az elsőnek, Y a másodiknak választott pontot. Nagyon fontos, hogy nem a kisebb és nagyobb, azaz lehet, hogy X a nagyobb, lehet, hogy Y. Ekkor (X, Y) egy pont a $[0, 1]^2$ egységnégyzetben, tehát a kísérlet kimenetelei megfeleltethetők az egységnégyzet pontjainak. Azaz egy geometriai valószínűségi mezőnk van, $\Omega = [0, 1]^2$, $\mathcal{A} = \mathcal{B}([0, 1]^2)$ (az egységnégyzet Borel-halmazai, vagy a szép halmazok, kinek mi), és mivel az egységnégyzet területe 1, ezért a valószínűség éppen a terület, azaz $\mathbf{P}(A) = \mathrm{ter}(A)$.



Azt kell meghatározni, hogy mi lesz az az esemény, hogy mindhárom szakasz nagyobb, mint 1/4. Az X, Y helyzetétől függően a három szakasz:

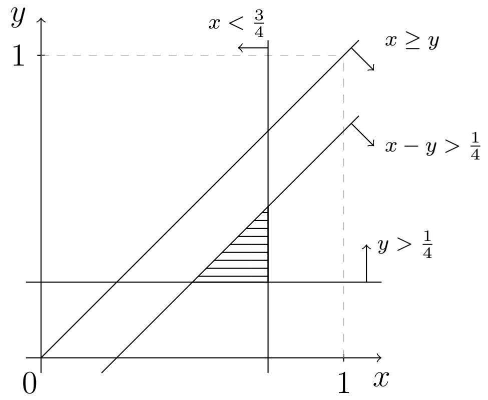
(i) ha $X \geq Y$: $[0, Y]$, $[Y, X]$, és $[X, 1]$.

(ii) ha $X \leq Y$: $[0, X]$, $[X, Y]$, és $[Y, 1]$.

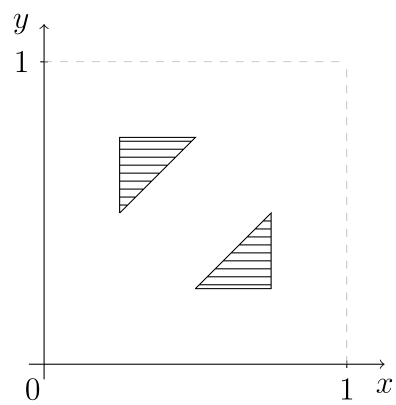
Nézzük az (i) esetet. Ekkor a jobb alsó háromszögben vagyunk. A szakaszok hosszai: $Y, X-Y, 1-X$. Tehát az kell, hogy

$$Y > \frac{1}{4}, \quad X - Y > \frac{1}{4}, \quad 1 - X > \frac{1}{4}.$$

Tehát az egységnégyzet azon részét kell meghatározni, ahol mindhárom egyenlőtlenség teljesül. Na de ez könnyű:



A (ii) eset hasonlóan tárgyalható (szimmetria okokból is következik). Azt kapjuk, hogy a kedvező síkrész két 1/4 befogójú egyenlő szárú derékszögű háromszög:



Azaz a kedvező terület 1/16 és a keresett valószínűség

$$\mathbf{P}(\text{ mindhárom szakasz 1/4-nél hosszabb }) = \frac{1}{16}.$$

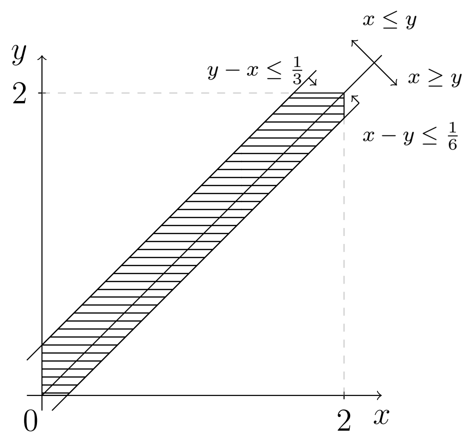
$\square$

1.4.2. András és Betti munkaideje egymástól függetlenül egy-egy du. 4 és 6 közötti egyenletes eloszlású időpontban ér véget. Munkaidejük végeztével mindketten elmennek egy, munkahelyüktől azonos távolságra levő kávézóba, ahol elfogyasztanak egy csésze kávét. András esetében ez 10 perc, Betti esetében 20. Mi a valószínűsége, hogy találkoznak?

Megoldás. A kísérlet két véletlen pont választásával írható le. Legyen $4+X$, ill. $4+Y$ az az időpont órában, amikor András, ill. Betti végez. Ekkor a kísérlet megfeleltethető egy véletlen pont választásának a $[0,2]^2$ négyzetben. Tehát $\Omega = [0,2]^2$, $\mathcal{A} = \mathcal{B}([0,2]^2)$, és $\mathbf{P} : \mathcal{A} \to [0,1]$, a normalizált terület, azaz

$$\mathbf{P}(A) = \frac{\text{ter}(A)}{4}.$$

Azt az eseményt kell leírnunk, hogy András és Betti találkoznak. Ha András végez hamarabb, azaz $X \le Y$, akkor mivel ő csak 10 percet (1/6 órát) kávézik, pontosan akkor találkoznak, ha $Y \le X + 1/6$. Ha viszont Betti véges hamarabb, azaz $Y \le X$, akkor pontosan akkor találkoznak, ha $X \le Y + 1/3$. Tehát ezt a síkrészt kell meghatározni, majd ennek a területét kiszámolni. A fentiek szerint a keresett síkrész:



A kedvező terület

$$4 - \frac{1}{2}\left(2 - \frac{1}{6}\right)^2 - \frac{1}{2}\left(2 - \frac{1}{3}\right)^2 = \frac{67}{72},$$

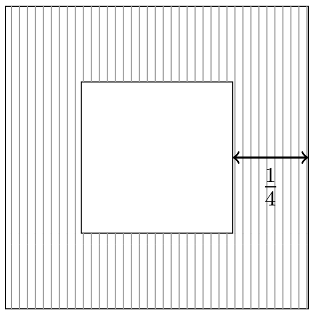
így a keresett valószínűség

$$\mathbf{P}(\text{találkoznak}) = \frac{67}{288}.$$

$\square$

1.4.3. Egy négyzet belsejében egyenletes eloszlás szerint választunk egy pontot. Mennyi a valószínűsége, hogy a választott pont közelebb van valamelyik oldalhoz, mint $1/4$?

Megoldás. A kísérlet egy geometriai valószínűségi mezőn írható le, ahol az eseménytér az egységnégyzet, az események az egységnégyzet Borel-halmazai (szép halmazai), és valószínűség pedig a terület, azaz $\Omega = [0,1]^2$, $\mathcal{A} = \mathcal{B}([0,1]^2)$, $\mathbf{P}(A) = \text{ter}(A)$. A kedvező síkrész:



A kedvező terület, ami éppen a keresett valószínűség

$$\mathbf{P}(\text{közelebb van mint } 1/4) = 4 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{4}.$$

□

1.4.4. Választunk egy véletlen számot 0 és 2 között, és egy másikat ettől függetlenül 1 és 2 között. Mennyi a valószínűsége, hogy az összegük kisebb, mint 2?

1.4.5. Válasszuk az X, Y pontokat egymástól függetlenül a $(0,1)$ intervallumban egyenletes eloszlás szerint. Mennyi a valószínűsége, hogy az

$$x^2 + Xx + Y = 0$$

egyenletnek valós gyökei lesznek?

1.4.6. Mátyás vonattal utazik Szegedről Debrecenbe, Cegléden kell átszállnia. A 10:45-kor induló vonat 12:12 és 12:22 között érkezik Ceglédre egyenletes eloszlás szerint, a Budapestről Debrecenbe induló vonat pedig 12:14 és 12:19 között egyenletes eloszlás szerint fut be, és vár Cegléden 2 percet. Mennyi a valószínűsége, hogy Mátyás eléri a csatlakozást? (A vonatok szomszédos vágányra érkeznek, és Mátyás nagyon gyors.)

1.4.7. Anna és Szabina minden szerdán fodrászhoz mennek. Anna 2 és 3 óra között, Szabina pedig 2 és fél 3 között végez egy véletlenszerű időpontban, egymástól függetlenül. Egymást megvárják, majd együtt indulnak haza. Mennyi annak a valószínűsége, hogy egy adott napon negyed 3 után indulnak haza? Mennyi ez a valószínűség, ha tudjuk, hogy Anna legalább 10 perccel korábban végzett, mint Szabina?

1.4.8. Tekintsünk egy egységnyi kerületű kört, és ennek egy rögzített pontját. Válasszunk további két pontot a körvonalon egymástól függetlenül egyenletes eloszlás szerint. Mennyi a valószínűsége, hogy a három pont által meghatározott háromszög fedi a kör középpontját?

1.4.9. Egy egységnégyzet két szemközti oldalán véletlenül választunk egy-egy pontot. Mi a valószínűsége, hogy távolságuk négyzete kisebb, mint $3/2$?

1.4.10. Egy egységnyi hosszú szakaszon találomra választunk kettő pontot. Mennyi a valószínűsége, hogy mindkét pont a szakasz egy előre adott végpontjához van közelebb, ha tudjuk, hogy a két választott pont távolsága kisebb, mint $1/3$?

1.4.11. A [0,1] intervallumban egyenletes eloszlás szerint véletlenszerűen egymástól függetlenül választunk két számot. Mennyi a valószínűsége, hogy legalább az egyik szám kisebb, mint 0,5, és ugyanakkor a távolságuk nagyobb, mint 0,25?

1.4.12. Egy kör kerületén egymástól függetlenül, egyenletesen választunk 4 pontot: A, B, C, D. Mennyi a valószínűsége, hogy az AB és CD húrok metszik egymást? (KöMaL 2012)

1.4.13. Egy kör kerületén válasszunk n pontot egymástól függetlenül, egyenletes eloszlás szerint. Mennyi annak a valószínűsége, hogy a pontok konvex burka tartalmazza a kör középpontját? Mennyi ez a valószínűség, ha a pontokat a kör belsejében választjuk függetlenül, egyenletes eloszlás szerint?

1.5. Vegyes

1.5.1. A vezeték nélküli hálózati kapcsolatot létrehozó egységek (wifi routerek) 3%-a 2 éven belül meghibásodik (ezt úgy tekinthetjük, hogy 0,03 annak a valószínűsége, hogy egy készülék meghibásodik 2 év alatt). A meghibásodott eszközt garanciálisan kicserélik. Az iskola 20 ilyen eszközt vásárolt. Mennyi a valószínűsége annak, hogy 2 év alatt legfeljebb egy hibásodik meg a vásárolt eszközök közül? (Érettségi 2018)

1.5.2. Egy szabályos érmét addig dobunk, amíg kétszer egymás után azonos oldalára nem esik. Adjuk meg a kísérletet leíró valószínűségi mezőt! Határozzuk meg a következő események valószínűségét:

(a) { legfeljebb 6-ot kell dobni };

(b) { páros sokat kell dobni }.

1.5.3. Egy csomag francia kártyát megkevertünk, majd egyesével kihúzzuk a lapokat. Hanyadik helyen legvalószínűbb a második ász?

1.5.4. Egy urnában 101 golyó van, közülük pontosan három piros. A golyókat visszatevés nélkül egyesével kihúzzuk. Hányadik helyen legvalószínűbb a második piros?

1.5.5. Száz gyerek közt egy nem szabályos érme többszöri feldobásával szeretnénk kisorsolni egy ajándékot. Az összes k hosszú dobássorozathoz meghatározzuk, hogy ki nyer, majd feldobjuk k-szor az érmét. Bizonyítsuk be, hogy a fej p valószínűségét és a k értékét alkalmasan megválasztva a 2^k kimenetelt fel lehet osztani a gyerekek közt úgy, hogy mindenki azonos valószínűséggel nyerjen. ([5, 3.41])

1.5.6. Veszünk egy elég nagy urnát, és éjfél előtt fél perccel 1-től 10-ig számozott golyókat rakunk bele, majd rögtön kiveszünk egyet. Éjfél előtt $1/4$ perccel az urnába 11-től 20-ig számozott golyókat teszünk, majd rögtön kiveszünk egyet. Ezt így folytatjuk éjfélig. Hány golyó lesz az urnában pontban éjfélkor, ha

(a) az i-edik lépésben az i-edik golyót vesszük ki?

(b) az i-edik lépésben a $10 \cdot i$-edik golyót vesszük ki?

(c) véletlenül vesszük ki a golyót?

1.5.7. Vegyünk két olyan kockát, melyeken annak a valószínűsége, hogy i-t dobunk, p_i, ill. q_i, $i = 1, 2, \ldots, 6$, ahol $p_i, q_i \geq 0$ és $\sum_{i=1}^{6} p_i = 1$ és $\sum_{i=1}^{6} q_i = 1$. Választhatók-e a p_i, q_i valószínűségek úgy, hogy két kockával dobva minden összeg 2-től 12-ig egyformán valószínű legyen?

Bolyongás. Olyan sorozatokat fogunk vizsgálni, melyek véges sok plusz egyből és mínusz egyből állnak. Tekintsünk az $\varepsilon_1, \ldots, \varepsilon_n$ sorozatot, melyben p db 1 és q db -1 szerepel, $p + q = n$. Jelölje $s_k = \varepsilon_1 + \cdots + \varepsilon_k$ ezek részletösszegét. Az $(\varepsilon_1, \ldots, \varepsilon_n)$ sorozatot egy törtvonallal ábrázoljuk: a törtvonal k-adik lépésének meredeksége ε_k és k-adik szögpontjának ordinátája s_k. Jelölje $N_{n,x}$ az origóból az (n, x) pontba vezető utak számát.

1.5.8. Határozzuk meg $N_{n,x}$ értékét!

1.5.9. Tükrözési elv. Legyen $A = (a, \alpha)$, $B = (b, \beta)$, és $A' = (a, -\alpha)$ az A pont x-tengelyre vonatkozó tükörképe. Mutassuk meg, hogy az x-tengelyt érintő vagy átmetsző $A \to B$ utak száma megegyezik az $A' \to B$ utak számával.

1.5.10. Ballot–tétel. Legyenek n és x pozitív egészek. Igazoljuk, hogy az origóból az (n, x) pontba vezető olyan $(s_1, s_2, \dots, s_n = x)$ utak száma, melyre $s_1 > 0, s_2 > 0, \dots, s_n > 0$ pontosan $\frac{x}{n} N_{n,x}$.

1.5.11. Ballot–tétel'. Tegyük fel, hogy egy választás során a P jelölt p számú, a Q jelölt q számú szavazatot kap, ahol $p > q$. Annak a valószínűsége, hogy a szavazatszámlálás során P végig vezetett $(p - q)/(p + q)$.

A továbbiakban a fent leírt törtvonalra úgy gondolunk, mint egy bolyongó részecske pályájára. Jelölje $X_1, X_2, \dots$ az egyes lépéseket és $S_k = X_1 + \dots + X_k$. Feltesszük, hogy a részecske n lépése során előforduló 2^n útvonal egyformán valószínű.

1.5.12. Mennyi $\mathbf{P}(S_n = r)$ valószínűség? Legyen $u_{2k} = \mathbf{P}(S_{2k} = 0)$ és $f_{2k} = \mathbf{P}(S_2 \neq 0, \dots, S_{2k-2} \neq 0, S_{2k} = 0)$. Határozzuk meg u_n-et, és igazoljuk, hogy

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0.$$

1.5.13. Annak a valószínűsége, hogy az origóba való visszatérés nem következik be a $(2n)$-edik időpillanatig az ugyanannyi, mint annak a valószínűsége, hogy a $(2n)$-edik időpillanatban a részecske visszatért az origóba.

1.5.14. Bizonyítsuk be, hogy

$$\Gamma(p+1) \sim \left(\frac{p}{e}\right)^p \sqrt{2\pi p}.$$

Ennek segítségével igazoljuk, hogy $\Gamma(p+h)/\Gamma(p) \sim p^h$, amint $p \to \infty$.

1.5.15. Mutassuk meg, hogy

$$\int_0^{\frac{\pi}{2}} \sin^n x \, \mathrm{d}x = \int_0^{\frac{\pi}{2}} \cos^n x \, \mathrm{d}x = \begin{cases} \frac{(n-1)!!}{n!!} \frac{\pi}{2}, & \text{ha } n \text{ páros}, \\ \frac{(n-1)!!}{n!!}, & \text{ha } n \text{ páratlan}. \end{cases}$$

Ez alapján igazoljuk a **Wallis–formulát**: $\lim_{n\to\infty} \left(\frac{(2n)!!}{(2n-1)!!}\right)^2 \frac{1}{2n} = \frac{\pi}{2}$.

1.5.16. De Moivre is tudta, hogy $n! \sim (n/e)^n \sqrt{n} c$, ahol c valamilyen pozitív konstans. A Wallis–formula segítségével igazoljuk, hogy $c = \sqrt{2\pi}$! (Stirling is így csinálta 1733-ban.)

1.5.17. A béta–függvény (vagy elsőfajú EULER–féle integrál).

(a) Bizonyítsuk be, hogy

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$$

véges minden $x > 0$, $y > 0$ esetén, és $B(x, y) = B(y, x)$.

(b) Mutassuk meg, hogy

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

(Vagyis, ha a Gamma–függvényt a faktoriális folytonos kiterjesztésének tekintjük, akkor a Béta–függvény a binomiális együttható folytonos kiterjesztése.)

1.6. Feltételes valószínűség, Bayes tétele

1.6.1. Az 52 lapos francia kártyából kiosztanak 13 lapot. Legyen A az az esemény, hogy pontosan 2 ászt kaptunk. Határozzuk meg a $\mathbf{P}(A|B_i)$ feltételes valószínűségeket, ha

(a) B_1 azt jelenti, hogy van legalább egy ászunk;

(b) B_2 azt jelenti, hogy a kőr ász nálunk van;

(c) B_3 azt jelenti, hogy a kiosztott lapok közül az első ász;

(d) B_4 azt jelenti, hogy a kiosztott lapok közül az első a kőr ász.

1.6.2. Magyar kártyából kapunk 12 lapot, a másik két játékos 10–10 lapot kap. Mennyi a valószínűsége, hogy pontosan 5 pirosat kapunk, közte a piros hetest? Feltéve, hogy pontosan 5 pirosat kaptunk közte a piros hetest, mennyi a valószínűsége, hogy a másik 3 piros egy kézben van?

1.6.3. Elhelyezünk N golyót n dobozba úgy, hogy az összes n^N elhelyezés egyformán valószínű. Feltéve, hogy az első dobozban van golyó, mennyi a valószínűsége, hogy K golyó van benne?

1.6.4. Egy kockával addig dobunk, míg először kapunk hatost. Feltéve, hogy a szükséges dobások száma páros, mennyi a valószínűsége, hogy kétszer kellett dobnunk?

1.6.5. Tegyük fel, hogy egy alkatrész meghibásodásának valószínűsége a $(t, t+h)$ intervallumban, feltéve, hogy t ideig működött, $a(t)h + o(h)$. Határozzuk meg annak a $p(t)$ valószínűségét, hogy az alkatrész legalább t ideig működött! ([2] 1.5.13)

1.6.6. *Doppingteszt.* Kifejlesztenek egy új doppingtesztet, mely a doppingolók 99%-ánál pozitív eredményt ad, azonban a nem doppingoló sportolók 1%-nál is tévesen pozitív eredményt ad. Tegyük föl, hogy a sportolók 1%-a doppingol. Mennyi annak a valószínűsége, hogy egy véletlenül kiválasztott sportoló

(a) doppingtesztje pozitív?

(b) doppingolt, ha tudjuk, hogy a doppingtesztje pozitív?

1.6.7. Két pénzérme közül az egyik szabályos, a másik cinkelt, 1/4 valószínűséggel ad fejet. Véletlenszerűen kiválasztjuk az egyiket, majd ezzel kétszer dobunk. Mennyi a valószínűsége, hogy két fejet kapunk? Ha két fejet kapunk, mennyi a valószínűsége, hogy a szabályos érmét választottuk?

1.6.8. Egy hallgató p valószínűséggel tudja a választ egy kérdésre. Ha nem tudja, akkor az n lehetséges válasz közül véletlenül választ egyet. Mennyi legyen a lehetséges válaszok n száma, hogy az oktató legalább $0,9$ valószínűséggel következtethessen arra a hallgató jó válaszából, hogy a hallgató tudta a választ?

1.6.9. Aladár a pénzét három egyforma borítékban tartja. Az elsőben két ezerforintos, a másodikban egy ezer- és egy kétezerforintos, a harmadikban egy ezer és három kétezerforintos van. Aladár találomra kivesz egy borítékot, és onnan egy bankjegyet. Mennyi a valószínűsége, hogy ezerforintost húzott?

1.6.10. Feldobunk n egyforma dobókockát, melyeken a hatos valószínűsége $p \in (0,1)$. Jelölje p_n annak a valószínűségét, hogy páratlan sok hatost dobtunk. Határozzuk meg p_n-et!

1.6.11. Egy cukrászdában 3 cukrász A, B és C süt süteményt, és a sütemények 2, 3 illetve 5%-át rontják el. A sütemények 50%-át A, 30 %-át B, 20%-át pedig C készíti. Mennyi a valószínűsége, hogy A sütötte a süteményt, feltéve, hogy az rossz?

1.6.12. Az iskolai karácsonyi vásárra készülődve Blanka,Csenge és Dóri feladata az volt, hogy különböző figurákat hajtogassanak színes papírból. Összesen 70 figurát hajtogattak. A figurák kétheted részét Dóri készítette, a maradékot pedig fele-fele arányban Blanka és Csenge. A Blanka által készített figurák 40%-a volt karácsonyfa, a Csenge által készített figuráknak 60%-a, a Dóri által készített figuráknak pedig 30%-a. Az első vásárló a vásáron Blanka édesanyja volt; ő megvett egy véletlenszerűen kiválasztott karácsonyfafigurát. Határozza meg annak a valószínűségét, hogy a figurát éppen Blanka készítette! (Érettségi 2018)

1.6.13. Aladár hétfő reggelenként 7:15-kor indul el otthonról, hogy 8:00-ra beérjen az egyetemre. Gyalog kimegy a buszmegállóba, 20 percet buszozik, aztán villamosra száll át, amelyen 15 percet utazik, végül ismét gyalogol az egyetemig. A buszon 0,5, a villamoson pedig 0,2 valószínűséggel hallja meg, ha csörög a mobiltelefonja, míg gyaloglás közben biztosan észreveszi, ha hívják. Ha hétfő reggel 7:15 és 8:00 között egy véletlen időpontban felhívjuk, és felveszi a telefonját, akkor mennyi a valószínűsége, hogy épp villamoson van?

1.6.14. Jókedvében Mátyás király kegyelmet ajánl velencei rabjának, ha a rab két egyforma urna közül az egyikből kihúz egy ezüstgolyót. Megengedi neki, hogy 50 ezüst- és 50 aranygolyót úgy osszon el a két urnába, ahogy akarja. Ezután Mátyás udvari bolondja találomra választ egy urnát, a rab pedig abból találomra egy golyót. Hogyan ossza el a rab a két urnába a golyókat, ha kedves az élete? Ekkor mekkora az esélye a szabadulásra?

1.6.15. Vándorlásai közben Odüsszeusz egyszer egy hármas útelágazáshoz ért. Tudta, hogy az egyik út Athénba, a másik Mükénébe, a harmadik pedig Spártába vezet. Azt is tudta, hogy az athéniak átlagosan minden harmadik alkalommal mondanak igazat, a mükénéiek minden második alkalommal, a spártaiak pedig becsületesek, sosem hazudnak. Kockadobással döntötte el, melyik utat válassza, egyforma esélyt adva mindegyiknek. Ezután ment, mendegélt, míg egy városba nem ért. Itt az első szembejövőtől megkérdezte, hogy mennyi kettő meg kettő, és azt a választ kapta, hogy négy. Mennyi a valószínűsége, hogy Athénba érkezett?

1.6.16. A szegedi Gabonakutatóban két új kukorica vetőmagot fejlesztettek ki. Az A típusú strapabíró vetőmag csapadékos, átlagos, és száraz időjárás esetén egyaránt 0,2 valószínűséggel hoz prémium termést. A B típusú vetőmag csapadékos időjárás estén 0,5, átlagos időjárás esetén 0,1 valószínűséggel hoz prémium termést, száraz időben viszont biztosan nem hoz jó termést. A meteorológusok 0,5 valószínűséggel átlagos, 0,2 valószínűséggel sok, 0,3 valószínűséggel pedig kevés csapadékot jósolnak. Melyik vetőmagot válasszuk, ha minél nagyobb valószínűséggel akarunk prémium termést?

1.6.17. Máté nagymamája meggylevest készít a vasárnapi ebédhez. Összesen 5 szem szegfűszeget tesz a levesbe. A levest 4 egyforma adagra osztják. Máté már az első kanál levesében talál egy szegfűszeget. Mennyi a valószínűsége, hogy legalább 3 szegfűszeg lesz a levesében, feltéve, hogy van a levesében szegfűszeg?

1.6.18. Egy könyvelő irodában három alkalmazott van, akik egyenlő arányban osztoznak a munkán. Felvesznek egy negyedik embert is az irodára, hogy csökkentsék a terhelést, így már négyfelé oszlik a munka egyenlő arányban. A tapasztalt könyvelők ritkán, az eseteknek mindössze 3%-ában hibáznak, míg az újonc, aki csak most tanult bele a könyvelésbe 10%-os hibarátával bír. A könyvelők főnöke kézbevesz egy adóbevallást és azt látja, hogy az hibásan van kitöltve, mekkora a valószínűsége, hogy az új alkalmazottat terheli a felelősség?

1.6.19. Egy matematikuscsaládban a fiú megfigyelte, hogy mikor hazaér, az esetek 20%-ában senki nincs otthon, 50%-ában csak az édesanyja, a maradék esetekben pedig mindkét szülője. Tudja, hogy édesanyja az esetek 80%-ában magára zárja az ajtót; ha mindketten otthon vannak, akkor már csak 40%-ban zárják be, de ha nincs otthon senki, szórakozottságból akkor is az esetek 5%-ában nyitva marad az ajtó. A fiú egy délután hazaér, és zárva találja az ajtót. Mekkora a valószínűsége, hogy van otthon valaki?

1.6.20. A sztochasztika tanszék egyik oktatója p valószínűséggel szokott bejönni a tanszékre. Ha ismerőseinek azt mondta, hogy aznap bejön, akkor annak a valószínűsége, hogy pontosan k-an keresik telefonon $e^{-\mu}\mu^k/k!$, ha pedig azt mondta, hogy nem, akkor $e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$, $0 < \lambda < \mu$. Feltéve, hogy k hívás érkezett, mennyi a valószínűsége, hogy aznap bent volt az oktató? Vizsgáljuk a $k \to \infty$ esetet. ([2])

1.6.21. Legyen $(1-p)p^n$ annak a valószínűsége, hogy egy almafán n virág van, $n = 0, 1, \ldots$. Tegyük fel, hogy minden virágból α valószínűséggel lesz érett gyümölcs. Feltéve, hogy a fán r alma van, mennyi a valószínűsége, hogy n virág volt? ([2])

1.6.22. Van n darab urnánk, melyek mindegyikében a fehér és b piros golyó van. Az első urnából kihúzunk egy golyót, áttesszük a másodikba; majd a másodikból húzunk egyet és átrakjuk a harmadikba, ... Végül az n-edik urnából húzunk egyet. Jelölje p_n annak a valószínűségét, hogy az utolsó urnából fehér golyót húzunk, feltéve, hogy az elsőből fehéret húztunk. Igazoljuk, hogy

$$p_n = \frac{a}{a+b} + \frac{b}{a+b}(a+b+1)^{1-n}.$$

1.6.23. Egy fiú és egy lány megbeszéli, hogy két utca kereszteződésénél találkoznak egy meghatározott időpontban. Elfelejtik megbeszélni, hogy a négy sarok közül melyiknél várnak egymásra. Az útkereszteződés nagyon forgalmas, nem lehet átlátni a többi sarokra. Mindketten pontosan érkeznek, és ha a másik nincs ott, akkor 2,5 perc után átmennek a szomszédos sarkok valamelyikére, 1/2–1/2 valószínűséggel. Ez fél percet vesz igénybe, majd ha megint nem találkoztak, akkor 2,5 perc után megint sarkot váltanak. Először mindketten 1/4 valószínűséggel választanak sarkot. Természetesen az is találkozásnak számít, ha egymással szembe jönnek az úttesten.

(a) Mennyi a valószínűsége, hogy az első percen belül találkoznak?

(b) Mennyi annak a p_n valószínűsége, hogy az első $3n$ percen belül találkoznak?

(c) Mennyi annak az r_n valószínűsége, hogy pontosan a $3n$-edik percben találkoznak?

(d) Igazoljuk, hogy egy valószínűséggel véges időn belül találkoznak.

([2])

1.6.24. Pinokkiónak kilenc akadályon kell sikerrel túljutnia, hogy fabábuból kisfiúvá változhasson. Ha egy akadályon elbukik, akkor vissza kell mennie az előzőhöz, ha az elsőn bukik el, örökre fabábu marad. Pinokkió nem tanul a kudarcokból, ezért az egyes akadályokon a siker valószínűsége $1/10, 2/10, \ldots,$ $9/10$. Milyen sorrendben helyezze el a Kékhajú Tündér az akadályokat, hogy Pinokkió a legnagyobb eséllyel lehessen igazi kisfiú. Mennyi ekkor a valószínűség? (KöMaL, 2001, B.3441)

1.6.25. Szindbádnak jogában áll N háremhölgy közül egyet kiválasztania oly módon, hogy az előtte egyenként elvonuló hölgyek valamelyikére rámutat. Tegyük fel, hogy egyértelmű szigorúan monoton szépségi sorrendet tud felállítani, és a háremhölgyek bármely elvonulási sorrendje egyformán valószínű. Szindbád k hölgyet elenged, majd kiválasztja az elsőt, aki szebb az összes előtte elvonultnál. Mennyi a valószínűsége, hogy a legszebb hölgyet választja ki? Milyen k esetén lesz ez a valószínűség a legnagyobb, ha N elég nagy?

1.6.26. Shanille O'Keal büntetőket dobál egy kosárpályán. Az elsőt bedobja, a másodikat nem. Ezek után annak a valószínűsége, hogy egy büntetőt bedob, megegyezik az eddig sikeres dobásainak részarányával. Mi annak a valószínűsége, hogy az első 100 dobásból pontosan 50 sikeres?

1.6.27. Egy százszemélyes repülőgépen száz ember utazik úgy, hogy mindenkinek van előre kiosztott helye. Az első utas ezzel nem törődve véletlenszerűen leül a száz közül egy helyre. Ezután minden utas a saját helyére próbál leülni, vagy ha az foglalt, véletlenszerűen választ egy másikat. Mennyi a valószínűsége annak, hogy a századik utas a helyére ül, ha egyszerre csak egy ember foglal helyet? (KöMaL B4771)

1.6.28. A számegyenesen egy bolha ugrál. A 0-ból indul és minden ugrásának hossza 1. Minden ugrásnál az előzőektől függetlenül p valószínűséggel jobbra, $1 - p$ valószínűséggel pedig balra ugrik. Mennyi annak a valószínűsége, hogy eljut az 1-be?

Jelölje ezt a valószínűséget u. Fejezzük ki így a -1-ből 1-be jutás valószínűségét, majd a kapott egyenletet oldjuk meg u-ra.

1.6.29. A számegyenesen egy bolha ugrál. A 0-ból indul és minden ugrásának hossza 1. Először az 1-be ugrik, majd a következő ugrás mindig p valószínűséggel az előzővel egyező, $1 - p$ valószínűséggel pedig ellentétes irányú. Mennyi annak a valószínűsége, hogy visszajut a 0-ba? (KöMaL B4676)

1.6.30. Egy bolha ugrál az ABCD négyzet csúcsain, az A csúcsról indulva. Minden egyes ugrásnál $1/2 - 1/2$ valószínűséggel valamelyik szomszédos csúcsba ugrik. A bolha akkor áll meg, ha az utolsó olyan csúcsot is eléri, amin addig még nem volt. Határozzuk meg, hogy melyik csúcs mekkora valószínűséggel lesz utolsó! (KöMaL B4783)

1.6.31. A parti tüzérség 1 km távolságban felfedez egy ellenséges cirkálót, és elkezd rá tüzelni, percenként egy lövést adva le. A cirkáló az első lövés leadásakor menekülni kezd 60 km/h sebességgel. A találat valószínűsége x km távolság esetén $0,75\,x^{-2}$. Ha egy lövés talált, akkor még mindig $1/4$ valószínűséggel a cirkáló nem süllyed el, és tovább menekül. Mekkora valószínűséggel menekül el a cirkáló? (Schweitzer, 1950)

1.7. Függetlenség

1.7.1. Legyen A önmagától független esemény. Mutassuk meg, hogy $\mathbf{P}(A) =$ 0 vagy 1!

1.7.2. Válasszunk találomra az $1, 2, \ldots, n$ számok közül úgy, hogy mindegyiket $1/n$ valószínűséggel választjuk. Jelölje A_p azt az eseményt, hogy a választott szám p-vel osztható.

(a) Igazoljuk, hogy ha p_1 és p_2 relatív prím és $p_1 p_2 | n$, akkor A_{p_1} és A_{p_2} függetlenek.

(b) Igazoljuk, hogy

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

ahol $\varphi(n)$ az Euler–féle függvény, azaz $\varphi(n)$ az n-nél kisebb n-hez relatív prím pozitív egészek száma.

1.7.3. Háromszor feldobunk egy szabályos érmét. Jelentse A azt az eseményt, hogy a dobások között fej és írás is előfordul, B pedig azt, hogy legfeljebb egy írás fordul elő. Függetlenek-e A és B.

1.7.4. Adjunk példát olyan A, B, C eseményekre, melyek páronként függetlenek, de nem függetlenek.

1.7.5. Legalább hány lottószelvényt kell kitölteni ahhoz, hogy egy sorsolásnál a telitalálat valószínűsége legalább 1/2 legyen? Legalább hány hétig kell játszani egyetlen szelvénnyel, hogy annak a valószínűsége, hogy legalább egyszer volt telitalálatunk legalább 1/2 legyen?

1.7.6. Egy dobókockával tízszer dobunk. Jelölje A azt az eseményt, hogy az első 5 dobás során nincs hatos, B pedig azt, hogy tíz dobás közt nincs egyes. Mekkora az A és a B események valószínűsége? Függetlenek-e A és B?

1.7.7. Egy dobókockával n-szer dobunk. Jelölje A azt az eseményt, hogy az első m dobás során nincs hatos, B pedig azt, hogy az n dobás közt nincs egyes, $m < n$. Mekkora az A és a B események valószínűsége? Igazoljuk, hogy $\mathbf{P}(A \cap B) < \mathbf{P}(A)\mathbf{P}(B)$.

1.7.8. Legyenek $x \in [0,1]$ és $m, n \in \mathbb{N}$. Bizonyítsuk be (lehetőleg valószínűségi gondolatmenettel), hogy

$$(1 - x^n)^m + (1 - (1 - x)^m)^n \geq 1.$$

1.7.9. A szultán birodalmának mind az 1024 matematikusát börtönbe záratta. Mindegyikük csak a saját réztalléros érméjét tarthatta meg. A matematikusok tudják, hányan vannak, de semmiféle módon nem képesek kommunikálni egymással. A szultán a születésnapján nagy kegyesen a következő játékot ajánlotta a matematikusoknak: az udvaron egyenként vagy 0-t, vagy 1-et mondanak. Ha a mondott számok összege 1, akkor szabadon bocsátja őket. (A matematikusok nem adhatnak jelet egymásnak, nem tudják, hogy őket hányadiknak vitték ki, vagy hogy az előttük az udvaron lévők mit csináltak.) Mekkora eséllyel szabadulhatnak ki a matematikusok? (KöMaL B5018)

1.7.10. Két játékos, Anna és Balázs játszanak. Annának a pont hiányzik a győzelemhez, Balázsnak pedig b. Egy-egy játékot Anna $p \in (0, 1)$ valószínűséggel nyer meg, Balázs pedig $1 - p$ valószínűséggel. Hogyan osszák el a tétet?

1.7.11. Máté nagymamája meggylevest készít a vasárnapi ebédhez. Összesen 5 szem szegfűszeget tesz a levesbe. A levest 4 egyforma adagra osztják. Legyen A az az esemény, hogy Máté levesében nincs szegfűszeg, B pedig az, hogy a nagymamája levesében nincs szegfűszeg. Független-e A és B?

1.7.12. Az amerikai kosárlabdabajnokság döntőjében a keleti és a nyugati konferencia bajnoka csap össze egy 4 győzelemig tartó mérkőzéssorozaton. 2016-ban a döntőt a Cleveland Cavaliers és a Golden State Warriors vívta.

A Warriors már $3-1$ arányban vezetett, de végül mégis elveszítette a mérkőzéssorozatot. Mennyi volt ennek a vereségnek az esélye a negyedik meccs után? Mennyi volt ez a valószínűsége a második meccs után, amikor $2-0$ volt a Warriors javára? (Az egyes meccseken a két csapatnak azonos a győzelmi esélye. Döntetlen nincs, minden mérkőzést eldöntenek hosszabbításokkal.)

2. Véletlen változók

2.1. Eloszlásfüggvény, eloszlás, sűrűségfüggvény

2.1.1. Eloszlásfüggvények-e a következő függvények?

(a) $F(x) = \begin{cases} 0, & \text{ha } x \le 0, \\ \frac{x}{1+x}, & \text{ha } x > 0. \end{cases}$

(b) $F(x) = e^{-e^{-x}}$.

(c) $F(x) = \begin{cases} 0, & \text{ha } x < 0, \\ 1 - \frac{1-e^{-x}}{x}, & \text{ha } x \ge 0. \end{cases}$

2.1.2. Milyen a, b értékek esetén lesz eloszlásfüggvény

(a) $F(x) = a + b \arctan x$?

(b) $F(x) = e^{-ae^{-bx}}$?

2.1.3. Legyen $F(x)$ eloszlásfüggvény. Igazoljuk, hogy

$$G_1(x) = \frac{1}{h}\int_x^{x+h} F(t)\,\mathrm{d}t \quad \text{és} \quad G_2(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,\mathrm{d}t$$

is az!

2.1.4. Adjunk példát olyan F eloszlásfüggvényre, mely tiszta ugrófüggvény, és bármely $a < b$ esetén $F(b) - F(a) > 0$!

2.1.5. Az alábbi számsorozatok közül melyek alkotnak valószínűségeloszlást?

(a) $p^k q^2$, ahol $p \in (0,1)$, $q = 1-p$, $k = 1, 2, \ldots$;

(b) $\frac{1}{k(k+1)}$, $k = 1, 2, \ldots$;

(c) $3^k/k!e^{-3}$, $k = 0, 1, 2, \ldots$.

2.1.6. Sűrűségfüggvény-e?

(a) $f(x) = (I_{(0,1)}(x) \sin x)/2$;

(b) $f(x) = I_{(1,\infty)}(x)x^{-2}$;

(c) $f(x) = I_{(0,\infty)}(x)\lambda e^{-\lambda x}$, ahol $\lambda > 0$.

(d) $f(x) = (\pi(1+x^2))^{-1}$.

2.2. Diszkrét véletlen változók

2.2.1. Egy embernek n egyforma kinézetű kulcsa van, melyek közül pontosan egy nyitja az ajtót. Emberünk véletlenül választva sorra próbálja a kulcsokat addig, amíg a jó kulcs elő nem kerül. Valamely $k \in \{1, 2, \ldots, n\}$ esetén mennyi a valószínűsége, hogy a k-adik próbálkozása sikeres, ha

(a) a kipróbált rossz kulcsokat mindig félreteszi?

(b) a kipróbált rossz kulcsokat sose teszi félre?

Várhatóan hányadik próbálkozása sikeres?

2.2.2. Egy szabályos érmét feldobunk egymás után háromszor. Jelölje X a dobott fejek és a dobott írások számának a különbségét. Határozzuk meg X eloszlását és eloszlásfüggvényét (készítsünk ábrát is)! Mi a valószínűsége, hogy X a $[-\sqrt{2}, \sqrt{2}]$ intervallumba esik?

2.2.3. Egy hallgató 26 tétel közül 16-ot jelesre tud, 8-at jóra, 2-t viszont csak közepesre. A vizsgatételek kiválasztásakor 26 tétel közül húz 1-et. Adjuk meg a kapott jegy értékének eloszlását, várható értékét és szórását!

2.2.4. Ötöslottón egy szelvénnyel játszva határozzuk meg a találataink számának eloszlását!

2.2.5. Határozzuk meg az ötöslottón kihúzott legnagyobb szám eloszlását, várható értékét és szórását!

2.2.6. Máté nagymamája meggylevest készít a vasárnapi ebédhez. Összesen 5 szem szegfűszeget tesz a levesbe. A levest 4 egyforma adagra osztják. Adjuk meg Máté levesében található szegfűszegek számának eloszlását, várható értékét, és szórását!

2.2.7. Mind a 4 pár különböző cipőmet az előszobában levő beépített szekrényben tartom. Az előszobában kiégett a villany, így sötétben keresgélve véletlenszerűen kiveszek 4 cipőt. Jelölje X a kivett összeillő párok számát! Adjuk meg X eloszlását, várható értékét és szórását!

2.2.8. Száz alma közül tíz férges. Véletlenül kiválasztunk ötöt! Adjuk meg a férges almák számának eloszlását!

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

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

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

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

Innen a várható értéket meghatározhatjuk a definíció alapján. Eszerint

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

Ezt kellene zárt alakra hozni. Vegyük észre, hogy

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

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

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

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

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

Ezt visszahelyettesítve a várható értékre kapott formulába

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

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

A szórás meghatározásához először a második momentum kell. Ez

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

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

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

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

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

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

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

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

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

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

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

Legyen $i = 1, 2, \ldots, 20$ esetén

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

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

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

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

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

Tehát

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

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

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

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

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

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

A definíció szerint

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

és

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

hiszen annak a valószínűsége, hogy az első és második golyó is piros

$$\frac{20 \cdot 19}{50 \cdot 49}.$$

Összegezve,

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

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

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

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

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

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

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

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

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

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

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

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

2.2.11. Egy szabályos érmével 100-szor dobunk. Jelölje X a különböző FF sorozatok számát. Határozzuk meg X várható értékét és szórását!

2.2.12. Egy csomag francia kártyát megkevertünk, majd egyesével kihúzzuk a lapokat. Adjuk meg a második ász helyének eloszlását!

2.2.13. Egy urnában 101 golyó van, közülük pontosan három piros. A golyókat visszatevés nélkül egyesével kihúzzuk. Jelölje X a második piros sorszámát. Adjuk meg X eloszlását!

2.2.14. Egy fiókban három pár kesztyű van összekeveredve: az egyik pár fekete, a másik szürke, a harmadik piros. (A három pár kesztyű csak a színében különböző.) A fiókból egyesével elkezdjük kihúzni a kesztyűket úgy, hogy húzás előtt nem nézzük meg a kesztyű színét, és a kihúzott kesztyűket nem tesszük vissza a fiókba. Addig folytatjuk a húzást, amíg lesz két azonos színű kesztyűnk. Határozza meg annak a hat eseménynek a valószínűségét, hogy ehhez 1, 2, 3, 4, 5, illetve 6 kesztyű kihúzására lesz szükség, majd számítsa ki a húzások számának várható értékét és szórását!

Megoldás. Ez egyszerű, pár éve érettségi feladat volt. Jelölje X azt, hogy hányszor kellett húzni, hogy legyen egy pár. Világos, hogy X lehetséges értékei 2,3,4. Tehát

$$\mathbf{P}(X=1)=\mathbf{P}(X=5)=\mathbf{P}(X=6)=0.$$

Elsőre húzunk valamit. Ha $X=2$, akkor a második kihúzott kesztyű éppen az első párja. Mivel 5 kesztyű maradt, ennek a valószínűsége

$$\mathbf{P}(X=2)=\frac{1}{5}.$$

Ha $X=3$, akkor másodikra nem húztuk ki az első párját, ennek a valószínűsége $4/5$, harmadikra viszont párt húztunk. A harmadik húzás előtt 4 kesztyű van, ebből 2 jó, tehát a pár valószínűsége $1/2$. Összegezve

$$\mathbf{P}(X=3)=\frac{4}{5}\cdot\frac{2}{4}=\frac{2}{5}.$$

Itt valójában egy feltételes valószínűséges gondolatmenetet követtünk. Úgy is gondolkodhatunk, hogy az első húzás után van 5 kesztyű, és kétszer húzunk. Mivel fontos a sorrend az összes esetek száma $5\cdot 4=20$. A kedvező eseteknél a második húzás nem pár, erre van 4 lehetőségünk, míg a harmadik húzásra párt húzunk, de már két kesztyűt húztunk ki, úgyhogy a jó esetek száma 2. Összegezve a kedvező esetek száma $4\cdot 2=8$. Persze azt kaptuk mint korábban, mert $\frac{8}{20}=\frac{2}{5}$.

Végül

$$\mathbf{P}(X=4)=1-\mathbf{P}(X=2)-\mathbf{P}(X=3)=\frac{2}{5}.$$

Persze ezt is kiszámolhatjuk a korábbiak szerint.

Tehát a valószínűségeloszlás:

$$\mathbf{P}(X=2)=\frac{1}{5},\quad \mathbf{P}(X=3)=\frac{2}{5},\quad \mathbf{P}(X=4)=\frac{2}{5}.$$

A várható érték

$$\mathbf{E}(X)=2\cdot\frac{1}{5}+3\cdot\frac{2}{5}+4\cdot\frac{2}{5}=\frac{16}{5}.$$

A második momentum

$$\mathbf{E}(X^2) = 2^2 \cdot \frac{1}{5} + 3^2 \cdot \frac{2}{5} + 4^2 \cdot \frac{2}{5} = \frac{54}{5},$$

ahonnan a szórásnégyzet

$$\mathbf{D}^2(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \frac{54}{5} - \frac{256}{25} = \frac{14}{25},$$

és a szórás $\mathbf{D}(X) = \sqrt{14}/5$. □

2.2.15. Földobunk n-szer egy szabályos pénzérmét. Határozzuk meg az F-I, I-F váltások számának eloszlását!

2.2.16. Legyenek $X_1, X_2, \ldots, X_n$ véletlen változók függetlenek és azonos eloszlásúak, melyre $\mathbf{P}(X_i = k) = 1/3$, $k \in \{0, 1, 2\}$. Adjuk meg az $Y = X_1 X_2 \cdots X_n$ szorzat eloszlását!

2.2.17. Egy városban 200 taxi közlekedik. Telefonon taxit rendelünk, és ha van szabad taxi, akkor a központ a legközelebbit hozzánk küldi. Feltesszük, hogy a taxik egymástól függetlenül, egyenletes eloszlás szerint helyezkednek el a városban, és mindegyik egymástól függetlenül $2/3$ valószínűséggel foglalt. Továbbá egy taxi helyzete a városon belül független attól, hogy foglalt-e vagy sem. Mennyi annak a valószínűsége, hogy a legközelebbi szabad taxi 1 km-es körzetünkben legyen (mely nem nyúlik ki a városból), feltéve, hogy van szabad taxi? A város területe $28,26\,\mathrm{km}^2$. ([2])

2.2.18. Egy urnában egy piros és egy fehér golyó van. Visszatevéssel húzunk az urnából, minden húzás után még egy piros golyót teszünk az urnába. Jelölje X_f annak a kísérletnek a sorszámát, amikor először húztunk fehéret. Adjuk meg X_f eloszlását, várható értékét!

2.2.19. Kupongyűjtő probléma. Egy N különböző elemből álló sokaságból visszatevéses mintát veszünk. Jelölje S_r azt a véletlen számot, ahány elemet kellett húznunk, hogy kapjunk r különböző elemet. Határozzuk meg S_r várható értékét, szórását, majd adjunk ezekre kezelhető aszimptotikus egyenlőséget.

Útmutatás: Vezessük be az $X_k = S_{k+1} - S_k$ változót.

2.2.20. Egy urnában N golyó van, megszámozva 1-től N-ig. Visszatevéssel n elemű mintát veszünk. Jelölje X a legnagyobb mintaelemet. Adjuk meg X eloszlását, várható értékét, és az utóbbira adjunk aszimptotikát. Ha N értéke ismeretlen, hogyan lehetne becsülni?

2.2.21. Egy gombafajta spórái nyolcelemű lánc alakjában keletkeznek. A lánc különböző részekre szakadhat, mind a hét lehetséges helyen egymástól függetlenül p valószínűséggel. Határozzuk meg az i spórából álló láncok várható értékét!

Egy tényleges kísérlet során 7251 spórát számoltak meg, melyek $N = 907$ láncból származtak (5 spóra elveszett). A spórák 1975 láncra szakadtak szét. Adjunk becslést p értékére! ([4])

2.2.22. Jelölje S_n n elem véletlen permutációja során a fixpontok számát. Határozzuk meg S_n várható értékét és szórását!

2.2.23. Stefan Banach erős dohányos volt. Mindkét zsebében egy-egy doboz gyufát tartott, 100-100 szál gyufával. Minden alkalommal, amikor rágyújtott találomra benyúlt az egyik zsebébe, és az abban levő gyufásskatulyából kivett egy szál gyufát, majd a dobozt visszatette abba a zsebébe, amelyikből kivette. Egyszer csak azt vette észre, hogy a gyufásskatulya kiürült. Határozzuk meg a másik zsebében levő gyufák számának eloszlását!

2.2.24. Vérvizsgálatot végeznek N embernél. Egyszerre k ember vérét összeöntik és analizálják. Ha az eredmény negatív, akkor egy vizsgálat elég k embernek, ha pozitív, akkor a k személyt egyenként is meg kell vizsgálni, és így $k + 1$ vizsgálat kellett. A vizsgálat eredménye egymástól függetlenül p valószínűséggel pozitív.

(a) Mennyi a valószínűsége, hogy k ember vérét együtt vizsgálva, az eredmény pozitív?

(b) Jelölje X a szükséges vizsgálatok számát! Mennyi X várható értéke?

(c) Minimalizáljuk $\mathbf{E}(X)$-et k függvényeként!

(d) Lássuk be, hogy a kapott k közel van $1/\sqrt{p}$-hez, és a várható érték $2N\sqrt{p}$ körül lesz.

2.2.25. Csodaország munka törvénykönyve szerint egy cég minden munkása fizetett szabadságot kap azokon a napokon, amikor legalább az egyiküknek születésnapja van. Ezen napok kivételével azonban az év minden napján mindenkinek dolgoznia kell. Minden munkás 1 TV-készüléket készít egy nap alatt. Hány alkalmazottat vegyen fel a cégtulajdonos, ha azt akarja, hogy a gyártott TV-készülékek számának a várható értéke maximális legyen?

2.2.26. Legyen X nemnegatív egész értékű véletlen változó, és tegyük fel, hogy $\mathbf{E}(X) < \infty$. Bizonyítsuk be, hogy

$$\mathbf{E}(X) = \sum_{i=1}^{\infty} \mathbf{P}(X \geq i).$$

2.2.27. Határozzuk meg $1/(X+1)$ várható értékét, ha

(a) $X \sim \text{Binom}(n,p)$;

(b) $X \sim \text{Poisson}(\lambda)$;

(c) X geometriai eloszlású;

(d) X hipergeometriai eloszlású.

2.2.28. Legyen X véletlen változó, melyre $\mathbf{E}(X^2) < \infty$. Mutassuk meg, hogy az $f(x) = \mathbf{E}[(X-x)^2]$ az $x = \mathbf{E}X$ pontban veszi föl a minimumát, ami éppen $\mathbf{D}^2(X)$.

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

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

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

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

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

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

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

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

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

Az i, j, k értékeket pedig $\binom{10}{3} = \frac{10\cdot 9\cdot 8}{2\cdot 3} = 120$ féleképpen választhatjuk, azaz összesen $6 \cdot 120 = 720$ ilyen tagot kapunk. Összegezve

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

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

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

Tehát

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

Végül kiszámoljuk a megfelelő momentumokat. Definíció szerint

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

Végül ezeket visszahelyettesítve,

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

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

2.2.30. Egy halastóban N hal van. Kihalászunk M halat, megjelöljük őket, és visszaeresztjük a tóba. Bizonyos idő elteltével, miután jól elkeveredtek, kihalászunk n-et. Ezek között legyen a megjelöltek száma X. A teljes halállomány N meghatározására az $Mn/(X+1)$ becslést használjuk. Számítsuk ki ennek a várható értékét és szórását! Miért nem a logikusabb Mn/X becslést használjuk?

2.3. Folytonos véletlen változók

2.3.1. A $(0, 1)$ intervallumon találomra kijelölünk három pontot. Határozzuk meg a középső nullától vett távolságának eloszlás- és sűrűségfüggvényét, várható értékét és szórását!

2.3.2. A $[0,1]$ intervallumon kijelölünk három pontot egymástól függetlenül, egyenletes eloszlás szerint. Határozzuk meg a legkisebb eloszlásfüggvényét és várható értékét!

2.3.3. Válasszunk az egységnégyzetben egy pontot véletlenszerűen. Legyen X a pontnak a négyzet határától vett távolsága. Adjuk meg X eloszlásfüggvényét, várható értékét, szórását!

2.3.4. Egy permetező szakaszoló szelep napokban mért élettartamának sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{3000}{x^4}, & \text{ha } x \geq 10, \\ 0, & \text{különben.} \end{cases}$$

Mennyi a valószínűsége, hogy a permetező szakaszoló szelep 20 napot túlél? Határozzuk meg a szelep élettartamának eloszlásfüggvényét, várható értékét és szórását!

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

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

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

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

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

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

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

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

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

Tehát

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

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

A várható értéket is definíció alapján számoljuk. Eszerint

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

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

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

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

A szórásnégyzet kiszámításához először a második momentumot kell meghatározni. Ez

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

És így a szórásnégyzet

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

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

2.3.5. Legyen a X véletlen változó sűrűségfüggvénye $f(x) = c/x^2$, ha $x > 1$.

(a) Határozzuk meg c értékét!

(b) Adjuk meg X várható értékét (ha létezik)!

(c) Mennyi $\mathbf{P}(X > 4)$?

(d) Legyen $Y = 1/X$. Adjuk meg Y eloszlás–, és sűrűségfüggvényét!

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

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

Tehát $c = 1$.

(b) A várható érték definíció szerint

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

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

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

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

Az eloszlásfüggvényt is meghatározzuk. A definíció alapján

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

Ez az 1 paraméterű Pareto-eloszlás.

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

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

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

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

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

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

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

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

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

2.3.6. Anna és Szabina minden szerdán fodrászhoz mennek. Anna 2 és 3 óra között, Szabina pedig 2 és fél 3 között végez egy véletlenszerű időpontban, egymástól függetlenül. Egymást megvárják, majd együtt indulnak haza. Jelölje X azt az időt amennyit a hamarabb végző lány vár a másikra! Határozzuk meg X eloszlását, várható értékét és szórását!

2.3.7. Legyen $X \sim \mathrm{E}(-1,1)$ eloszlású véletlen változó. Határozzuk meg a következő véletlen változók sűrűségfüggvényeit:

(a) $|X|$;

(b) X^2;

(c) e^X.

2.3.8. Legyen $X \sim \mathrm{Exp}(\lambda)$ eloszlású véletlen változó. Határozzuk meg a következő véletlen változók sűrűségfüggvényeit:

(a) $2X + 3$;

(b) X^3;

(c) $\sqrt{X}$.

2.3.9. Válasszunk két számot egymástól függetlenül az egyenletességi hipotézis szerint a $(-1,1)$ intervallumból! Adjuk meg a két szám maximumának eloszlásfüggvényét! Számoljuk ki a várható értéket és a szórást is!

2.3.10. Válasszunk $2n+1$ pontot egymástól függetlenül, egyenletes eloszlás szerint a $[0,1]$ intervallumban. Adjuk meg a középső pont Z eloszlásfüggvényét! Mutassuk meg, hogy tetszőleges $\varepsilon > 0$ esetén

$$\mathbf{P}(|Z-1/2| > \varepsilon) \to 0$$

amint $n \to \infty$.

2.4. Vektorváltozók

2.4.1. Egy vállalat egy hónapra eső profitja a havi teljes bevétel és a havi teljes kiadás különbségeként áll elő, ahol a bevétel és a kiadás is véletlen változó. A bevétel várható értéke 120 millió forint 30 millió forint szórással, míg a kiadás várható értéke 80 millió forint 20 millió forint szórással. Határozzuk meg az egy hónapra jutó profit várható értékét és szórását akkor, ha a bevétel és a kiadás független, illetve akkor, ha a közöttük lévő korrelációs együttható 0,8. A korreláció függvényében írjuk fel formulával és ábrázoljuk grafikonon a profit várható értékét és varianciáját.

2.4.2. A megtakarított pénzünket értékpapírba fektetjük, 20 darabot vásárolunk az A vál- lalat és 10 darabot a B vállalat részvényeiből. Egy év múlva a két vállalat részvényei várható értékben 700 illetve 1500 dollárt fognak majd érni, az árfolyamok szórása pedig 20 illetve 80 dollár.

(a) Tegyük fel, hogy a részvények árfolyama független egymástól. Várhatóan menynyit fog majd érni a portfóliónk egy év múlva? Mennyi a portfólió értékének a szórása?

(b) Tegyük fel, hogy a részvények árfolyama nem független egymástól. Az árfolyamok közötti korrelációs együttható függvényében írjuk fel formulával és ábrázoljuk a portfólió értékének várható értékét és varianciáját!

(c) Milyen kapcsolat van a korrelációs együttható és a befektetés kockázata között? Ha én egy kockázatkerülő befektető vagyok, akkor pozitív vagy negatív korrelációjú értékpapírokból állítsak össze portfóliót?

Megoldás. Jelölje X az A vállalat, Y pedig a B vállalat részvényének értékét egy év múlva. A feladat szerint $\mathbf{E}(X) = 700$, $\mathbf{D}(X) = 20$, $\mathbf{E}(Y) = 1500$, és $\mathbf{D}(Y) = 80$. Egy év múlva a portfóliónk értéke $20X + 10Y$.

(a) Az összeg várható értéke *mindig* a várható értékek összege, az összeg szórásnégyzete pedig akkor egyenlő a szórásnégyzetek összegével, ha a változók függetlenek (na jó, korrelálatlanság is elég). Tehát, *mindig*

$$\mathbf{E}(20X + 10Y) = 20\mathbf{E}(X) + 10\mathbf{E}(Y) = 29000,$$

és ha X és Y függetlenek, akkor

$$\begin{aligned}\mathbf{D}^2(20X+10Y) &= \mathbf{D}^2(20X)+\mathbf{D}^2(10Y) = (20)^2\mathbf{D}^2(X)+(10)^2\mathbf{D}^2(Y)\\ &= 160000+640000 = 800000,\end{aligned}$$

azaz $\mathbf{D}(20X+10Y) = 400\sqrt{5}$.

(b) Láttuk,

$$\mathbf{E}(20X+10Y) = 20\mathbf{E}(X)+10\mathbf{E}(Y) = 29000,$$

azaz a várható érték nem függ a korrelációtól.

A szórásnégyzet (variancia) az általános esetben, használva a korreláció $\rho(X,Y) = \mathbf{Cov}(X,Y)/(\mathbf{D}(X)\mathbf{D}(Y))$ definícióját, és hogy a kovariancia olyan mint a szorzás, bilineáris

$$\begin{aligned}\mathbf{D}^2(20X+10Y) &= \mathbf{Cov}(20X+10Y, 20X+10Y)\\ &= \mathbf{Cov}(20X,20X)+2\mathbf{Cov}(20X,10Y)+\mathbf{Cov}(10Y,10Y)\\ &= \mathbf{D}^2(20X)+2\mathbf{D}(X)\mathbf{D}(Y)\rho+\mathbf{D}^2(10Y)\\ &= 400^2+2\cdot 400\cdot 800\cdot\rho+800^2\\ &= 400^2(5+4\rho).\end{aligned}$$

Ez egyszerű lineáris függvény. Figyeljünk, hogy $\rho\in[-1,1]$.

(c) Nyilván a kockázatos befektetés azért kockázatos, mert nagyon ingadozik a várható értéke körül, azaz a szórása nagyobb. Tehát, ha én kockázatkerülő vagyok, akkor minél kisebb szórásnégyzetű portfóliót szeretnék összerakni. Ezt úgy érhetem el, ha ρ a korreláció negatív. $\square$

2.4.3. Bence és Luca testvérek. Ebéd után mindketten véletlentől függő ideig alszanak. Bence esetében ez átlagosan 2 óra, a szórás 30 perc, míg Luca esetében 1,5 óra, 20 perc szórással. Határozzuk meg Bence és Luca együttes (Bence + Luca) alvásának várható értékét és szórását, ha az alvásmennyiségek egymástól függetlenek, illetve akkor, ha a közöttük lévő korrelációs együttható 0,5.

2.4.4. A Real Madrid 2018/2019-es idényben az egy mérkőzésen lőtt góljainak száma Poisson-eloszlást követ $\lambda = 3$ paraméterrel, míg a kapott gólok száma Poisson-eloszlást követ $\mu =$0,7 paraméterrel. Adjuk meg a Real Madrid egy mérkőzésén esett összes gól számának várható értékét és szórásnégyzetét abban az esetben ha

(a) a lőtt és kapott gólok száma függetlenek;

(b) a lőtt és kapott gólok számának korrelációs együtthatója 0,4.

2.4.5. Egy dobókockával az első hatosig dobunk. Jelölje X a szükséges dobások számát, Y pedig a dobott egyesek számát. Adjuk meg az együttes eloszlást és a kovarianciát!

2.4.6. Legyen az (X, Y) véletlen változó. eloszlása egyenletes az egységkörben. Határozzuk meg az együttes eloszlásfüggvényt és a peremeloszlások sűrűségfüggvényeit!

2.4.7. Két szabályos kockával játszunk. Jelölje X az első kockával dobott számot és Y a dobott számok nagyobbikát. Adjuk meg az együttes eloszlást és az összeg várható értékét, szórását!

Megoldás. Mind az X, mind az Y véletlen változó lehetséges értékei $1, 2, \ldots,$ 6. Tehát X, Y diszkrét véletlen változók. Az (X, Y) vektorváltozó lehetséges értékei számpárok. Mivel $Y \geq X$, hiszen Y a két szám közül a nagyobb, ezért a lehetséges értékek halmaza

$$\{(i, j) \,:\, 1 \leq i \leq j \leq 6\}.$$

Először meghatározzuk a $\mathbf{P}((X, Y) = (1, 1))$ valószínűséget. Mivel X az első kockával dobott szám, Y pedig a két szám nagyobbika, ezért csak úgy lehet mindekettő 1, ha mindkét kockával 1-et dobtunk. Mivel az összes eset száma $6 \cdot 6 = 36$, így

$$\mathbf{P}((X, Y) = (1, 1)) = \mathbf{P}(X = 1, Y = 1) = \frac{1}{36}.$$

Most vizsgáljuk a $\mathbf{P}(X = 1, Y = 2)$ valószínűséget. Az első kockával megint 1-et dobtunk, és a dobott számok maximuma 2. Ez csak úgy lehet, ha az első kockával 1-est, a másodikkal 2-est dobtunk. Megint 1 kedvező eset, ezért

$$\mathbf{P}((X, Y) = (1, 2)) = \mathbf{P}(X = 1, Y = 2) = \frac{1}{36}.$$

Látjuk, hogy tetszőleges $j = 1, 2, \ldots, 6$ esetén

$$\mathbf{P}((X, Y) = (1, j)) = \mathbf{P}(X = 1, Y = j) = \frac{1}{36}.$$

Nézzük a $\mathbf{P}(X = 2, Y = 2)$ valószínűséget. (Persze, ha $X = 2$ akkor $Y \geq 2$.) Ez azt jelenti, hogy az első kockával 2-est dobtunk, és a dobott számok nagyobbika 2. Ez úgy lehet, ha a második kockával 1-est vagy 2-est dobtunk. Azaz, most két kedvező eset van,

$$\mathbf{P}((X, Y) = (2, 2)) = \mathbf{P}(X = 2, Y = 2) = \frac{2}{36}.$$

Ha $X = 2$ és $Y = 3$, akkor az első kockával 2-est dobtunk, a nagyobbik szám pedig 3, tehát a második kockával 3-ast dobtunk. Ez megint egy eset, tehát

$$\mathbf{P}(X = 2, Y = 3) = \frac{1}{36}.$$

Ugyanígy tetszőleges $j \geq 3$ esetén

$$\mathbf{P}(X = 2, Y = j) = \frac{1}{36}.$$

Most már látjuk az általános képet. Ha $X = i$, $Y = i$, akkor az azt jelenti, hogy az első kockával i-t dobtunk, és a nagyobbik szám i, tehát a második kockával az $1, 2, \ldots, i$ számok valamelyikét dobtuk. Ez i lehetőség, azaz

$$\mathbf{P}(X = i, Y = i) = \frac{i}{36}.$$

Ha pedig $X = i$ és $Y = j > i$, akkor az első szám i, a nagyobbik j, és mivel ez nem lehet az első, ezért a második j. Tehát

$$\mathbf{P}(X = i, Y = j) = \frac{1}{36}.$$

Röviden

$$\mathbf{P}(X = i, Y = j) = \begin{cases} \frac{i}{36}, & \text{ha } i = j, \\ \frac{1}{36}, & \text{ha } i < j. \end{cases}$$

Ez az együttes eloszlás.

A peremeloszlásokat úgy kapjuk meg, hogy rögzítjük az egyik változó értékét, és kiösszegzünk a másik összes lehetséges értékére (folytonos esetben ugyanez, csak összegzés helyett, a sűrűséget kell integrálni). Azaz

$$\mathbf{P}(X = i) = \sum_{j=i}^{6} \mathbf{P}(X = i, Y = j) = \frac{i}{36} + \sum_{j=i+1}^{6} \frac{1}{36} = \frac{6}{36} = \frac{1}{6}.$$

Ez persze nem nagy meglepetés, ehhez nem kellett volna ennyit számolni. A szabályos kockával $1/6$ a valószínűsége minden lehetséges értéknek. Az Y eloszlása pedig

$$\mathbf{P}(Y = j) = \sum_{i=1}^{j} \mathbf{P}(X = i, Y = j) = \sum_{i=1}^{j-1} \frac{1}{36} + \frac{j}{36} = \frac{2j - 1}{36}.$$

Innen a várható értékek definíció szerint számolhatók. Valóban

$$\mathbf{E}(X) = \sum_{i=1}^{6} \frac{1}{6} \cdot i = \frac{1}{6} \frac{6 \cdot 7}{2} = \frac{7}{2},$$

és

$$\mathbf{E}(Y) = \sum_{j=1}^{6} \frac{2j-1}{36} \cdot j = \frac{1}{36}\left(2\frac{6\cdot 7\cdot 13}{6} - \frac{6\cdot 7}{2}\right) = \frac{182-21}{36} = \frac{161}{36}.$$

Az összeg várható értéke az a várható értékek összege, azaz

$$\mathbf{E}(X+Y) = \frac{7}{2} + \frac{161}{36} = \frac{126+161}{36} = \frac{287}{36}.$$

Az összeg szórásnégyzetének kiszámítása macerásabb, hiszen X és Y nyilván nem függetlenek (hát persze, a lehetséges értékek $1, 2, \ldots, 6$, ugyanakkor $X \leq Y$). Az összeg szórásnégyzete akkor egyenlő a szórásnégyzetek összegével, ha *a változók függetlenek* (vagy korrelálatlanok). Általában, a kovariancia tulajdonságai szerint (vegyük észre, hogy a kovariancia úgy viselkedik, mint egy szorzás; pontosabban ő egy bilineáris funkcionál)

$$\begin{aligned}\mathbf{D}^2(X+Y) &= \mathbf{Cov}(X+Y, X+Y)\\ &= \mathbf{Cov}(X,X) + 2\mathbf{Cov}(X,Y) + \mathbf{Cov}(Y,Y)\\ &= \mathbf{D}^2(X) + 2\mathbf{Cov}(X,Y) + \mathbf{D}^2(Y).\end{aligned}$$

A szórásnégyzetek definíció alapján számolhatók

$$\begin{aligned}\mathbf{D}^2(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 &= \sum_{i=1}^{6} \frac{1}{6} \cdot i^2 - \left(\frac{7}{2}\right)^2\\ &= \frac{6\cdot 7\cdot 13}{6\cdot 6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}.\end{aligned}$$

Ugyanígy

$$\mathbf{D}^2(Y) = \mathbf{E}(Y^2) - (\mathbf{E}(Y))^2,$$

és

$$\mathbf{E}(Y^2) = \sum_{j=1}^{6} \frac{2j-1}{36} \cdot j^2 = \frac{1}{36}\left(2\left(\frac{6\cdot 7}{2}\right)^2 - \frac{6\cdot 7\cdot 13}{6}\right) = \frac{791}{36},$$

tehát

$$\mathbf{D}^2(Y) = \frac{791}{36} - \left(\frac{161}{36}\right)^2 = \frac{2555}{1296}.$$

Maradt a kovariancia. A kovariancia tulajdonságai szerint

$$\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - (\mathbf{E}(X) \cdot \mathbf{E}(Y)).$$

A várható értékeket már kiszámoltuk, tehát már csak a szorzat várható értéke kell. A szorzat egy kétváltozós függvény, tehát a várható érték tulajdonságai szerint

$$\mathbf{E}(XY) = \sum_{i,j} \mathbf{P}(X = i, Y = j) \cdot i \cdot j.$$

Ez egy 21 tagú összeg, ki kell számolni:

$$\begin{aligned}
\mathbf{E}(XY) &= \sum_{i=1}^{6} \sum_{j=i}^{6} \mathbf{P}(X = i, Y = j) \cdot i \cdot j \\
&= \sum_{i=1}^{6} i \left(i \cdot \frac{i}{36} + \frac{1}{36} \frac{6 \cdot 7 - i(i+1)}{2} \right) \\
&= \frac{1}{36} \sum_{i=1}^{6} i \cdot \left(i^2 + 21 - \frac{i(i+1)}{2} \right) \\
&= \frac{1}{36} \sum_{i=1}^{6} \left[\frac{i^3}{2} - \frac{i^2}{2} + 21 \cdot i \right] \\
&= \frac{1}{36} \left[\frac{1}{2} \left(\frac{6 \cdot 7}{2} \right)^2 - \frac{1}{2} \frac{6 \cdot 7 \cdot 13}{6} + 21 \cdot \frac{6 \cdot 7}{2} \right] = \frac{154}{9}
\end{aligned}$$

Tehát

$$\mathbf{Cov}(X, Y) = \frac{154}{9} - \frac{7}{2} \cdot \frac{161}{36} = \frac{35}{24}.$$

A kovariancia pozitivitása azt jelenti, hogy ha az egyik változó nagy, akkor a másik is hajlamos nagyobb lenni. Ez intuitíven világos ebben a példában.

□

2.4.8. Válasszunk két számot egymástól függetlenül az egyenletességi hipotézis szerint a $(0, 1)$ intervallumból! Adjuk meg a maximum és a minimum együttes eloszlását, és számoljuk ki a kovarianciájukat!

Megoldás. Legyen U_1 és U_2 a két, függetlenül egyenletes eloszlás szerint választott pont, és jelölje X a maximumot, Y pedig a minimumot. Először meghatározzuk X és Y eloszlását külön-külön.

Világos, hogy X lehetséges értékei a $[0, 1]$ intervallum elemei, és $x \in [0, 1]$ esetén $\{X \leq x\}$ pontosan akkor teljesül, ha mindkét pont a $[0, x]$ intervallumba esik, aminek valószínűsége x^2. Tehát

$$F(x) = \mathbf{P}(X \leq x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ x^2, & \text{ha } x \in [0, 1], \\ 1, & \text{ha } x \geq 1. \end{cases}$$

Innen deriválással kapjuk a sűrűségfüggvényt, azaz

$$f(x) = \begin{cases} 2x, & \text{ha } x \in (0,1), \\ 0, & \text{különben.} \end{cases}$$

A várható érték

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_0^1 x \cdot 2x \mathrm{d}x = \frac{2}{3}.$$

Hasonlóan, az Y lehetséges értékei a $[0,1]$ intervallum elemei, és $\{Y \leq y\}$ pontosan akkor teljesül, ha legalább 1 pont a $[0,y]$ intervallumba esik, azaz vagy pontosan 1 esik oda, aminek a valószínűsége

$$2 \cdot y \cdot (1-y),$$

melyik pont esik oda, az a $[0,y]$-ba esik, a másik pedig nem; vagy pontosan 2 pont esik oda, aminek a valószínűsége

$$y^2,$$

hiszen ekkor mindkettő a $[0,y]$-ba esik. Tehát az eloszlásfüggvény

$$G(y) = \mathbf{P}(Y \leq y) = \begin{cases} 0, & \text{ha } y \leq 0, \\ 2y(1-y) + y^2 = 2y - y^2, & \text{ha } y \in [0,1], \\ 1, & \text{ha } y \geq 1. \end{cases}$$

Innen a sűrűségfüggvény

$$g(y) = \begin{cases} 2 - 2y, & \text{ha } y \in (0,1), \\ 0, & \text{különben.} \end{cases}$$

A várható érték

$$\mathbf{E}(Y) = \int_0^1 y(2-2y)\mathrm{d}y = \frac{1}{3}.$$

Ez heurisztikusan világos, hiszen 2 véletlen pont nagyjából 3 egyforma hosszú intervallumra osztja az egységnyi intervallumot. Tehát a kisebbik pont várható értéke 1/3, a nagyobbiké pedig 2/3, ahogy kiszámoltuk.

Nézzük most az együttes eloszlást. Legyen

$$\mathbf{P}(X \leq x, Y \leq y) = H(x,y)$$

az együttes eloszlásfüggvény. Mivel a lehetséges értékek a halmaza $[0,1] \times [0,1]$, ezért

$$H(x,y) = \begin{cases} 1, & \text{ha } x \geq 1, y \geq 1, \\ \mathbf{P}(X \leq x), & \text{ha } y \geq 1, \\ \mathbf{P}(Y \leq y), & \text{ha } x \geq 1, \\ 0, & \text{ha } x \leq 0 \text{ vagy } y \leq 0. \end{cases}$$

Így az érdekes eset amikor $x, y \in [0,1]$. Sőt, mivel $X \geq Y$, így ha $x \leq y$ akkor

$$\mathbf{P}(X \leq x, Y \leq y) = \mathbf{P}(X \leq x).$$

Hát persze, ha a maximum kisebb, mint x, akkor a minimum is kisebb, így az a feltétel elhagyható. Legyen tehát $1 \geq x > y \geq 0$. Ekkor az $\{X \leq x, Y \leq y\}$ esemény akkor következik be, ha mindkét pont $\leq x$, és legalább egy pont $\leq y$. Ez úgy lehet, ha vagy mindkét pont a $[0,y]$-ba esik, aminek a valószínűsége y^2, vagy az egyik a $[0,y]$-ba, a másik $(y,x]$-be, aminek a valószínűsége $2 \cdot y \cdot (x-y)$ (melyik pont, hova, hova). Tehát

$$\mathbf{P}(X \leq x, Y \leq y) = y^2 + 2y(x-y) = 2xy - y^2.$$

Összegezve,

$$H(x,y) = \mathbf{P}(X \leq x, Y \leq y) = \begin{cases} 1, & \text{ha } x \geq 1, y \geq 1, \\ \mathbf{P}(X \leq x), & \text{ha } y \geq 1, \\ \mathbf{P}(Y \leq y), & \text{ha } x \geq 1, \\ \mathbf{P}(X \leq x), & \text{ha } x \leq y, \\ 2xy - y^2, & \text{ha } 0 \leq y \leq x \leq 1, \\ 0, & \text{ha } x \leq 0 \text{ vagy } y \leq 0. \end{cases}$$

Látjuk, hogy kétváltozós eloszlásfüggvény meghatározása nehezebb. A sűrűségfüggvény többváltozós esetben is az eloszlásfüggvény deriváltja, csak most minden változó szerint egyszer kell deriválni. Először x szerint deriválunk, az y-t konstansnak tekintve, majd az eredmény deriváljuk y szerint. Az eloszlásfüggvény alakjából látjuk, hogy csak ott nem 0 a sűrűség, ahol x-től és y-tól is függ az eredmény, azaz

$$h(x,y) = \frac{\partial^2}{\partial x \partial y} = \begin{cases} 2, & \text{ha } 0 \leq y \leq x \leq 1, \\ 0, & \text{különben.} \end{cases}$$

Persze a sűrűség éppen azon a tartomány nem tűnik el, ami az (X, Y) vektor lehetséges értékeit adják, ami most $\{(x, y) : 0 \le y \le x \le 1\}$. Ezek után kiszámolhatjuk a kovarianciát. A formula szerint

$$\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y),$$

és a várható értékeket már ismerjük. A szorzat várható értéke, a várható érték tulajdonságainál megismert formula szerint (szorzat, mint kétváltozós függvény)

$$\mathbf{E}(XY) = \int\int h(x, y)xy\mathrm{d}x\mathrm{d}y = \int_0^1 \left(\int_0^x 2xy\mathrm{d}y\right)\mathrm{d}x = \int_0^1 x \cdot x^2\mathrm{d}x = \frac{1}{4}.$$

Az előbbi integrálásnál figyeljünk a határokra: rögzített $x \in [0, 1]$ esetén $y \in [0, x]$-re lesz $h = 2$! Tehát

$$\mathbf{Cov}(X, Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{36}.$$

□

2.4.9. Legyenek X, Y, Z független, $\mathrm{Geom}(p)$ eloszlású véletlen változók. Adjuk meg a következő valószínűségeket:

(a) $\mathbf{P}(X = Y)$;

(b) $\mathbf{P}(X \ge 2Y)$;

(c) $\mathbf{P}(X + Y \ge Z)$.

2.4.10. Legyen $U = \min\{X, Y\}$ és $V = X - Y$, ahol X, Y független, $\mathrm{Geom}(p)$ véletlen változók. Igazoljuk, hogy U és V függetlenek. [Ez jellemzi is a geometriai eloszlást.]

2.4.11. Legyen az X és Y véletlen változók együttes sűrűségfüggvénye

$$h(x, y) = \begin{cases} \frac{4}{5}(x + xy + y), & \text{ha } (x, y) \in (0, 1)^2\,, \\ 0, & \text{különben}\,. \end{cases}$$

Határozzuk meg a peremeloszlásokat!

2.4.12. Láttuk, hogy abszolút folytonos véletlen vektorváltozó peremeloszlásai abszolút folytonosak. Igazoljuk, hogy ez nem megfordítható, azaz mutassunk X, Y abszolút folytonos véletlen változókat, melyek együttes eloszlása nem abszolút folytonos!

2.4.13. Lássuk be, hogy ha az (X,Y) véletlen vektorváltozó abszolút folytonos, akkor $\mathbf{P}(X=Y)=0$. Az együttes sűrűségfüggvénnyel írjuk fel a $\mathbf{P}(X\le Y)$ valószínűséget!

2.4.14. Legyenek $X_1, X_2, \ldots, X_n$ független véletlen változók F_1, F_2, …, F_n eloszlásfüggvénnyel. Adjuk meg a minimum $m_n = \min\{X_1,\ldots,X_n\}$, és a maximum $M_n = \max\{X_1,\ldots,X_n\}$ eloszlását és együttes eloszlását!

2.4.15. Legyen $f(x,y)=c(x+y)$, $0\le x,y\le 1$, egy (X,Y) vektorváltozó sűrűségfüggvénye. Mennyi c értéke? Adjuk meg a peremeloszlásokat, várható érték vektort, kovarianciamátrixot! Számoljuk ki Xe^Y várható értékét!

2.4.16. Legyen X és Y független Poisson eloszlású véletlen változó λ illetve μ paraméterrel. Határozzuk meg $X+Y$ és XY várható értékét és szórását, valamint a két változó kovarianciáját.

2.4.17. Egy szabályos kockával n-szer dobunk. Jelölje X_1 az egyesek, X_2 a kettesek számát. Határozzuk meg az együttes eloszlásukat! Határozzuk meg X_1, X_2 kovarianciáját és korrelációját!

2.4.18. A Jonas Brothers nevű együttes újra összeáll és koncertet adnak. A PepsiCo cég a következő ötlettel áll elő: a kólásüvegeik kupakjában elrejtik a banda egy-egy tagjának a nevét és azok között, akik összegyűjtik mindhárom nevet kisorsolnak egy VIP belépőt. Kevin neve a kupakok felén szerepel, Joe-val a kupakok egyharmadában találkozhatunk és Nick a legritkább, neve átlagosan minden hatodik kupakban szerepel. Veszünk 10 kólát. Adjuk meg a Joe feliratú kupakok számának várható értékét és szórását! Adjuk meg a Joe feliratú és a Kevin feliratú kupakok számának kovarianciáját és korrelációs együtthatóját!

2.4.19. Legyen az X és Y változók együttes sűrűségfüggvénye

$$f(x,y)=6e^{-2x-3y}\quad (x,y>0)\,,$$

0, különben. Határozzuk meg az együttes és marginális eloszlásfüggvényeket! Adjuk meg a kovarianciamátrixot!

2.4.20. Legyen X és Y együttes sűrűsége h, ahol

(a) $h(x,y)=xe^{-x(1+y)}$, ha $x,y\ge 0$;

(b) $h(x,y)=6xy^2$, ha $x,y\in[0,1]$;

(c) $h(x,y)=2xy+x$, ha $x,y\in(0,1)$;

(d) $h(x,y)=(x+y)^2-(x-y)^2$, ha $x,y\in(0,1)$.

Határozzuk meg a kovarianciamátrixot!

Megoldás. Ez egyszerű számolás. A definíciót kell tudni. Megcsináljuk az (a), (b) részt.

(a) Az X sűrűségét jelölje f, Y-ét g. Ekkor a peremeloszlás sűrűségére vonatkozó formula szerint $x > 0$ esetén

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} h(x,y)\mathrm{d}y = \int_0^{\infty} xe^{-x(1+y)}\mathrm{d}y \\ &= e^{-x}\int_0^{\infty} xe^{-xy}\mathrm{d}y = e^{-x}[-e^{-xy}]_{y=0}^{\infty} = e^{-x}, \end{aligned}$$

és, a parciális integrálás formulája szerint $y > 0$ esetén ($\int f'g = fg - \int fg'$)

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} h(x,y)\mathrm{d}x = \int_0^{\infty} xe^{-x(1+y)}\mathrm{d}x \\ &= \left[x(-1)(1+y)^{-1}e^{-x(1+y)}\right]_{x=0}^{\infty} - \int_0^{\infty}(-1)(1+y)^{-1}e^{-x(1+y)}\mathrm{d}x \\ &= \frac{1}{(1+y)^2}. \end{aligned}$$

Tehát X sűrűsége

$$f(x) = \begin{cases} e^{-x}, & \text{ha } x \ge 0, \\ 0, & \text{különben.} \end{cases},$$

és Y-é

$$g(y) = \begin{cases} (1+y)^2, & \text{ha } y \ge 0, \\ 0, & \text{különben.} \end{cases}$$

A várható értékek

$$\mathbf{E}(X) = \int_0^{\infty} xe^{-x}\mathrm{d}x = 1,$$

és

$$\mathbf{E}(Y) = \int_0^{\infty} y\frac{1}{(1+y)^2}\mathrm{d}y > \int_3^{\infty}\frac{2}{y}\mathrm{d}y = \infty.$$

Itt annyit használtunk, hogy $y/(1+y)^2 > 1/(2y)$ ha $y > 3$, és hogy az $1/y$ integrálja végtelen a $(3,\infty)$ intervallumon. Tehát Y várható értéke nem létezik. Vannak olyan változók, amiknek nincs várható értéke, ezen ne lepődjünk meg. Ekkor a szórás sem létezik, így kovariancia sem.

(b) Egyszerű, a korábbi jelölésekkel, ha $x \in [0,1]$, $y \in [0,1]$,

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} h(x,y)\mathrm{d}y = \int_0^1 6xy^2\mathrm{d}y = 2x, \\ g(y) &= \int_{-\infty}^{\infty} h(x,y)\mathrm{d}x = \int_0^1 6xy^2\mathrm{d}x = 3y^2. \end{aligned}$$

Látjuk, hogy $h(x,y) = f(x)g(y)$, ami azzal ekvivalens, hogy a változóink függetlenek. Ekkor kovarianciájuk 0. Várható értékek és a második momentumok:

$$
\begin{aligned}
\mathbf{E}(X) &= \int_0^1 xf(x)\mathrm{d}x = \int_0^1 2x^2\mathrm{d}x = \frac{2}{3},\\
\mathbf{E}(X^2) &= \int_0^1 x^2f(x)\mathrm{d}x = \int_0^1 2x^3\mathrm{d}x = \frac{1}{2},\\
\mathbf{E}(Y) &= \int_0^1 yg(y)\mathrm{d}y = \int_0^1 3y^3\mathrm{d}y = \frac{3}{4},\\
\mathbf{E}(Y^2) &= \int_0^1 y^2g(y)\mathrm{d}y = \int_0^1 3y^4\mathrm{d}y = \frac{3}{5}.
\end{aligned}
$$

És a szórásnégyzet:

$$
\begin{aligned}
\mathbf{D}^2(X) &= \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18},\\
\mathbf{D}^2(Y) &= \mathbf{E}(Y^2) - (\mathbf{E}(Y))^2 = \frac{3}{4} - \frac{9}{25} = \frac{39}{100}.
\end{aligned}
$$

□

2.4.21. Legyen az (X,Y) véletlen vektor sűrűsége $f(x,y) = 3/x^5$, ha $x \geq y \geq 0$, $x \geq 1$. Adjuk meg a kovarianciamátrixot!

2.4.22. Legyenek X és Y független standard normálisok. Határozzuk meg az $(X-Y, X+Y)$ vektor várhatóérték-vektorát és kovarianciamátrixát!

2.5. Nevezetes eloszlások

2.5.1. Húsvétra dobta piacra a Kinder Meglepetés új, matematikusfigurákat tartalmazó Kinder tojásait. Átlagosan minden 4-edik tojás rejt matematikusfigurát. Aladár 10 Kinder tojást kapott. Adjuk meg annak a valószínűségét, hogy Aladár matematikusfigurának örülhet! Adjuk meg Aladár matematikusfigurái számának eloszlását, várható értékét!

2.5.2. Egy könyvben az egyes oldalakon a sajtóhubák száma egymástól független, Poisson(2) eloszlást követ. Határozzuk meg annak a valószínűségét, hogy a 30. és 31. oldalon sincs hiba! Adjuk meg az ezeken az oldalakon található sajtóhubák várható értékét és szórását!

Megoldás. Először is felidézzük a Poisson-eloszlás definícióját (rosszabb esetben megnézzük a képletgyűjteményben). Eszerint X Poisson-eloszlású $\lambda = 2$ paraméterrel, ha

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

Jelölje X a 30. oldalon, Y a 31. oldalon levő hibák számát. A feladat szerint ezek független, Poisson-eloszlásúak 2 paraméterrel. Az, hogy egyik oldalon sincsen hiba, azt jelenti, hogy $X = 0$ és $Y = 0$. Ennek a valószínűsége, a függetlenség miatt

$$\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X = 0) \cdot \mathbf{P}(Y = 0) = e^{-2} \cdot e^{-2} = e^{-4}.$$

Tudjuk, hogy λ-paraméterű Poisson várható értéke és szórásnégyzete is λ, ezért

$$\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y) = 4,$$

és mivel független változók összegének szórásnégyzete egyenlő a szórásnégyzetek összegével, ezért

$$\mathbf{D}^2(X + Y) = \mathbf{D}^2(X) + \mathbf{D}^2(Y) = 4,$$

azaz $\mathbf{D}(X + Y) = 2$. □

2.5.3. Egy augusztusi éjszakán megfigyelhető csillaghullások száma Poisson–eloszlást követ. Annak a valószínűsége, hogy egy éjszaka egyetlen hullócsillagot sem látunk 0,1. Várhatóan hány hullócsillag figyelhető meg egy éjszaka?

Megoldás. Itt is Poisson-eloszlás van. Most ki kell találnunk az adatokból a paramétert, mert nincs expliciten megadva, mint az előbb. Legyen X a megfigyelt hullócsillagok száma egy este. Ekkor $X \sim$Poisson(λ). Az, hogy nem látunk hullócsillagot azt jelenti, hogy $X = 0$. Azaz, azt tudjuk, hogy

$$\mathbf{P}(X = 0) = 0,1.$$

Na de a formula szerint ez éppen $\lambda^0/0! \cdot e^{-\lambda} = e^{-\lambda}$. Tehát $e^{-\lambda} = 0,1$ ahonnan $\lambda = -\log 0,1 = \log 10$ (itt a $\log = \ln$ a természetes alapú logaritmus). A Poisson paramétere tetszőleges pozitív szám lehet, nem kell, hogy egész legyen! Poisson várható értéke a paramétere, azaz

$$\mathbf{E}(X) = \lambda = \log 10.$$

□

2.5.4. Egy biztosítótársaság felmérte, hogy egy év során egy családi ház 0,0002 valószínűséggel gyullad ki. Mennyi a valószínűsége, hogy 2008-ban egy faluban, ahol 15000 ház van, négynél kevesebb tűz üt ki? (Közelítsünk Poisson–eloszlással!)

Megoldás. Itt is Poisson-eloszlással kell számolni. A pontos eloszlás binomiális $n = 15000$ és $p = 0{,}0002$ paraméterekkel, de a binomiális jól közelíthető Poisson-eloszlással ha p kicsi és n nagy.

Először meg kell határoznunk a λ paramétert. Mivel a tűz valószínűsége egy háznál 0,0002, és a faluban 15000 ház van, ezért az összes tűz várható értéke

$$15000 \cdot 0{,}0002 = 3.$$

Tehát egy olyan Poisson-eloszlás kell nekünk, aminek a várható értéke 3. A várható érték éppen a paraméter, tehát $\lambda = 3$. Ezek szerint annak a valószínűsége, hogy négynél kevesebb tűz üt ki az

$$\begin{aligned}\mathbf{P}(X < 4) &= \mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2) + \mathbf{P}(X = 3)\\ &= \sum_{k=0}^{3} \frac{3^k}{k!} e^{-3} = e^{-3}\left(1 + 3 + \frac{9}{2} + \frac{27}{6}\right)\\ &= e^{-3} 13.\end{aligned}$$

□

2.5.5. Egy szövet 100 méterében átlagosan 5 hiba van. Három méteres darabokra vágnak 300 m hosszú szövetet. Várhatóan hány hibátlan darab lesz?

2.5.6. Legyen $X_n \sim \mathrm{Binom}(n, p_n)$, ahol $np_n \to \lambda > 0$. Határozzuk meg X_n határeloszlását, azaz a $\lim_{n\to\infty} \mathbf{P}\{X_n = k\}$ értékeket!

2.5.7. Valamely növényfajta magjaiból álló mintában a hibás magok száma λ paraméterű Poisson–eloszlású véletlen változó. Minden mintát 3 technikus vizsgál meg egymás után, hogy eltávolítsák a hibás magokat. Az i-edik technikus $p_i < 1$ valószínűséggel veszi észre a hibás magokat; döntései az egyes magokra nézve függetlenek, és az egyes technikusok is egymástól függetlenül döntenek. Határozzuk meg az el nem távolított hibás magok eloszlását! ([2])

Megoldás. Azt nézzük, hogy egy technikus után mi a helyzet. Legyen $p = p_1$. Jelölje X az eredeti mintában a hibás magok számát, és Y a technikus ellenőrzése után megmaradt hibás magok számát. Tudjuk, hogy X Poisson-eloszlású λ paraméterrel, azaz lehetséges értékei $0, 1, \ldots$. Világos, hogy Y-nak is ezek a lehetséges értékei.

Így nehéz a feladat. Az eredeti hibás magok száma is véletlen. Mi lenne, ha tudnánk a hibás magok számát. Tegyük fel, hogy az eredeti hibás magok száma k, azaz $X = k$. Jön az első technikus, minden egyes hibás magot p valószínűséggel vesz észre, azaz $1 - p$ valószínűséggel nem veszi észre. Ez minden hibás magra egymástól független, ezért a bennemaradt hibás magok száma binomiális eloszlású k és $1 - p$ paraméterekkel. Tehát, az $X = k$ feltétel mellett (ez egy kutya közönséges feltételes valószínűség)

$$\mathbf{P}(Y = \ell | X = k) = \binom{k}{\ell}(1-p)^{\ell} p^{k-\ell}, \quad \ell = 0, 1, \ldots, k.$$

Akkor alkalmazzuk a teljes valószínűség tételét az $\{X = 0\}$, $\{X = 1\}$, $\ldots$, teljes eseményrendszerrel. Eszerint

$$\mathbf{P}(Y = \ell) = \sum_{k=0}^{\infty} \mathbf{P}(Y = \ell | X = k)\mathbf{P}(X = k).$$

Írjuk be a Poisson-eloszlás formuláját, és a fent kapott formulát. Kicsit számolunk (persze ha $Y = \ell$ akkor X is legalább ℓ, azaz $\mathbf{P}(Y = \ell | X = k) = 0$, ha $k < \ell$)

$$\begin{aligned}
\mathbf{P}(Y = \ell) &= \sum_{k=0}^{\infty} \mathbf{P}(Y = \ell | X = k)\mathbf{P}(X = k) \\
&= \sum_{k=\ell}^{\infty} \binom{k}{\ell}(1-p)^{\ell} p^{k-\ell} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= (1-p)^{\ell} \lambda^{\ell} e^{-\lambda} \frac{1}{\ell!} \sum_{k=\ell}^{\infty} \frac{1}{(k-\ell)!} (\lambda p)^{k-\ell} \\
&= [\lambda(1-p)]^{\ell} e^{-\lambda} \frac{1}{\ell!} e^{\lambda p} \\
&= \frac{[\lambda(1-p)]^{\ell}}{\ell!} e^{-(1-p)\lambda}.
\end{aligned}$$

Az utolsó előtti egyenlőségnél használtuk, hogy az exponenciális függvény Taylor-sora jelent meg ($e^x = \sum_{k=0}^{\infty} x^k / k!$). Látjuk, hogy ez éppen egy $(1-p)\lambda$ paraméterű Poisson-eloszlás.

Tehát az első technikus után maradt hibás magok száma is Poisson. Azaz a második technikus is Poisson darab hibás magot kap, csak a paraméter $(1 - p_1)\lambda$. A fentiek szerint a második technikus után maradt hibás magok száma is Poisson, a paraméter pedig $(1-p_1)(1-p_2)\lambda$, végül a harmadik után maradt hibás magok száma is Poisson-eloszlású, $(1 - p_1)(1 - p_2)(1 - p_3)\lambda$ paraméterrel. □

2.5.8. Tegyük fel, hogy egy rovar petéinek száma λ-paraméterű Poisson-eloszlást követ, és egy petéből $p \in (0,1)$ valószínűséggel lesz lárva, továbbá a peték egymástól függetlenül fejlődnek lárvává, vagy sem. Mutassuk meg, hogy a lárvák száma λp-paraméterű Poisson-eloszlást követ! ([4])

2.5.9. Egy szabályos kockával N-szer dobunk, ahol $N \sim \text{Poisson}(\lambda)$. Jelölje X_1 az egyesek, X_2 a kettesek számát. Adjuk meg az együttes eloszlást!

2.5.10. Mutassuk meg, hogy két független Poisson–eloszlású véletlen változó összege is Poisson–eloszlású!

Megoldás. Itt először is a Poisson-eloszlás definícióját kell tudni. Az X Poisson-eloszlású $\lambda > 0$ paraméterrel, ha nemnegatív egész értékű, és

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

Legyenek tehát X, Y független Poisson-eloszlású véletlen változók $\lambda > 0$ és $\mu > 0$ paraméterekkel. Ekkor a $Z = X + Y$ változó lehetséges értékei is $0, 1, \ldots$, és ha $Z = n$ akkor $X = k$ és $Y = n - k$ valamilyen k-ra, ahol persze $k = 0, 1, \ldots, n$. Vagyis

$$\mathbf{P}(Z = n) = \mathbf{P}(X + Y = n) = \sum_{k=0}^{n} \mathbf{P}(X = k, Y = n - k).$$

Eddig nem használtunk igazából semmit. Mivel X, Y függetlenek, ezért az utóbbi összeg tagjait szorzattá alakíthatjuk, azaz

$$= \sum_{k=0}^{n} \mathbf{P}(X = k) \cdot \mathbf{P}(Y = n - k).$$

Most használjuk, hogy X és Y is Poisson, és egy kicsit számolunk. Ezért

$$\begin{aligned}
&= \sum_{k=0}^{n} \frac{\lambda^k}{k!}e^{-\lambda} \cdot \frac{\mu^{n-k}}{(n-k)!}e^{-\mu} \\
&= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k} \\
&= \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}.
\end{aligned}$$

Itt a binomiális tételt használtuk, és a binomiális együttható alakját, semmi extra. Tehát azt kaptuk, hogy tetszőleges $n = 0, 1, \ldots$ esetén

$$\mathbf{P}(Z = n) = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)},$$

na de ez éppen a $(\lambda+\mu)$ paraméterű Poisson-eloszlás definíciója. Kész vagyunk. $\square$

2.5.11. Egy urnában 3 piros és 5 fehér golyó van. Visszatevéssel húzunk N-szer, ahol N Poisson-eloszlású véletlen változó 2 paraméterrel. Határozzuk meg a kihúzott piros golyók számának N-re vett feltételes várható értékét! Adjuk meg a kihúzott piros golyók számának eloszlását!

2.5.12. Hipergeometrikus eloszlás. Egy urnában van f fekete és z zöld golyó. Vegyünk egy r elemű véletlen mintát visszatevés nélkül, és jelölje S_r a fekete golyók számát! Határozzuk meg S_r eloszlását, várható értékét és szórását!

2.5.13. Egy villanykörte élettartama exponenciális eloszlású, átlagosan 2 évig működik. Mennyi a valószínűsége, hogy legalább egy évig fog működni egy új villanykörte? Mennyi a valószínűsége, hogy legalább még egy évig fog működni egy már fél éve működő? Mennyi időt él meg a villanykörték 90%-a?

Megoldás. Exponenciális eloszlásunk van, először meg kell határozni a paramétert.

Legyen egy villanykörte évben mért élettartama X. Ekkor a feladat szerint X exponenciális eloszlású és várható értéke 2. Tudjuk (vagy megnézzük a képletgyűjteményben), hogy az exponenciális várható értéke a paraméter reciproka. Azaz, ha $X \sim \mathrm{Exp}(\lambda)$, akkor $\mathbf{E}(X) = 1/\lambda$. Az, hogy egy villanykörte átlagosan 2 évig működik, éppen azt jelenti, hogy az élettartam várható értéke 2. Azaz $2 = 1/\lambda$, ahonnan kapjuk, hogy $\lambda = 1/2$.

Az, hogy egy új villanykörte legalább egy évig jó, azt jelenti, hogy az élettartam nagyobb, mint 1. Tehát a kérdés, hogy mennyi $\mathbf{P}(X > 1)$ (mivel folytonos eloszlásunk van, mindegy, hogy ≥ 1 vagy > 1). Az exponenciális eloszlásfüggvénye $\mathbf{P}(X \leq x) = 1 - e^{-\lambda x}$, ha $x \geq 0$, és 0 különben, ezért

$$\mathbf{P}(X > 1) = 1 - \mathbf{P}(X \leq 1) = e^{-\frac{1}{2}} \approx 0{,}6.$$

A második kérdésnél tudjuk, hogy már fél éve működik az izzónk. Ez egy feltételes valószínűség, hiszen van egy részinformációm, nevezetesen $\{X > 1/2\}$. A kérdés, hogy mennyi a valószínűsége, hogy még legalább 1 évig él, azaz $\{X > 3/2\}$. A feltételes valószínűség definíciója szerint

$$\mathbf{P}(X > 3/2 | X > 1/2) = \frac{\mathbf{P}(X > 3/2, X > 1/2)}{\mathbf{P}(X > 1/2)} = \frac{\mathbf{P}(X > 3/2)}{\mathbf{P}(X > 1/2)} = \frac{e^{-\frac{1}{2}\cdot\frac{3}{2}}}{e^{-\frac{1}{2}\cdot\frac{1}{2}}} = e^{-\frac{1}{2}}.$$

Most lényegében beláttuk, hogy az exponenciális eloszlás örökifjú, ami azt jelenti, hogy tetszőleges $s, t > 0$ esetén $\mathbf{P}(X > s + t | X > s) = \mathbf{P}(X > t)$, azaz ha a változóm már s időt megélt (feltétel!), akkor annak a valószínűsége, hogy még t-t megél az ugyanannyi, mint annak a valószínűsége, hogy t-t megél (nincs feltétel!).

A harmadik kérdésen kicsit el kell mélázni. Az, hogy a időt megél az izzók 90%-a azt jelenti, hogy ha egyszerre becsavarok 100 izzót, akkor a idő után még 90 db világít. Másképpen, 10 égett ki a-ig, azaz annak a valószínűsége, hogy egy izzó élettartam a-nál kisebb, az 0,1. Tehát a kérdés az az a szám, melyre

$$\mathbf{P}(X \leq a) = 0{,}1.$$

(Mivel folytonos eloszlásunk van, ezért mindegy, hogy hol van szigorú egyenlőtlenség, és hol nincs.) Beírva az eloszlásfüggvény képletét ($\lambda = 1/2$, ezt már kiszámoltuk!)

$$0{,}1 = 1 - e^{-\frac{a}{2}},$$

vagyis $a = 2 \log \frac{10}{9} \approx 0{,}21$. Azaz 0,21 évet él meg a villanykörték 90%-a. □

2.5.14. A skót bakák mellkasának körmérete $N(88, 10)$ eloszlást követ. Mekkora hányaduk fér bele 84-es zubbonyba?

Megoldás. Végre egy normális eloszlás. Most mindkét paraméter meg van adva. A várható érték $\mu = 88$, a szórásnégyzet $\sigma^2 = 10$. Arra figyeljünk, hogy a szórásnégyzet a 10, nem a szórás! Azt fogjuk használni, hogy ha X normális eloszlású μ és σ^2 paraméterekkel, akkor $Z = (X - \mu)/\sigma$ standard normális eloszlású, ami meg táblázatba van szedve. Figyeljünk oda, hogy mi σ és mi σ^2!

Legyen X egy skót baka mellkasának körmérete. Erre úgy gondolunk, hogy a skót katonák összességéből véletlenszerűen kiválasztunk egyet. Ezzel a jelöléssel a feladat éppen a $\{X \leq 84\}$ esemény valószínűségét kérdezi. Ez pedig

$$\mathbf{P}(X \leq 84) = \mathbf{P}\left(\frac{X - 88}{\sqrt{10}} \leq \frac{84 - 88}{\sqrt{10}}\right) = \mathbf{P}(Z \leq -1{,}29) = \Phi(-1{,}26).$$

Az standard normális eloszlásfüggvény táblázatában negatív x-ek nincsenek. De nem baj, tudjuk, hogy a sűrűség páros függvény, amiből következik, hogy

$$\Phi(-x) = 1 - \Phi(x), \quad x > 0.$$

Tehát

$$\Phi(-1{,}26) = 1 - \Phi(1{,}26) = 1 - 0{,}8962 = 0{,}1038.$$

Azaz a skót bakák kb. 10%-a fér bele egy 84-es zubbonyba. □

2.5.15. Egy munkadarabokat készítő gép 40 cm-re van beállítva. A hiba normális eloszlást követ 0 várható értékkel. Annak a valószínűsége, hogy egy munkadarab nagyobb, mint 40,5 cm, 0,05. Mennyi a szórás?

Megoldás. Jelölje X egy munkadarab hibáját (előjellel, tehát lehet negatív is), azaz a munkadarab hossza $40+X$ cm. Tudjuk, hogy X normális eloszlású. Most nincs megadva mindkét paramétere a normális eloszlásnak. Azt tudjuk, hogy 0 a várható érték, azaz $\mu = 0$. Legyen σ^2 a szórásnégyzet. Azt tudjuk még, hogy

$$\mathbf{P}(X > 0{,}5) = 0{,}05,$$

hiszen a munkadarab pontosan akkor nagyobb, mint 40,5, ha a hiba nagyobb, mint 0,5. Megint a normális eloszlás skálázási tulajdonságát használjuk. Eszerint $X/\sigma = Z$ standard normális. Tehát

$$\mathbf{P}(X > 0{,}5) = \mathbf{P}(X/\sigma > 0{,}5/\sigma) = \mathbf{P}(Z > 0{,}5/\sigma) = 1 - \Phi(0{,}5/\sigma) = 0{,}05.$$

Innen adódik, hogy

$$\Phi(0{,}5/\sigma) = 0{,}95$$

Most azt kell megkeresni a normális eloszlás táblázatában, hogy hol veszi fel a 0,95 értéket. Ez valahol 1,64 és 1,65 között történik, legyen 1,65. Tehát azt kaptuk, hogy

$$\frac{0{,}5}{\sigma} = 1{,}65,$$

ahonnan

$$\sigma = 0{,}3.$$

Azaz a szórás 0,3. □

2.5.16. Frankenstein professzor vámpír denevéreket tenyészt a laboratóriumában. A denevérek tépőfogainak a hossza normális eloszlást követ $\mu = 28$ mm átlaggal és $\sigma = 4$ mm szórással. Frankenstein tudja, hogy azoknak az állatoknak a harapása halálos, akiknek a tépőfogmérete a populáció felső 5%-ába esik. Számítsuk ki, hogy ez hány mm-es fogméretet jelent!

2.5.17. A házimacskák testsúlya jó közelítéssel normális eloszlást követ. A macskák 10%-a könnyebb, mint 1,5 kg, és 20%-a nehezebb, mint 7 kg. Mekkora a 6 kg-nál nehezebb macskák aránya?

2.5.18. Tegyük fel, hogy Ausztriában a munkavállalók keresete normális eloszlást követ. Tudjuk, hogy a munkavállalók fele keres havi 3000 eurót vagy kevesebbet, míg 5%-uk keres 8000 eurónál többet. Egy törvénytervezet szerint változna az adókulcs az 5000 eurónál többet keresők számára. A munkavállalók mekkora hányadát érinti ez a változtatás?

2.5.19. Egy telefonfülke előtt állunk, és várjuk, hogy az előttünk beszélő befejezze a beszélgetést. Az illető véletlentől függő ideig beszél, az időtartam sűrűségfüggvénye (percben mérve) $e^{-(x/3)}/3$, $x > 0$.

(a) Mennyi a valószínűsége, hogy a beszélgetés 3 percnél tovább tart?

(b) Mennyi a valószínűsége, hogy a beszélgetés $t + 3$ percnél tovább tart, feltéve, hogy t percnél tovább tart?

2.5.20. Anna 30-ik születésnapjára azt a 6 darabos pohárkészletet kapja nagymamájától, mely már 100 éve a család tulajdona. A poharak élettartamai egymástól függetlenek, exponenciális eloszlást követnek 50 év várható értékkel. Adjuk meg annak a valószínűségét, hogy 50 év múlva Anna sértetlenül adhatja tovább unokájának a családi ereklyét (azaz mind a hat poharat)!

2.5.21. A Texpo áruházakban az i-edik kasszánál egy vásárló percben számolva i paraméterű exponenciális időt tölt el. A kiszolgálási idők az egyes kasszáknál egymástól függetlenek.

- András éppen üresen találja az 1-es kasszát. Mennyi a valószínűsége, hogy 2 percen belül végez?
- Andrással pontosan egyidőben Béla beáll az ugyancsak üres 2-es kasszához. Mennyi a valószínűsége, hogy mindketten 2 percen belül végeznek? Mennyi a valószínűsége, hogy Béla 2 percen belül végez, de András nem?
- Mennyi a valószínűsége, hogy valamelyikük 2 percen belül végez? Határozzuk meg a hamarabb végző kiszolgálási idejének eloszlásást! Tehát András és Béla kiszolgálási idejének a minimumára vagyunk kíváncsiak.
- Mennyi a valószínűsége, hogy András végez hamarabb?

Megoldás. (a) Legyen X András kiszolgálási ideje. Ekkor X exponenciális eloszlású 1 paraméterrel, ezért az a valószínűség, hogy András 2 percen belül végez

$$\mathbf{P}(X \leq 2) = 1 - e^{-1\cdot 2} = 1 - e^{-2} = 0,86.$$

(b) Legyen Y Béla kiszolgálási ideje. Ez exponenciális eloszlású 2 paraméterrel, és független X-től. Ezért az a valószínűség, hogy mindketten 2 percen belül végeznek

$$\mathbf{P}(X \leq 2, Y \leq 2) = \mathbf{P}(X \leq 2) \cdot \mathbf{P}(Y \leq 2) = (1 - e^{-2}) \cdot (1 - e^{-4}) = 0,84.$$

Ha Béla 2 percen belül végez, András pedig nem, akkor $Y \leq 2$ és $X > 2$, aminek a valószínűsége

$$\mathbf{P}(X > 2, Y \leq 2) = \mathbf{P}(X > 2) \cdot \mathbf{P}(Y \leq 2) = e^{-2} \cdot (1 - e^{-4}) = 0,13.$$

(c) Jelölje Z az X és Y közül a kisebbet, azaz $Z = \min\{X, Y\}$. Ekkor Z nemnegatív értékeket vesz fel. Valamely $z > 0$ esetén

$$\begin{aligned}\mathbf{P}(Z \leq z) &= 1 - \mathbf{P}(Z > z) = 1 - \mathbf{P}(X > z, Y > z)\\ &= 1 - \mathbf{P}(X > z) \cdot \mathbf{P}(Y > z) = 1 - e^{-z} \cdot e^{-2z}\\ &= 1 - e^{-3z}.\end{aligned}$$

Ez pontosan azt jelenti, hogy Z exponenciális eloszlású 3 paraméterrel.

Lényegében beláttuk (1 és 2 helyett λ és μ-t írva), hogy független exponenciálisok minimuma exponenciális, és a paraméter a paraméterek összege. Ez az egyszerű tény fontos lesz Sztochasztikus modellek kurzuson.

(d) Mivel X és Y függetlenek, ezért együttes sűrűségfüggvényük az egyes sűrűségfüggvények szorzata, azaz

$$h(u, v) = f_X(u) f_Y(v) = \begin{cases} e^{-u} 2e^{-2v}, & \text{ha } u, v > 0,\\ 0, & \text{különben.}\end{cases}$$

Az, hogy András végez hamarabb pontosan azt jelenti, hogy $X < Y$, azaz $(X, Y) \in \{(x, y) : x < y\} = A$. Ennek a valószínűsége

$$\begin{aligned}\mathbf{P}(X < Y) = \mathbf{P}((X, Y) \in A) &= \iint_A h(u, v)\mathrm{d}u\mathrm{d}v\\ &= \int_0^\infty \left(\int_u^\infty 2e^{-u} e^{-2v} \mathrm{d}v \right) \mathrm{d}u\\ &= \int_0^\infty e^{-u} \left[-e^{-2v} \right]_{v=u}^{v=\infty} \mathrm{d}u\\ &= \int_0^\infty e^{-u} e^{-2u} \mathrm{d}u = \frac{1}{3}.\end{aligned}$$

□

2.5.22. Legyenek X és Y független exponenciális eloszlású véletlen változók λ és μ paraméterekkel. Határozzuk meg a minimumuk eloszlását! Mennyi a valószínűsége, hogy Y a kisebb?

2.5.23. Számítsuk ki az (a, b) intervallumon egyenletes eloszlás ferdeségét és lapultságát! Miért nem függ az eredmény a-tól és b-től?

2.5.24. Számítsuk ki a λ paraméterű exponenciális eloszlás ferdeségét és lapultságát!

2.5.25. Számítsuk ki az $\mathrm{N}(\mu, \sigma^2)$ paraméterű normális eloszlás ferdeségét és lapultságát!

2.5.26. Tegyük fel, hogy az X véletlen változó örökifjú, azaz tetszőleges $t, s > 0$ esetén

$$\mathbf{P}\left(X > t + s | X > t\right) = \mathbf{P}\left(X > s\right).$$

Határozzuk meg X eloszlásfüggvényét!

2.5.27. Geometriai eloszlás. Legyen $\mathbf{P}(A) = p \in (0, 1)$. Egy kísérletet addig ismétlünk, míg az A esemény be nem következik. Jelölje X a szükséges ismétlések számát. Adjuk meg X eloszlását, várható értékét, szórását!

2.5.28. Legyen X pozitív egész értékű véletlen változó, melyre teljesül a diszkrét örökifjú tulajdonság, azaz

$$\mathbf{P}\{X > k + l | X > l\} = \mathbf{P}\{X > k\}.$$

Mutassuk meg, hogy $X \sim \mathrm{Geom}(p)$!

2.5.29. Diszkrét örökifjúból folytonosat. Legyen $X_n \sim \mathrm{Geom}(\lambda/n)$. Határozzuk meg X_n/n határeloszlását, azaz adjuk meg a

$$\lim_{n \to \infty} \mathbf{P}\left(\frac{X_n}{n} \leq x\right)$$

határértéket minden $x \in \mathbb{R}$ esetén!

2.5.30. Folytonos örökifjúból diszkrétet. Legyen $X \sim \mathrm{Exp}(\lambda)$ eloszlású véletlen változó. Határozzuk meg $\lfloor X \rfloor$ eloszlását! (A geometriai eloszlás a diszkrét örökifjú.)

2.6. Feltételes eloszlás

2.6.1. Egy szabályos kockával N-szer dobunk, ahol $N \sim \text{Poisson}(\lambda)$. Jelölje X_1 az egyesek, X_2 a kettesek számát. Adjuk meg az együttes eloszlást!

2.6.2. Legyen az (X, Y) véletlen vektorváltozó eloszlása egyenletes az egységkörben. Határozzuk meg az Y feltételes sűrűségfüggvényét az $X = x$ feltétel mellett! Számítsuk ki az $\mathbf{E}[Y^2|X = x]$ feltételes várható értéket. ([2] 2.3.5.)

Megoldás. Jelölje $B = \{(x, y) : x^2 + y^2 \leq 1\}$ a zárt egységkört. Mivel (X, Y) egyenletes eloszlású, ezért az együttes h sűrűségfüggvény

$$h(x, y) = \begin{cases} \frac{1}{\pi}, & (x, y) \in B, \\ 0, & \text{különben.} \end{cases}$$

Innen a peremeloszlások sűrűségei

$$f(x) = \int_{-\infty}^{\infty} h(x, y) \mathrm{d}y = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2}, & x \in [-1, 1], \\ 0, & \text{különben,} \end{cases}$$

és a szimmetria miatt $f(x) = g(x)$. A feltételes sűrűségfüggvény

$$f_{Y|X}(y|x) = \frac{h(x, y)}{f(x)} = \frac{1}{2\sqrt{1 - x^2}}, \quad y \in [-\sqrt{1 - x^2}, \sqrt{1 - x^2}],$$

azaz feltéve, hogy $X = x$ az Y eloszlása egyenletes a $[-\sqrt{1 - x^2}, \sqrt{1 - x^2}]$ intervallumon. Nyilván

$$\mathbf{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) \mathrm{d}y = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y \frac{1}{2\sqrt{1 - x^2}} \mathrm{d}y = 0,$$

és

$$\mathbf{E}[Y^2|X = x] = \int_{-\infty}^{\infty} y^2 f_{Y|X}(y|x) \mathrm{d}y = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 \frac{1}{2\sqrt{1 - x^2}} \mathrm{d}y = \frac{1 - x^2}{3}.$$

□

2.6.3. Legyenek X, Y, Z független exponenciális eloszlású véletlen változók, λ, μ, ν paraméterekkel. Határozzuk meg a $\mathbf{P}(X > Y)$, $\mathbf{P}(X > Y > Z)$ valószínűségeket.

Megoldás. Jelölje f az Y sűrűségfüggvényét. A teljes valószínűség tétele szerint

$$\begin{aligned}
\mathbf{P}(X>Y) &= \int_0^\infty \mathbf{P}(X>y|Y=y)f(y)\mathrm{d}y \\
&= \int_0^\infty e^{-\lambda y}\mu e^{-\mu y}\mathrm{d}y \\
&= \frac{\mu}{\lambda+\mu}\int_0^\infty (\lambda+\mu)e^{-(\lambda+\mu)y}\mathrm{d}y \\
&= \frac{\mu}{\lambda+\mu}.
\end{aligned}$$

Ezt már meghatároztuk korábban, kicsit más számolással.

Hasonlóan

$$\begin{aligned}
\mathbf{P}(X>Y>Z) &= \int_0^\infty \mathbf{P}(X>y>Z|Y=y)f(y)\mathrm{d}y \\
&= \int_0^\infty e^{-\lambda y}(1-e^{-\nu y})\mu e^{-\mu y}\mathrm{d}y \\
&= \frac{\mu}{\lambda+\mu}\int_0^\infty (\lambda+\mu)e^{-(\lambda+\mu)y}\mathrm{d}y \\
&\quad - \frac{\mu}{\lambda+\mu+\nu}\int_0^\infty (\lambda+\mu+\nu)e^{-(\lambda+\mu+\nu)y}\mathrm{d}y \\
&= \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\nu+\mu}.
\end{aligned}$$

□

2.6.4. Legyenek X, Y független 1 paraméterű exponenciális eloszlású véletlen változók. Legyen $U = X \wedge Y$, $V = X \vee Y$. Határozzuk meg a maximum minimumra vett feltételes sűrűségét, és fordítva, azaz adjuk meg a $g_{U|V}(u|v)$, $g_{V|U}(v|u)$ feltételes sűrűségeket! Ismerjünk rá a kapott eloszlásokra!

2.6.5. Legyenek X, Y független azonos eloszlású véletlen változók, f sűrűségfüggvénnyel. Legyen $U = X \wedge Y$, $V = X \vee Y$. Határozzuk meg a maximum minimumra vett feltételes sűrűségét, és fordítva, azaz adjuk meg a $g_{U|V}(u|v)$, $g_{V|U}(v|u)$ feltételes sűrűségeket!

Megoldás. Mivel $U \le V$, így az $f_{U,V}(u,v)$ együttes sűrűség 0, ha $v < u$. Ha $v \ge u$, akkor

$$\begin{aligned}
\mathbf{P}(U \le u, V \le v) &= 2\mathbf{P}(X \le u, Y \in (u,v]) + \mathbf{P}(X \le u, Y \le u) \\
&= 2F(u)(F(v)-F(u)) + F(u)^2 \\
&= F(u)(2F(v)-F(u)),
\end{aligned}$$

ahol F a közös eloszlásfüggvény. Így

$$f_{U,V}(u,v) = \frac{\partial^2}{\partial u \partial v}\mathbf{P}(U \le u, V \le v) = 2f(u)f(v),\ \ u \le v.$$

A marginális sűrűséget innen integrálással, vagy direkt számolással meghatározhatjuk. Kapjuk, hogy

$$f_U(u) = \int_{-\infty}^{\infty} f(u,v)\mathrm{d}v = \int_u^{\infty} 2f(u)f(v)\mathrm{d}v = 2f(u)[1 - F(u)],$$

és

$$f_V(v) = \int_{-\infty}^{\infty} f(u,v)\mathrm{d}u = \int_{-\infty}^{v} 2f(u)f(v)\mathrm{d}u = 2f(v)F(v).$$

A feltételes sűrűségek

$$g_{V|U}(v|u) = \frac{f_{U,V}(u,v)}{f_U(u)} = \frac{f(v)}{1 - F(u)},\ \ v \ge u,$$
$$g_{U|V}(u|v) = \frac{f_{U,V}(u,v)}{f_V(v)} = \frac{f(u)}{F(v)},\ \ u \le v.$$

Vegyük észre, hogy a $g_{U|V}(\cdot|v)$ sűrűség egy olyan F eloszlású változó sűrűsége, amiről feltesszük, hogy v-nél kisebb, a $g_{V|U}(\cdot|u)$ pedig egy olyan F eloszlású változó sűrűsége, amiről feltesszük, hogy u-nál nagyobb. Hát persze, pontosan ezt kellett kapjuk. $\square$

2.6.6. Egyenletes eloszlás szerint választok egy p értéket a $[0,1]$ intervallumon, majd gyártok egy olyan érmét, mely p valószínűséggel ad fejet. Jelölje X annak a dobásnak a sorszámát, mikor először dobok fejet. Adjuk meg X eloszlását és várható értékét. Ugyanez lesz a várható érték, ha egy olyan érmét dobálok, mely $\mathbf{E}(p)$ valószínűséggel ad fejet? *Szűcs Gábor feladata*

Megoldás. Nyilván X lehetséges értékei $1, 2, \ldots$, így

$$\begin{aligned}\mathbf{P}(X = k) &= \int_0^1 \mathbf{P}(X = k | p = x) f(x)\mathrm{d}x \\ &= \int_0^1 (1-x)^{k-1} x \mathrm{d}x = \int_0^1 x^{k-1}(1-x)\mathrm{d}x \\ &= \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}.\end{aligned}$$

A várható érték

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} k\mathbf{P}(X = k) = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty.$$

$\square$

2.6.7. Legyen X egyenletes eloszlású a $(0, 1)$ intervallumon, és $X = x$ esetén legyen Y egyenletes eloszlású a $(0, x)$ intervallumon. Adjuk meg (X, Y) eloszlását, a peremeloszlásokat, várható érték vektort és a kovarianciamátrixot! *Szűcs Gábor feladata*

2.6.8. Legyen λ E(0,1) eloszlású véletlen változó. Legyen az X a $\lambda = \lambda_0$ feltétel mellett $\mathrm{Exp}(\lambda_0)$ eloszlású véletlen változó. Adjuk meg X eloszlásfüggvényét! ([2])

2.7. Vegyes

A következő néhány feladat a kombinatorikában *véletlen módszerként* ismert bizonyítási módszerre mutat példát. Bővebben, lásd Alon és Spencer [1] könyvét.

2.7.1. Legyen $G = (V, E)$ egy egyszerű gráf, és $v_1, v_2, \dots, v_n$ a csúcsok egy sorrendje. Minden csúcsra feldobunk egy szabályos érmét, ha az fej akkor a csúcs az A halmazba, különben a B halmazba kerül. Határozzuk meg az A és B halmazok közt futó élek számának várható értékét!

Megoldás. Legyenek a gráf élei $e_1, \dots, e_k$. Ekkor az A és B között futó élek száma

$$X = \sum_{i=1}^{k} I_i,$$

ahol $I_i = 1$, ha e_i él, 0, különben. Nyilván $I_i = 1$ pontosan akkor, ha az e_i él két végpontja különböző halmazba esik. Tehát

$$\mathbf{E}I_i = \mathbf{P}(I_i = 1) = \frac{1}{2}.$$

Innen

$$\mathbf{E}X = k\mathbf{E}I_1 = \frac{k}{2}.$$

□

2.7.2. Mutassuk meg, hogy tetszőleges hurokélmentes gráfból páros gráfot kaphatunk legfeljebb az élek felének elhagyásával!

2.7.3. A $G = (V, E)$ gráf jó síkra rajzolása egy olyan lerajzolás, ahogy mindenki lerajzol egy gráfot, azaz két él véges sok pontban metszi egymást és egy ponton kettőnél több él nem halad át. Egy gráf metszési száma, $\mathrm{cr}(G)$, a lehető legkevesebb metszést adó jó lerajzolásnál keletkezett metszések száma. Igazoljuk, hogy

$$\operatorname{cr}(G) \geq \frac{e^3}{64v^2},$$

ahol e az élek v a csúcsok száma. (Ajtai, Chvatal, Newborn, Szemerédi)

Útmutatás: Tekintsünk egy véletlen G' részgráfot, melyben minden csúcsot egymástól függetlenül $p \in (0,1)$ valószínűséggel tartunk meg. Jelölje X a G optimális lerajzolásában a G' metszéseinek számát.(Nyilván $X \geq \operatorname{cr}(G')$.) Határozzuk meg az $\mathbf{E}(v(G')), \mathbf{E}(e(G'))$ és $\mathbf{E}(X)$ értékeket, alkalmazzuk az Euler–tételből adódó $\operatorname{cr}(G) \geq e - 3v + 6$ becslést, végül legyen $p = 4v/e$.

2.7.4. Ramsey tételkör. $R(k)$ a legkisebb olyan N, hogy ha egy N csúcsú teljes gráf éleit pirossal és kékkel színezzük, akkor lesz egyszínű k-klikk. Mutassuk meg, hogy $2^{k/2} \leq R(k) \leq 2^{2k}$ (Erdős)!

Segítség. A fölső becslés konstruktív, nem kell hozzá véletlen. Az alsó becsléshez színezzük véletlenül az éleket, és számoljuk ki a k-klikkek számának várható értékét! (Konstruktív bizonyítás az alsó becslésre nem ismert.)

Megoldás. Színezzünk minden élet egymástól függetlenül 1/2 valószínűséggel pirosra, 1/2 valószínűséggel kékre. Válasszunk ki k csúcsot az N-ből. Annak a valószínűsége, hogy ez éppen egy egyszínű k-klikk, $2^{1-\binom{k}{2}}$. Tehát az egyszínű k-klikkek várható értéke $2^{1-\binom{k}{2}}\binom{N}{k}$. Na most, ha ez kisebb, mint 1, akkor szükségképpen van olyan konstrukció, ahol az egyszínű k-klikkek száma 0, azaz ekkor $R(k) > N$. Egyszerű számolás adja, hogy $N = 2^{k/2}$ esetén a várható érték kisebb, mint 1. $\square$

2.7.5. A $[0,1]$ intervallumból függetlenül, egyenletes eloszlás szerint választunk pontokat addig, míg az összeg meghaladja t-t. Jelölje $\alpha(t)$ a szükséges választások várható számát. Határozzuk meg $\alpha(t)$-t, $t \in [0,1]$ esetén.

Bővebben [3].

2.7.6. Bergengóciában felütötte fejét a madárinfluenza legújabb változata, a HπNe. Egy influenzás bergengóc 1/7 valószínűséggel pontosan egy, és 4/7 valószínűséggel pontosan kettő másik bergengócot fertőz meg; három, vagy annál több személyt pedig biztosan nem fertőz meg. Mekkora valószínűséggel terjeszti el a járványt egyetlen beteg? (Azaz mekkora annak a valószínűsége, hogy a vírus sosem tűnik el a bergengóc társadalomból?) Mekkora ugyanez a valószínűség, ha kezdetben 3 beteg van?

2.7.7. Legyen $X \in [0,1]$. Adjunk szükséges és elegendő feltételt arra, hogy

$$(X|X \leq a) \overset{\mathcal{D}}{=} aX, \text{ minden } a \in [0,1] \text{ esetén,}$$

azaz feltéve, hogy $X \le a$, X ugyanolyan eloszlású, mint aX.

Adjunk szükséges és elegendő feltételt arra, hogy

$$(X|X > a) \stackrel{\mathcal{D}}{=} (1-a)X + a, \text{ minden } a \in [0,1] \text{ esetén.}$$

Mutassuk meg, hogy ha X-re teljesül mindkét feltétel, akkor X egyenletes eloszlású $[0,1]$-en!

2.7.8. Legyen $X \in [0,1]$. Mi a szükséges és elegendő feltétele az $I(X \le 1/2)$ és $\min\{X, 1-X\}$ változók függetlenségének?

2.7.9. Legyen Y véletlen változó folytonos F eloszlásfüggvénnyel. Mutassuk meg, hogy $F(Y) \sim \text{Uniform}(0,1)$.

Legyen Y véletlen változó F eloszlásfüggvénnyel (*nem feltétlenül folytonos*). Legyen $V \sim \text{Uniform}(0,1)$, mely független Y-tól. Legyen

$$\tilde{F}(x, V) = F(x-) + V(F(x) - F(x-)).$$

Mutassuk meg, hogy $\tilde{F}(Y, V) \sim \text{Uniform}(0,1)$. Bővebben Rüschendorf [7].

2.7.10. Ketten céllövésben versenyeznek, a két versenyző p_1, ill. p_2 valószínűséggel ér el találatot, $p_1 < p_2$. Az ügyetlenebb kezd, majd felváltva lőnek. Mennyi a valószínűsége, hogy az ügyesebb nyer? Mennyi a játék várható időtartama, ha percenként egyet lőnek? ([2])

2.8. De Moivre–Laplace tétel

2.8.1. Az FC Barcelona passzolási hatékonysága 2019. áprilisában $p = 0{,}85$ (azaz egy passz ekkora valószínűséggel sikeres). Adjuk meg annak a valószínűségét, hogy a Liverpool elleni 525 passzból legalább 460 sikeres.

Megoldás. Ez a legklasszikusabb példa a de Moivre–Laplace-tétel alkalmazására. Jelölje S a sikeres passzok számát. Mivel összesen 525 passz van, és mindegyik passz egymástól függetlenül $0,85$ valószínűséggel sikeres, ezért S eloszlása binomiális $n = 525$ és $p = 0{,}85$ paraméterekkel. A de Moivre–Laplace-tétel szerint

$$\frac{S - np}{\sqrt{np(1-p)}}$$

közelítőleg standard normális eloszlású, azaz

$$\mathbf{P}\left(a \le \frac{S - np}{\sqrt{np(1-p)}} \le b\right) \approx \Phi(b) - \Phi(a).$$

A pontos állítás az, hogy ha $n \to \infty$, akkor az $\approx$ helyett $=$ van. Ha csak felső korlát van, akkkor $a = -\infty$ és $\Phi(-\infty) = 0$, ha csak alsó, akkor $b = \infty$ és $\Phi(\infty) = 1$. A feladat kérdése a $\mathbf{P}(S \geq 460)$ valószínűség. Egyszerű átalakítással elérjük, hogy $(S - np)/\sqrt{np(1-p)}$ jelenjen meg. Az $np = 525 \cdot 0{,}85 \approx 446$ és $\sqrt{np(1-p)} \approx 8{,}2$ értékeket beírva

$$\mathbf{P}(S \geq 460) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} \geq \frac{460 - np}{\sqrt{np(1-p)}}\right) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} \geq 1{,}7\right)$$
$$\approx 1 - \Phi(1{,}7) = 1 - 0{,}955 = 0{,}045.$$

Tehát a keresett valószínűsége 0,045. □

2.8.2. Magyarországon, és mindenütt a világon, több fiúgyermek születik, mint lány. Az újszülöttek 52%-a fiú, 48%-a lány. Nevezzük *lányos napoknak/heteknek* azokat a napokat/heteket, amikor több lány születik, mint fiú.

- Szegeden naponta 9 gyermek születik. Mennyi a pontos valószínűsége, hogy Szegeden egy adott nap lányos nap? Milyen eloszlású az egy héten bekövetkezett lányos napok száma? Várhatóan hány lányos nap van egy héten?
- Budapesten naponta 100 gyermek születik. Mennyi a közelítő (normális közelítés, de Moivre–Laplace-tétel) valószínűsége, hogy Budapesten egy adott nap lányos nap?
- Egész Magyarországon egy héten 2500 gyermek születik. Mennyi a lányos hét bekövetkezésének közelítő valószínűsége? Várhatóan hány lányos hét lesz 2020-ban? Adjuk meg annak a közelítő valószínűségét (Poisson-közelítés), hogy legalább 3 lányos hét lesz 2020-ban.

Megoldás. (a) Egy nap 9 gyermek születik. Annak a valószínűsége, hogy egy gyermek lány 0,48. Ezek az események egymástól függetlenek. Ezért, ha X jelöli a lányok számát, akkor X binomiális eloszlású $n = 9$ és $p = 0,48$ paraméterekkel. Ha X a lányok száma, akkor $9 - X$ a fiúké, és pontosan akkor születik több lány, ha $X \geq 5$. Ennek a valószínűsége

$$\mathbf{P}(X \geq 5) = \sum_{k=5}^{9} \mathbf{P}(X = k) = \sum_{k=5}^{9} \binom{9}{k} (0,48)^k \, (0,52)^{9-k} = 0,45.$$

Tehát annak a valószínűsége, hogy egy nap lányos nap, 0,45. A egyes napokon történt születések egymástól függetlenek, ezért az egy héten bekövetkezett lányos napok száma binomiális eloszlást követ, $n_1 = 7$ és $p_1 = 0,45$

paraméterekkel. A binomiális várható értéke $n_1 \cdot p_1 = 3,15$. Várhatóan 3,15 lányos nap van egy héten.

(b) Jelölje most X_B a Budapesten egy napon született lányok számát. Ekkor X_B binomiális $n = 100$ és $p = 0,48$ paraméterekkel. Akkor lesz több lány, ha $X_B \geq 51$. A de Moivre–Laplace-tétel szerint (most $np = 48$ és $\sqrt{np(1-p)} = 5$)

$$\begin{aligned}\mathbf{P}\left(X_B \geq 51\right) &= \mathbf{P}\left(\frac{X_B - np}{\sqrt{np(1-p)}} \geq \frac{51 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \mathbf{P}(Z > 0,6) = 1 - \Phi(0,6) = 0,27.\end{aligned}$$

Itt Z standard normális véletlen változó. Tehát közelítőleg 27% a lányos nap valószínűsége. A pontos valószínűség

$$\sum_{k=51}^{100} \binom{100}{k} (0,48)^k (0,52)^{100-k} = 0,30.$$

Egy héten várhatóan $0,27 \cdot 7 = 1,89$ lányos nap van.

(c) Ez ugyanaz, mint az előbb csak $n = 2500$ és $p = 0,48$ paraméterekkel. Legyen X_M a Magyarországon egy héten született lányok száma. Akkor van lányos hét, ha $X_M \geq 1251$. Most $np = 1200$, $\sqrt{np(1-p)} = 25$, így

$$\begin{aligned}\mathbf{P}\left(X_M \geq 1251\right) &= \mathbf{P}\left(\frac{X_M - np}{\sqrt{np(1-p)}} \geq \frac{1251 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \mathbf{P}(Z > 2,04) = 1 - \Phi(2,04) = 0,02.\end{aligned}$$

Tehát a magyarországi lányos hét valószínűsége 0,02. Egy évben 52 hét van, így az egy évben levő lányos hetek száma binomiális eloszlású $n = 52$ és $p = 0,02$ paraméterekkel. Mivel p kicsi n pedig nagy, ezt közelíthetjük Poisson-eloszlással. Mivel a várható érték megegyezik a paraméterrel, ezért $\lambda = 52 \cdot 0,02 = 1,04$. Legyen tehát Y Poisson-eloszlású $\lambda = 1,04$ paraméterrel. Annak a valószínűsége, hogy legalább 3 lányos hét lesz egy évben

$$\mathbf{P}(Y \geq 3) = 1 - \mathbf{P}(Y \leq 2) = 1 - \left(e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2} e^{-\lambda}\right) = 0,09.$$

□

2.8.3. Chicago és Los Angeles között két vasútvonal van, melyek mindegyikén egy-egy vonat közlekedik. Mindkét vonat egyidőben indul, lényegében egyformán kényelmes és k személyes. Tegyük fel, hogy 1000 utas egymástól függetlenül $1/2-1/2$ valószínűséggel választ vonatot. Legalább mekkora legyen az ülőhelyek k száma, hogy 0,01-nél kisebb legyen annak a valószínűsége, hogy lesz olyan utas, akinek nem jut ülőhely? ([4, 186.o])

Megoldás. Ez is de Moivre–Laplace. Jelölje S az A társasággal utazók számát. Ekkor a B-vel utazók száma $1000 - S$. Mivel $n = 1000$ ember egymástól függetlenül $1/2 - 1/2$ valószínűséggel dönt A ill. B mellett, ezért S binomiális eloszlású $n = 1000$ és $p = 1/2$ paraméterekkel. Legyen k az ülőhelyek száma (mindkét vonaton). Az, hogy lesz olyan, akinek nem jut hely, azt jelenti, hogy vagy az A társaságnál túl sokan vannak, vagy a B-nél. Azaz, vagy $S > k$ (A vonat betelik) vagy $1000 - S > k$ (B vonat betelik). A kérdés a legkisebb olyan k érték, melyre

$$\mathbf{P}(S > k \text{ vagy } S < 1000 - k) \leq 0{,}01.$$

Vegyük észre, hogy $k \geq 500$ kell legyen, és ekkor csak az egyik vonat telhet be, azaz a fenti valószínűségben szereplő két esemény egymást kizáró, ahonnan

$$\mathbf{P}(S > k \text{ vagy } S < 1000 - k) = \mathbf{P}(S > k) + \mathbf{P}(S < 1000 - k).$$

Használjuk a de Moivre–Laplace-tételt. Beírjuk az $np = 500$, $\sqrt{np(1-p)} = 5\sqrt{10} \approx 15{,}8$ értékeket. Ekkor

$$\mathbf{P}(S > k) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} > \frac{k - np}{\sqrt{np(1-p}}\right) \approx 1 - \Phi((k-500)/15{,}8),$$

és ugyanígy (csak felhasználjuk, hogy $\Phi(-x) = 1 - \Phi(x)$ ha $x > 0$, és hogy $500 - k \leq 0$)

$$\begin{aligned}\mathbf{P}(S < 1000 - k) = \mathbf{P}\left(\frac{S - np}{\sqrt{np(1-p)}} > \frac{1000 - k - np}{\sqrt{np(1-p}}\right)\\ \approx 1 - \Phi((k-500)/15{,}8).\end{aligned}$$

Azaz, keressük azt a legkisebb k értéket melyre

$$2\left[1 - \Phi((k-500)/15{,}8)\right] \leq 0{,}01.$$

Átrendezve,

$$\Phi((k-500)/15{,}8) \geq 0{,}995.$$

Kikeressük a normális eloszlás táblázatából, hogy hol veszi fel Φ a 0,995 értéket. Ez 2,57, azaz

$$\frac{k-500}{15{,}8} \geq 2{,}57,$$

amit átrendezve

$$k \geq 540{,}6.$$

Mivel k egész, ezért 541 a legkisebb olyan k, amire a feltétel teljesül. $\square$

2.8.4. Budapesten meg akarják állapítani a dohányosok p arányát. Ehhez kiválasztanak n egyént úgy, hogy minden választásnál mindenki ugyanakkora valószínűséggel kerül kiválasztásra, és csak ezek közt nézik meg a dohányosok k számát. Legalább mekkora legyen az n, hogy a kapott $p' = k/n$ arány legalább 0,95 valószínűséggel legfeljebb 0,005 hibával közelítse a valódi p arányt, akármi is $p \in (0,1)$? ([4, 187.o])

Megoldás. Ez már érdekesebb feladat, ugyanis semmi nincs megadva. Világos, hogy a feladat nagyon fontos, ugyanis a közvéleménykutatásokhoz pontosan ilyen típusú kérdést kell feltenni. Van valami ismeretlen valószínűség p, ami azt mutatja meg, hogy az emberek ilyen aránya szavazna az A párt jelöltjére. Nem tudjuk mi a p, de erről szeretnénk valamit mondani. Hány embert kell megkérdezni, hogy valami okosat mondhassunk?

Jelölje S a dohányzók számát a megkérdezettek között. Ekkor S binomiális eloszlású véletlen változó n (meghatározandó, de ismert) és p (ismeretlen) paraméterekkel. Világos, hogy az ismeretlen p értékre az S/n becslést adjuk. Elég összetett a kérdés, kicsit el kell rajta gondolkodni. A becslés hibája $|S/n - p|$. Azt akarjuk, hogy ez nagy valószínűséggel (0,95) kicsi legyen (0,005-nél kisebb), azaz olyan n értéket keresünk, amire

$$\mathbf{P}\left(\left|\frac{S}{n} - p\right| < 0{,}005\right) \geq 0{,}95.$$

(Annak a valószínűsége, hogy a hiba 0,005-nél kisebb, legalább 0,95.) A neheze megvan. Használjuk a de Moivre–Laplace-tételt. Eszerint

$$\mathbf{P}\left(a \leq \frac{S-np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

Tehát be kell erőltetni az $(S-np)/\sqrt{np(1-p)}$ kifejezést. Tegyük meg:

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{S}{n}-p\right|<0{,}005\right) &= \mathbf{P}\left(\left|\frac{S-np}{n}\right|<0{,}005\right)\\
&= \mathbf{P}\left(\left|\frac{S-np}{\sqrt{np(1-p)}}\right|<0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)\\
&\approx \Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)-\Phi\left(-0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)\\
&= 2\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)-1.
\end{aligned}$$

Annyit használtunk, hogy $|x|\le a$ pontosan akkor, ha $-a<x<a$, és hogy $\Phi(-x)=1-\Phi(x)$. Ezek szerint az kell, hogy

$$2\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)-1\ge 0{,}95,$$

azaz

$$\Phi\left(0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)\ge 0{,}975.$$

A táblázatból kikeresve azt kapjuk, hogy

$$0{,}005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\ge 1{,}96.$$

Átrendezve

$$n\ge 392^2\cdot p(1-p). \tag{$*$}$$

Na de nem ismerjük p értékét. Úgy kell n-et választani, hogy a fenti egyenlőtlenség minden p-re igaz legyen. Tehát válasszuk p-t úgy, hogy a jobb oldal maximális legyen. Ez $p=1/2$-nél van, értéke $1/4$. Tehát, ha

$$n\ge 396^2\frac{1}{4}=38416,$$

akkor $(*)$ teljesül minden $p\in[0,1]$ esetén. □

2.8.5. Egy szabályos dobókockát feldobunk 200-szor. Jelölje S_n a dobott hatosok számát. Adjuk meg pontosan, majd a de Moivre– Laplace tétellel közelítve a $\mathbf{P}(30<S_n\le 40)$ valószínűséget!

2.8.6. Egy szabályos érmét n-szer földobunk. Adjuk meg a

$$\mathbf{P}\left\{\frac{n}{2} - c\sqrt{n} < S_n < \frac{n}{2} + c\sqrt{n}\right\}$$

valószínűségek közelítő értékét! Mit kapunk a $c \to 0$ ill. $c \to \infty$ esetben?

2.8.7. Egy étteremben kétféle menü közül lehet választani. A vendégek 5/6 valószínűséggel A menüt, 1/6 valószínűséggel B menüt választanak. Egy adott napon 500 vendég érkezik. A vendéglős 420 A és 100 B menüt készített elő. Feltételezve, hogy a vendégek egymástól függetlenül választanak, mi a valószínűsége, hogy mindenkinek jut olyan menü, amilyet kér?

2.8.8. Egy szerencsejátékon a nyerési esélyed 1/11. Ha nyersz, visszakapod a feltett tétet és még nyereményként annak kilencszeresét. Elegendő sok kezdőtőkével indulva ezer alkalommal felteszel 1–1 petákot. Mi a valószínűsége, hogy ezer játszma után még legalább annyi pénzed van, mint kezdetben volt?

2.8.9. Az utóbbi években felmerült a mozilátogatókban az igény arra, hogy eredeti hanggal feliratosan is megnézhessék a filmeket. Egy friss felmérés szerint az emberek 2/7-e választaná a feliratos filmet a szinkronizált változattal szemben. A szegedi Cinema City vezetősége minket kért fel arra, hogy segítsünk dönteni Christopher Nolan The Dark Knight Rises című filmjének premierje kapcsán. A premierre 1000 látogatót várnak, és a filmet 9 egyenként 130 fős teremben vetítik. Hány teremben kell feliratosan vetíteni a filmet, hogy a lehető legnagyobb valószínűséggel tudjon minden látogató arra a változatra beülni, amire szeretne? Mekkora ez a valószínűség?

2.8.10. Egy általános iskolában egy és két forintosok gyűjtését hirdetik meg a pénzérmék bevonása előtti fél évben. Megkérik az oda járó diákokat és szüleiket, hogy az otthoni felesleges apórópénzüket az iskolának adják, hogy az így befolyt összegből játszóteret építhessenek az iskolaudvaron. A játszótér megépítéséhez 1,5 millió Ft-ra van szükségük. A gyűjtés során egymillió darab pénzérmét adományoztak az iskolának. Ha ezen pénzérmék mindegyike a többitől függetlenül $1/2 - 1/2$ valószínűséggel egy illetve két forintos, akkor mennyi a közelítő valószínűsége, hogy az igazgatónak legfeljebb 1000 Ft-tal kell hozzájárulnia a játszótér megépüléséhez?

2.8.11. A héten jelenik meg George R. R. Martin új könyve Winds of Winter címmel. A könyvesboltunkba rendeltünk a keményborítós kiadásból 180-et a puhakötésűből pedig 240-et. Tapasztalataink alapján az emberek 40%-a választja a tartósabb, de valamivel drágább keményborítású kiadást. Ha 400 vevőre számítunk, akik egymástól függetlenül döntenek, akkor milyen valószínűséggel tudunk mindenkit kiszolgálni?

2.8.12. Kávézót szeretnénk nyitni egy város forgalmas utcáján. A közelben van egy kávézó, ahol naponta 300 ember megfordul. Mi ezen emberek egy részét szeretnénk elcsábítani, hogy hozzánk térjenek be. Az eltérő hangulat, árak és üzletstratégia alapján úgy gondoljuk, hogy az emberek 27%-át sikerül erre rávennünk. Hány férőhelyesre tervezzük a kávézónkat, ha azt szeretnénk, hogy 90%-os biztonsággal minden betérőnek jusson szabad hely?

3. Statisztika

3.1. Alapstatisztikák, pontbecslések

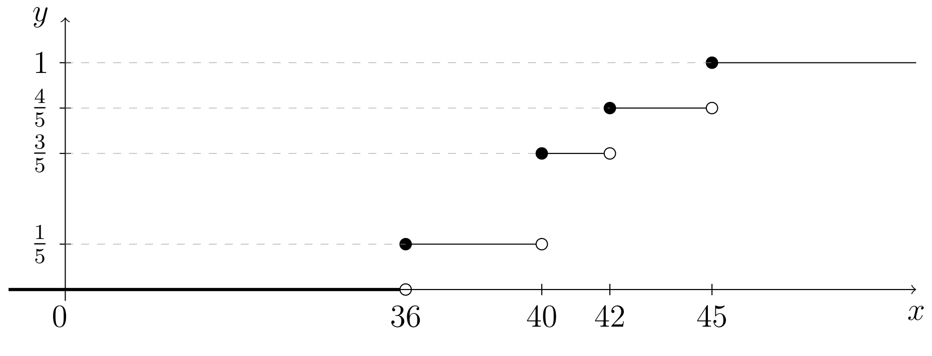
3.1.1. Egy játék fizikai terhelhetőségére elvégzett tesztek kg-ban a következő eredményeket adták: 40, 45, 40, 42, 36. A minta alapján adjuk meg a terhelhetőség empirikus eloszlásfüggvényét, mintaátlagát, korrigált/korrigálatlan empirikus szórásnégyzetét és a mediánt!

Megoldás. Ez egy $n = 5$ elemű minta, ahol a mintaelemek $x_1 = 40$, $x_2 = 45$, $x_3 = 40$, $x_4 = 42$, $x_5 = 36$. A jelölés kicsit szokatlan. A véletlen mintát mindig nagy X-ekkel jelöljük, ugyanakkor egy konkrét mintát ami a kezünkben van, a véletlen minta egy *realizációját* kis x-ekkel jelöljük.

A definíciókat kell tudni. Az empirikus eloszlásfüggvény

$$F_n(x) = \frac{1}{n} \left|\{i : X_i \leq x\}\right|,$$

azaz rögzített x-re megszámoljuk, hogy hány x-nél kisebb, vagy vele egyenlő mintaelemünk van, és ezt elosztjuk a minta elemszámával. Vagyis az empirikus eloszlásfüggvény a mintaelemekben ugrik $1/n$-et (ha több mintaelem egyenlő, akkor annyiszor $1/n$-et, ahányszor az adott értéket felveszi). Most 5 elemű a minta, azaz $n = 5$. A legkisebb mintaelem 36, addig függvény 0, 36-ban pedig $1/5$-öt ugrik. A 40-et kétszer veszi fel, így ott $2/5$-öt ugrik, vagyis $3/5$-re ugrik föl, stb. Tehát így néz ki az empirikus eloszlásfüggvény:



Formulával,

$$F_5(x) = \begin{cases} 0, & \text{ha } x < 36, \\ \frac{1}{5}, & \text{ha } x \in [36, 40), \\ \frac{3}{5}, & \text{ha } x \in [40, 42), \\ \frac{4}{5}, & \text{ha } x \in [42, 45), \\ 1, & \text{ha } x \geq 45. \end{cases}$$

A mintaátlag az

$$\overline{x}_5 = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(40 + 45 + 40 + 42 + 36) = 40,6$$

az empirikus szórásnégyzet

$$s_5^2 = \frac{1}{5}\sum_{i=1}^{5}(x_i - \overline{x}_5)^2 = 8,64,$$

míg a korrigált empirikus szórásnégyzet

$$z_5^2 = \frac{1}{4}\sum_{i=1}^{5}(x_i - \overline{x}_5)^2 = 10,8.$$

A medián a középső mintaelem. Most 5 mintaelem van, ennek tényleg van közepe, így a medián értéke 40. Ha páros a mintaelemszám, akkor a középső két elemnek kell a számtani közepét venni. □

3.1.2. Legyenek $X_1, \dots, X_n$ független azonos eloszlású véletlen változók véges μ várható értékkel és $\sigma > 0$ szórással. Igazoljuk, hogy $\hat{\mu}_1 = X_1$ torzítatlan becslése a várható értéknek! Igazoljuk, hogy $\hat{\mu}_2 = (X_1 + X_2)/2$ is torzítatlan becslése a várható értéknek. Melyik a hatásosabb?

Megoldás. Az, hogy a becslés torzítatlan azt jelenti, hogy a várható értéke megegyezik a becsült paraméterrel. Ez teljesül, hiszen

$$\mathbf{E}(\hat{\mu}_1) = \mathbf{E}(X_1) = \mu,$$

és

$$\mathbf{E}(\hat{\mu}_2) = \mathbf{E}\left(\frac{X_1 + X_2}{2}\right) = \mu.$$

Azaz mindkét becslés torzítatlan. A hatásossághoz szórásnégyzetet kell számolni. Ez

$$\mathbf{D}^2(\hat{\mu}_1) = \mathbf{D}^2(X_1) = \sigma^2,$$

és
$$\mathbf{D}^2\left(\frac{X_1+X_2}{2}\right)=\frac{1}{4}\mathbf{D}^2\left(X_1+X_2\right)=\frac{\sigma^2}{2}.$$
Azaz $\hat{\mu}_2$ szórásnégyzete kisebb, ez a hatásosabb becslés.

Az előadásjegyzetben láttuk, hogy a mintaátlag a leghatásosabb lineáris becslése a várható értéknek. Ezen nagyon nem lepődünk meg. □

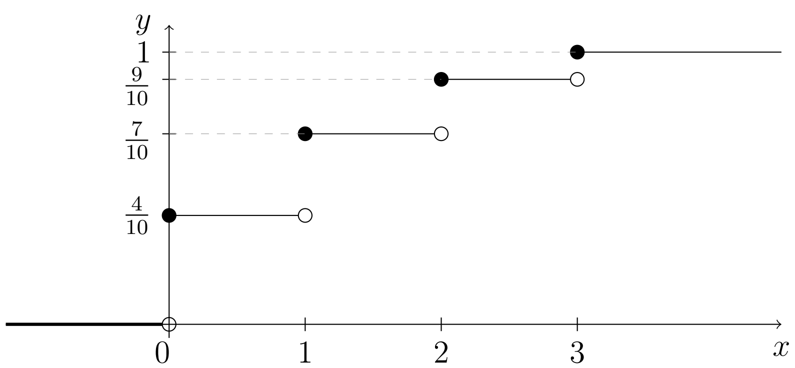
3.1.3. Egy alkatrészekből álló sokaság 6 mintapéldányának következő volt a hónapokban mért teljes élettartama: 39, 45, 67, 50, 50, 60. Adjuk meg az empirikus eloszlásfüggvényt, a mintaátlagot, és a korrigált/korrigálatlan empirikus szórásnégyzetet!

3.1.4. Egy almáskertben véletlenszerűen, egymástól függetlenül találhatók fertőzött fák. Tíz egyforma nagy, egyenként három sorból álló ültetvényben rendre 0, 3, 0, 1, 0, 0, 2, 1, 1, 2 beteg fát találtak.

(a) Adjuk meg az empirikus eloszlásfüggvényt, a mintaátlagot és az empirikus szórásnégyzetet!

(b) Tegyük fel, hogy a beteg fák száma Poisson-eloszlást követ. Adjunk maximum likelihood becslést és momentumbecslést az egy sorban található fák számának várható értékére!

Megoldás. Az (a) részhez a definíciókat kell tudni. Az empirikus eloszlásfüggvény
$$F_{10}(x)=\begin{cases}0, & \text{ha } x<0,\\ \frac{4}{10}, & \text{ha } x\in[0,1),\\ \frac{7}{10}, & \text{ha } x\in[1,2),\\ \frac{9}{10}, & \text{ha } x\in[2,3),\\ 1, & \text{ha } x\geq 1.\end{cases}$$



A mintaátlag

$$\overline{x}_{10} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1,$$

és a korrigálatlan és korrigált empirikus szórásnégyzet

$$s_{10}^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = 1$$

$$z_{10}^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = \frac{10}{9}.$$

(b) Tegyük fel, hogy az adatok Poisson-eloszlásból jönnek, ahol a paraméter $\lambda > 0$ ismeretlen. A megadott realizáció valószínűsége, a függetlenség és a Poisson-eloszlás definícióját felhasználva

$$\begin{aligned}
&\mathbf{P}(X_1 = 0, X_2 = 3, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 0,\\
&\qquad X_7 = 2, X_8 = 1, X_9 = 1, X_{10} = 2)\\
&= \prod_{i=1}^{10}\mathbf{P}(X_i = x_i) = \prod_{i=1}^{10}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\\
&= \left(\prod_{i=1}^{10}\frac{1}{x_i!}\right)\lambda^{\sum_i x_i}e^{-10\,\lambda}\\
&= \left(\prod_{i=1}^{10}\frac{1}{x_i!}\right)\lambda^{10}e^{-10\,\lambda}.
\end{aligned}$$

Ez annak a valószínűsége, hogy ha adott λ-paraméterű Poisson-eloszlásból jönnek az adataink, akkor a konkrét realizáció bekövetkezik. Fordítsuk meg! Tudjuk, hogy a konkrét realizáció bekövetkezett (persze, megszámoltuk a fákat). Melyik λ paraméter a legvalószínűbb? Az, amelyre a fenti valószínűség a legnagyobb. Ez a *maximum likelihood* módszer. Tehát a

$$\lambda^{10}e^{-10\lambda}$$

kifejezés maximumát kell megtalálnunk, ahol $\lambda > 0$. Ez pont ott maximális, ahol a logaritmusa maximális, azaz a

$$\ell(\lambda) = 10\log\lambda - 10\lambda$$

függvény. Ezt lederiválva

$$\ell'(\lambda) = \frac{10}{\lambda} - 10.$$

Látjuk, hogy a függvény monoton nő $\lambda = 1$-ig, ott maximuma van, utána csökken. Tehát a paraméter ML becslése

$$\widehat{\lambda}_{ML} = 1.$$

Ez persze éppen a $\overline{x}_{10}$ mintaátlag. A levezetés ugyanaz, mint az általános esetben az előadásjegyzetben.

A momentumbecslésnél azt a paraméter választjuk, amelyhez tartozó elméleti várható érték megegyezik a mintaátlaggal. Na de a Poisson várható értéke éppen λ, tehát a momentumbecslés éppen a mintaátlag lesz,

$$\widehat{\lambda}_m = 1.$$

A momentumbecslés és a ML becslés sok ismert eloszlás esetében megegyezik. Ilyen a Poisson, exponenciális, normális. Az egyenletes eloszlásnál a kettő nem ugyanaz. □

3.1.5. Egy adott típusú izzó élettartamára öt mérés alapján a következő adataink vannak: 2,3, 4, 1,7, 3,2, 2,8.

(a) Adjuk meg az empirikus eloszlásfüggvényt, a mintaátlagot és az empirikus szórásnégyzetet!

(b) Tegyük fel, hogy a háttéreloszlás exponenciális ismeretlen paraméterrel. Adjunk ML és momentumbecslést a paraméterre!

3.1.6. Egy adatszerverre a lekérdezések exponenciális időközönként érkeznek, ahol ismeretlen paraméterrel. Az időközökre percben mérve a következő adatokat kaptuk: 1,94, 0,33, 2,51, 5,27, 1,73, és 0,61. Adjunk becslést a paraméterre a maximum likelihood és a momentumbecslés alkalmazásával!

Megoldás. Folytonos eloszlás esetén a maximum likelihood becslésnél a sűrűségfüggvények szorzatát kell venni, és azt maximalizálni az adott realizációra. A λ-paraméterű exponenciális sűrűségfüggvénye $\lambda e^{-\lambda x}$. A likelihood függvény ezek szorzata, mint λ függvénye a realizációt beírva az x-ek helyére, azaz ha $x_1, \ldots, x_n$ az adott realizáció, akkor

$$L_{x_1,\ldots,x_n}(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

Ezt kell maximalizálni λ-ban. Véve a logaritmust (ez azért jó, mert a logaritmus szigorúan monoton függvény, ezért a maximumhely nem változik, viszont a nagy szorzatból nagy összeget csinál, amit könnyebb deriválni),

$$\ell(\lambda) = \log L_{x_1,\ldots,x_n}(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i.$$

Deriváljuk

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

Ez pozitív, ha $\lambda < 1/\overline{x}_n$, és negatív ha $\lambda > 1/\overline{x}_n$, tehát

$$\hat{\lambda}_{ML} = \frac{1}{\overline{x}_n} = \frac{n}{\sum_{i=1}^{n} x_i}$$

maximumhely. Ez a paraméter maximum likelihood becslése.

A momentumbecslés egyszerűbb. Egy λ-paraméterű exponenciális eloszlás várható értéke $1/\lambda$. A momentumbecslésnél azt a paramétert választjuk, amihez tartozó elméleti várható érték megegyezik a mintaátlaggal. Azaz keressük az a λ paramétert, melyhez tartozó várható érték, ami $1/\lambda$ éppen az $\overline{x}_n$ mintaátlag. Vagyis

$$\hat{\lambda}_m = \frac{n}{\sum_{i=1}^{n} x_i}.$$

Ez ugyanaz, mint a ML becslés. Most ez

$$\hat{\lambda}_{ML} = \hat{\lambda}_m = 2,065.$$

□

3.1.7. Legyenek $X_1, \dots, X_n$ független, Egyenletes$(0, \theta)$ eloszlású véletlen változók, ahol $\theta > 0$. Határozzuk meg $X_{n,n} = \max\{X_1, \dots, X_n\}$ eloszlás-, és sűrűségfüggvényét, várható értékét és szórását!

Mutassuk meg, hogy

$$T_1(\mathbf{X}) = \frac{n+1}{2n} X_{n,n}$$

torzítatlan becslése $(\theta/2)$-nek, ami hatásosabb, mint a $\overline{X}_n$ mintaátlag! Igazoljuk, hogy T_1 gyengén konzisztens!

Megoldás. A feladat első feléhez semmi statisztika nem kell. Világos, hogy $X_{n,n}$ a $[0, \theta]$ intervallumba esik, ezért

$$F_n(x) = \mathbf{P}(X_{n,n} \le x) = \begin{cases} 0, & \text{ha } x \le 0, \\ 1, & \text{ha } x \ge \theta, \end{cases}$$

vagyis a $[0, \theta]$-n érdekes a dolog. Legyen tehát $x \in [0, \theta]$. Ekkor $X_{n,n} \le x$ pontosan akkor teljesül, ha $X_i \le x$ minden $i = 1, 2, \dots, n$ esetén. Hát persze, a maximum akkor $\leq x$ ha minden elem $\leq x$. Mivel X_i-k függetlenek, és egyenletes eloszlásúak $[0,\theta]$-n ezért

$$\begin{aligned}\mathbf{P}(X_{n,n} \leq x) &= \mathbf{P}(X_1 \leq x, \dots, X_n \leq x) = \mathbf{P}(X_1 \leq x) \dots \mathbf{P}(X_n \leq x) \\ &= (\mathbf{P}(X_1 \leq x))^n = \left(\frac{x}{\theta}\right)^n.\end{aligned}$$

Innen a sűrűségfüggvény

$$f_n(x) = F_n'(x) = \begin{cases} nx^{n-1}\theta^{-n}, & \text{ha } x \in (0,\theta), \\ 0, \text{különben.} \end{cases}$$

Emlékszünk, hogy a sűrűségfüggvény értéke egy-egy pontban (egészen pontosan megszámlálható sok pontban) mindegy, hogy micsoda.

Várható értéket és szórást a definíció szerint számolunk. Így

$$\mathbf{E}(X_{n,n}) = \int_{-\infty}^{\infty} x f_n(x)\mathrm{d}x = \int_0^{\theta} nx^n\theta^{-n}\mathrm{d}x = \frac{n}{n+1}\theta^{n+1}\theta^{-n} = \frac{n}{n+1}\theta.$$

Érdemes meggondolni, hogy mit kaptunk. Egy θ hosszúságú intervallumot felosztottunk n véletlen egyenletes eloszlás szerint választott ponttal. Így $n+1$ intervallumot kapunk, és intuitíven világos, hogy nagyjából egyforma hosszúak. Ekkor minden intervallum hossza várhatóan $\theta/(n+1)$. Ez pedig azt jelenti, hogy a legnagyobb pont várhatóan $\theta n/(n+1)$.

A szóráshoz kell a második momentum, ami

$$\mathbf{E}(X_{n,n}^2) = \int_{-\infty}^{\infty} x^2 f_n(x)\mathrm{d}x = \int_0^{\theta} nx^{n+1}\theta^{-n}\mathrm{d}x = \frac{n}{n+2}\theta^2,$$

így a szórásnégyzet, némi számolás után

$$\begin{aligned}\mathbf{D}^2(X_{n,n}) &= \mathbf{E}(X_{n,n}^2) - (\mathbf{E}(X_{n,n}))^2 \\ &= \theta^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) \\ &= \theta^2\frac{n}{(n+2)(n+1)^2} \sim \frac{\theta^2}{n^2}.\end{aligned}$$

Most rátérünk a statisztika részre. Az, hogy T_1 torzítatlan becslése $\theta/2$-nek, az pontosan azt jelenti, hogy $\mathbf{E}(T_1) = \theta/2$. (Azért $\theta/2$-t becsüljük, mert az az egyenletes eloszlás várható értéke. És láttuk előadáson (nagyon egyszerű), hogy a mintaátlag *mindig* torzítatlan becslése a várható értéknek. Persze becsülhetnénk a θ-t is.) Na de az előbb kiszámoltak alapján

$$\mathbf{E}(T_1) = \mathbf{E}\left(\frac{n+1}{2n}X_{n,n}\right) = \frac{n+1}{2n}\frac{n}{n+1}\theta = \frac{\theta}{2},$$

azaz a becslés tényleg torzítatlan.

A torzítatlanság önmagában nem sokat jelent, hiszen egyetlen mintaelem is torzítatlan becslése a várható értéknek, mégsem jó. Azt, hogy a becslés milyen közel van becsült paraméterhez éppen a szórás jellemzi, ha a becsült paraméter éppen a várható érték, azaz a becslés torzítatlan. Tehát egy torzítatlan becslés annál jobb, minél kisebb a szórása. Két torzítatlan becslés esetén az egyik akkor hatásosabb, mint a másik, ha kisebb a szórása (egészen pontosan minden paraméter esetén kisebb, vagy egyenlő, és legalább egy paraméter esetén szigorúan kisebb). Az előbbiek szerint

$$\mathbf{D}^2(T_1) = \frac{(n+1)^2}{4n^2}\mathbf{D}^2(X_{n,n}) = \frac{(n+1)^2}{4n^2}\frac{\theta^2 n}{(n+2)(n+1)^2} = \frac{\theta^2}{4n(n+2)}.$$

A mintaátlag szórásnégyzete

$$\mathbf{D}^2(\overline{X}_n) = \frac{1}{n^2}\mathbf{D}^2(\sum_{i=1}^{n} X_i) = \frac{1}{n}\mathbf{D}^2(X_1) = \frac{\theta^2}{12\,n}.$$

Látjuk, hogy ha $n \geq 2$ akkor

$$\mathbf{D}^2(T_1) < \mathbf{D}^2(\overline{X}_n).$$

Igazából egy $1/n$-es szorzó a különbség, vagyis a T_1 tényleg sokkal jobb becslés.

Végül lássuk a gyenge konzisztencia bizonyítását! Ez azt jelenti, hogy tetszőleges $\varepsilon > 0$ esetén annak a valószínűsége, hogy a becslés a valódi paramétertől ε-nál többel tér el, 0-hoz tart, amint a mintaelemszám tart végtelenbe, formulával

$$\lim_{n\to\infty} \mathbf{P}\left(\left|T_1(\mathbf{X}) - \frac{\theta}{2}\right| > \varepsilon\right) = 0.$$

Vegyük észre, hogy $T_1(\mathbf{X}) = T_1(\mathbf{X}_n)$ függ n-től, csak eddig nem jelöltük a függést. Nagyjából világos a dolog, hiszen ha egyre több mintaelemet veszünk, akkor a legnagyobb mintaelem egyre közelebb kerül az intervallum jobboldali végpontjához, θ-hoz, ekkor pedig T_1 közel kerül $\theta/2$-höz. Valóban, beírva a T_1 definícióját

$$\begin{aligned}&\mathbf{P}\left(\left|T_1(\mathbf{X}) - \frac{\theta}{2}\right| > \varepsilon\right)\\ &= \mathbf{P}\left(\frac{n+1}{n}X_{n,n} > \theta + 2\varepsilon\right) + \mathbf{P}\left(\frac{n+1}{n}X_{n,n} < \theta - 2\varepsilon\right).\end{aligned}$$

Mivel $X_{n,n} \leq \theta$ ezért a jobb oldalon első valószínűség 0, ha $n > \theta/(2\varepsilon)$. A második pedig, beírva $X_{n,n}$ eloszlásfüggvényére kapott formulát

$$\mathbf{P}\left(\frac{n+1}{n}X_{n,n} < \theta - 2\varepsilon\right) = \mathbf{P}\left(X_{n,n} \leq \frac{n}{n+1}(\theta - 2\varepsilon)\right)$$
$$= \left(\frac{n}{n+1}\left(1 - \frac{2\varepsilon}{\theta}\right)\right)^n \to 0,$$

hiszen az alap határértéke $1-2\varepsilon/\theta$, ami 1-nél kisebb. Azaz a becslés gyengén konzisztens.

Fontos, hogy megértsük, a torzítatlanság nem sokat jelent, az a fontos, hogy a becslés konzisztens legyen, mert az jelenti azt, hogy ha nagy a mintaelemszám, akkor a becslés valóban közel kerül a becsült paraméterhez. □

3.1.8. Legyenek $X_1, \ldots, X_n$ független, Egyenletes$(0,\theta)$ eloszlású véletlen változók, ahol $\theta > 0$. Határozzuk meg $X_{1,n} = \min\{X_1, \ldots, X_n\}$ eloszlás-, és sűrűségfüggvényét, várható értékét és szórását!

Mutassuk meg, hogy

$$T_2(\mathbf{X}) = \frac{X_{1,n} + X_{n,n}}{2}$$

torzítatlan becslése $(\theta/2)$-nek. Igaz, hogy T_2 hatásosabb, mint a $\overline{X}_n$ mintaátlag, vagy mint T_1 az előző feladatból? Igazoljuk, hogy T_2 gyengén konzisztens!

3.1.9. Még mindig egyenletes eloszlás. Legyenek $X_1, \ldots, X_{2n+1}$ független, Egyenletes$(0,\theta)$ eloszlású véletlen változók, ahol $\theta > 0$. Jelölje Y a nagyság szerint középső mintaelemet. Határozzuk meg Y eloszlását, várható értékét és szórását! Igazoljuk, hogy Y konzisztens becslése a $\theta/2$-nek!

3.1.10. Legyenek $X_1, \ldots, X_n$ független véletlen változók, $f(x) = \frac{2x}{3\theta^2}$, $\theta \leq x \leq 2\theta$, sűrűségfüggvénnyel. Adjunk becslést θ-ra momentum módszerrel és ML módszerrel is!

Megoldás. Először megadjuk a paraméter momentum becslését. Ehhez a várható értéket kell meghatározni. Definíció szerint

$$\mathbf{E}_\theta(X) = \int_{-\infty}^{\infty} x f_\theta(x)\mathrm{d}x = \int_{\theta}^{2\theta} x\frac{2x}{3\theta^2}\mathrm{d}x$$
$$= \frac{2}{3\theta^2}\left[\frac{x^3}{3}\right]_{x=\theta}^{x=2\theta} = \frac{14\theta}{9}.$$

Tehát úgy kell választani a $\hat{\theta}$ becslést, hogy a hozzá tartozó elméleti várható érték éppen a mintaátlag legyen, azaz

$$\frac{14\hat{\theta}}{9} = \overline{X}_n,$$

azaz a paraméter momentum becslése

$$\hat{\theta} = \frac{9}{14}\overline{X}_n.$$

Térjünk rá a maximum likelihood becslésre. A minta együttes sűrűségfüggvénye (ne feledkezzünk el az intervallumról, hiszen pont az függ a paramétertől)

$$f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{2^n}{3^n\theta^{2n}} \prod_{i=1}^{n} x_i I(\theta \leq x_i \leq 2\theta).$$

Rögzített $x_1, \ldots, x_n$ realizáció esetén ennek kell meghatározni a maximumhelyét, mint θ függvénye, azaz

$$L(\theta) = \frac{2^n}{3^n\theta^{2n}} \prod_{i=1}^{n} x_i I(\theta \leq x_i \leq 2\theta).$$

Ezt egy kicsit nézegessük, ne deriváljunk ész nélkül. Látjuk, hogy minél kisebb θ annál nagyobb θ^{-2n}, a többi meg konstans. Na de figyelni kell, hogy θ olyan, hogy $\theta \leq \min x_i$, persze, hiszen minden mintaelem legalább θ, és $2\theta \geq \max x_i$, hiszen minden mintaelem legfeljebb 2θ. Tehát a legkisebb olyan θ paraméter kell, ami (1) $\theta \leq \min x_i$ és (2) $\theta \geq \max x_i/2$. Összegezve a maximum likelihood becslés

$$\widehat{\theta}_{ML} = \frac{1}{2} \max_{1 \leq i \leq n} x_i.$$

□

3.1.11. Legyenek $X_1, \ldots, X_n$ független, azonos eloszlású véletlen változók, melyek közös eloszlásfüggvénye F_θ, várható értéke $\mu(\theta)$, a szórásnégyzete σ^2. A szórásnégyzet ismert, a várható értéket becsüljük a mintaátlaggal. Legalább mekkora legyen n, hogy

$$\mathbf{P}_\theta\left(|\overline{X}_n - \mu| > 0,01\right) \leq 0,01.$$

Használjuk a Csebisev-egyenlőtlenséget! És ha normális közelítést használunk?

3.1.12. Egy gyárban a termékek minőségét úgy ellenőrzik, hogy minden nap n terméket vizsgálnak meg. Az adott napi összes gyártmányt akkor fogadják el, ha minden megvizsgált gyártmány jó. Azt tapasztalták, hogy m nap alatt összesen x-szer fogadták el a napi gyártmányokat. Adjunk maximum likelihood becslést annak a valószínűségére, hogy egy termék selejtes!

3.1.13. Mendel törvényei szerint egy növény AA, AB, BB genotípusa rendre θ^2, $2\theta(1-\theta)$, $(1-\theta)^2$ arányban fordul elő. Egy területen a három gyakoriságra x_{AA}, x_{AB}, és x_{BB} adódott. Adjunk maximum likelihood becslést θ-ra! Mutassuk meg, hogy a kapott becslés torzítatlan!

Megoldás. Legyen n a megvizsgált növények száma, és jelölje X az AA genotípusúak, Y az AB genotípusúak számát. Ekkor persze a BB genotípusúak száma $n-X-Y$. Ha θ a valódi paraméter, akkor

$$\mathbf{P}_\theta(X=k, Y=\ell) = \binom{n}{k}\binom{n-k}{\ell}(\theta^2)^k(2\theta(1-\theta))^\ell((1-\theta)^2)^{n-k-\ell},$$

$k+\ell \le n$, azaz (X,Y) trinomiális eloszlású. Valóban, az n növény közül először kiválasztom a k db AA-t, aztán a maradék $n-k$ közül az ℓ db AB-t, a többi pedig BB. Az AA valószínűsége θ^2, az AB-é $2\theta(1-\theta)$, míg a BB-é $(1-\theta)^2$. Nyilván, ha arányokat számolunk, akkor k/n, ℓ/n a lehetséges értékek. Azaz ha X_r és Y_r jelöli a megfelelő arányokat, akkor

$$\mathbf{P}_\theta(X_r=k/n, Y_r=\ell/n) = \binom{n}{k}\binom{n-k}{\ell}(\theta^2)^k(2\theta(1-\theta))^\ell((1-\theta)^2)^{n-k-\ell},$$

$k+\ell \le n$. Adott x_{AA} és x_{AB} realizáció esetén $k = nx_{AA}$ és $\ell = nx_{AB}$ értékeket írva

$$\begin{aligned}
&\mathbf{P}_\theta(X_r = x_{AA}, Y_r = x_{AB}) \\
&= \binom{n}{nx_{AA}}\binom{n-nx_{AA}}{nx_{AB}}(\theta^2)^{nx_{AA}}(2\theta(1-\theta))^{nx_{AB}}((1-\theta)^2)^{n-nx_{AA}-nx_{AB}} \\
&= \binom{n}{nx_{AA}}\binom{n-nx_{AA}}{nx_{AB}}2^{nx_{AB}}\left[\theta^{x_{AA}+x_{AB}/2}(1-\theta))^{x_{AB}/2+x_{BB}}\right]^{2n}.
\end{aligned}$$

Ezt kell maximalizálni θ-ban. Látjuk, hogy elég a szögletes zárójelen belüli értéket maximalizálni θ-ban, hiszen a szorzó nem függ θ-tól és az n-edik hatvány szigorúan monoton. Felhasználva, hogy $x_{AA}+x_{AB}+x_{BB} = 1$, a maximalizálandó függvény

$$h(\theta) = \theta^u(1-\theta)^{1-u},$$

ahol $u = x_{AA} + x_{AB}/2$. Deriváljuk h-t

$$h'(\theta) = \left(\frac{u}{\theta} - \frac{1-u}{1-\theta}\right)\theta^u(1-\theta)^{1-u}.$$

Látjuk, hogy $h'(\theta) = 0$ akkor és csak akkor ha $\theta = u$, és ez valóban maximum hely. Azaz a θ ML becslése

$$\widehat{\theta}_{ML} = x_{AA} + \frac{x_{AB}}{2}.$$

Az, hogy ez a becslés torzítatlan azt jelenti, hogy

$$\mathbf{E}(X + \frac{1}{2}Y) = n\theta.$$

Az együttes eloszlás nem is kell ehhez. Nyilván X és Y is binomiális eloszlásúak, n és θ^2, illetve n és $2\theta(1-\theta)$ paraméterekkel. Tehát a várható érték

$$\mathbf{E}(X + \frac{1}{2}Y) = \mathbf{E}(X) + \frac{1}{2}\mathbf{E}(Y) = n\theta^2 + \frac{1}{2}n2\theta(1-\theta) = n\theta,$$

amint állítottuk. □

3.1.14. A színtévesztés leggyakoribb fajtája az X kromoszómához kapcsoltan, nemhez kötötten öröklődik. Emiatt a férfiaknál a gyakoriság p, míg a nőknél csak p^2. Adjunk maximum likelihood becslést p-re, az alapján, hogy M férfiból m, N nőből pedig n volt színtévesztő!

3.1.15. Egy gombafajta spórái nyolcelemű lánc alakjában keletkeznek. A lánc különböző részekre szakadhat, mind a hét lehetséges helyen egymástól függetlenül p valószínűséggel. Határozzuk meg, hogy egyetlen lánc várhatóan hány darabra szakad!

Egy tényleges kísérlet során 7251 spórát számoltak meg, melyek $N = 907$ láncból származtak (5 spóra elveszett). A spórák 1975 láncra szakadtak szét. Adjunk becslést p értékére! Torzítatlan-e a kapott becslés? ([4])

3.1.16. Egy tóban a halakat egy betegség támadta meg, mely ismert p valószínűséggel pusztítja el az egyes egyedeket. A kifogott haltetemek k számából adjunk maximum likelihood becslés a betegség előtt a tóban élt halak számára!

3.1.17. Családok jövedelmét egy olyan skálán mérjük, ahol $X = 1$ a létminimumnak felel meg. Feltételezzük, hogy a jövedelem eloszlása $f(x) = \frac{\theta}{x^{\theta+1}}$, $x \geq 1$, sűrűségfüggvénnyel adható meg. Adjunk maximum likelihood becslést θ-ra, ha 10 véletlenszerűen választott család jövedelme: 1,53, 2,76, 19,65, 4,16, 7,31, 1,21, 254,2, 5,43, 1,12, 1,63.

3.1.18. Egy alkatrész élettartama exponenciális eloszlású θ/t várható értékkel, ha t hőmérsékleten működtetjük. Tegyük fel, hogy az n megfigyelést $t_1, \ldots, t_n$ hőmérsékleten végeztük és $x_1, \ldots, x_n$ élettartamokat figyeltünk meg. Adjunk ML becslést θ-ra!

3.1.19. Augusztusban 5 éjszakán át figyeltük meg a hullócsillagok számát. A következő mintát kaptuk: 4, 3, 7, 2, 4. A hullócsillagok száma egy este Poisson-eloszlású. Adjunk ML becslést az eloszlás paraméterére!

3.1.20. Egy céllövő ismeretlen p valószínűséggel talál el egy célpontot. Adjunk ML-becslést p-re, ha az első sikeres lövés k-adikra következett be. Torzítatlan-e a kapott becslés? A második sikeres lövésre további ℓ lövésig kellett várni. Ezt figyelembe véve adjunk ML-becslést p-re!

3.1.21. A müncheni Oktoberfesten minden évben az első hordó sört München főpolgármestere nyitja ki. Ez úgy történik, hogy a hordóba egy sörcsapolót üt bele az Oberbürgermeister. Tegyük fel, hogy minden ütés után p valószínűséggel indul meg a sör. Dieter Reiter főpolgármesternek 2018. szeptember 22-én a 3. ütésre sikerült kinyitni a hordót. Ezek alapján adjunk maximum likelihood becslést p-re!

3.2. Konfidenciaintervallumok, próbák

3.2.1. Egy játék fizikai terhelhetőségére elvégzett tesztek kg-ban a következő eredményeket adták: 40, 45, 40, 42, 36. Tegyük fel, hogy a háttérváltozó normális eloszlást követ 2 szórással. (a) Adjunk 95%-os konfidenciaintervallumot a várható értékre! (b) Oldjuk meg a feladatot ismeretlen szórás esetén is!

Megoldás. (a) Tegyük fel, hogy a háttéreloszlás normális, ismeretlen μ várható értékkel, és ismert $\sigma = 2$ szórással. (Figyeljünk oda, hogy ez a szórás, és nem a szórásnégyzet.) Az, hogy az adatok normális eloszlásúak egy teljesen természetes feltevés, ami általában teljesül. A centrális határeloszlás-tétel miatt van ez. A statisztikai probléma az az, hogy határozzuk meg a várható értéket. A szórás ismerete bizonyos esetekben indokolható: méréseknél ismerjük a mérőeszköz tulajdonságait. Nekünk az ismert szórás azért kényelmes, mert ez a legegyszerűbb eset.

Tegyük fel, hogy $X_1, X_2, \ldots, X_n$ független normális véletlen változók ismeretlen μ várható értékkel és $\sigma = 2$ szórással. Persze az ismeretlen μ várható értékre a becslésünk a mintaátlag, azaz

$$\widehat{\mu} = \overline{X}_n.$$

Ehhez nem kellett volna ennyi mindent tanulni. A kérdés az, hogy ez mennyire pontos. Adjunk meg egy olyan intervallumot, amire teljesül az, hogy az *igazi paraméter* 0,95 valószínűséggel beleesik ebbe az intervallumba. Mivel a normális eloszlás szimmetrikus a várható értékére, ezért természetes az intervallumot $[\widehat{\mu}-\delta, \widehat{\mu}+\delta]$ alakban keresni, ahol $\delta > 0$. Közben tartsuk észben, hogy ez az intervallum a véletlentől függ, hiszen $\widehat{\mu}$ egy véletlen mennyiség. Tehát úgy szeretnénk megválasztani δ értékét, hogy a $\mu \in [\widehat{\mu}-\delta, \widehat{\mu}+\delta]$ esemény valószínűsége 0,95 legyen. Átírva

$$\widehat{\mu}-\mu = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) = \frac{\sigma}{n}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}.$$

Az $(X_i-\mu)/\sigma$ változó normális eloszlású, a várható értéke $\mu-\mu=0$, szórása $\sigma/\sigma = 1$. Azaz ő standard normális. Ezt láttuk a normális eloszlás tulajdonságainál. Azt kell felhasználnunk, hogy *független normális eloszlású véletlen változók összege normális eloszlású.* Ezt nem bizonyítjuk sem itt, sem az előadáson. Mivel függetlenek, várható értékük és szórásnégyzetük is összeadódik. Ezt az állítást bizonyítottuk tetszőleges változókra. Összegezve

$$\widehat{\mu}-\mu$$

véletlen változó normális eloszlású, várható értéke 0, szórásnégyzete pedig σ^2/n. Ezek szerint

$$\frac{\sqrt{n}}{\sigma}(\widehat{\mu}-\mu)$$

standard normális véletlen változó. Tehát

$$\begin{aligned}\mathbf{P}(|\widehat{\mu}-\mu| \le \delta) &= \mathbf{P}\left(\frac{\sqrt{n}}{\sigma}(\widehat{\mu}-\mu) \in [-\delta\sqrt{n}/\sigma, \delta\sqrt{n}/\sigma]\right) \\ &= \Phi(\delta\sqrt{n}/\sigma) - \Phi(-\delta\sqrt{n}/\sigma) \\ &= 2\Phi(\delta\sqrt{n}/\sigma) - 1.\end{aligned}$$

Ez utóbbi kell 0,95 legyen, azaz

$$\Phi(\delta\sqrt{n}/\sigma) = 0,975.$$

A normális eloszlás táblázatából kapjuk, hogy

$$\frac{\delta\sqrt{n}}{\sigma} = \Phi^{-1}(0,975) = 1,96$$

vagyis

$$\delta = \frac{\sigma}{\sqrt{n}} 1,96 = 1,75.$$

Most használjuk csak a konkrét adatainkat. Eszerint $n = 5$ és $\overline{x}_n = 40,6$, ezért a 95%-os konfidenciaintervallum [40,6−1,75, 40,6+1,75] = [38,85, 42,35]. Tehát azt kaptuk, hogy a valódi várható érték, ami determinisztikus de ismeretlen, 95% valószínűséggel beleesik a [38,85, 42,35] (véletlen!) intervallumba.

Ez ugyanaz a levezetés, mint az előadásjegyzetben a konfidenciaintervallum résznél, csak konkrét számokkal. És persze ez van a képletgyűjteményben is.

(b) Minden ugyanaz mint az előbb, csak nem ismerjük a σ szórást. Ezért becsüljük a korrigált empirikus szórásnégyzet gyökével, azaz a

$$Z_n = \left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \right)^{1/2}$$

mennyiséggel. Az előbb, ismert szórás esetén az volt a kulcs, hogy a

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

véletlen változó standard normális. Most nem ismerjük σ értékét, ezért nem tudunk leosztani. Ezért a

$$\frac{1}{Z_n(X)\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \tag{1}$$

mennyiséget tekintjük. Előadáson bizonyítás nélkül volt, hogy a fenti mennyiség számlálója és nevezője függetlenek. Azaz, *normális eloszlás esetén a mintaátlag és az empirikus szórásnégyzet függetlenek.* Ez egy nagyon meglepő állítás, hiszen ugyanabból mintából gyártjuk le őket. Sőt, a (1)-ben szereplő hányados Student-eloszlású $n-1$ szabadsági fokkal. Ezek után minden ugyanúgy megy, mint ismert szórásnál, csak más táblázatból kell kinézni az adatokat. Tehát

$$\begin{aligned}
\mathbf{P}(|\widehat{\mu} - \mu| \le \delta) &= \mathbf{P}\left(\frac{\sqrt{n}}{Z_n}(\widehat{\mu} - \mu) \in [-\delta\sqrt{n}/Z_n, \delta\sqrt{n}/Z_n] \right) \\
&= \Phi_{n-1}(\delta\sqrt{n}/Z_n) - \Phi_{n-1}(-\delta\sqrt{n}/Z_n) \\
&= 2\Phi_{n-1}(\delta\sqrt{n}/Z_n) - 1.
\end{aligned}$$

Itt most Φ_{n-1} az $n-1$ szabadsági fokú Student-eloszlás eloszlásfüggvénye. Ez utóbbi kell 0,95 legyen, azaz

$$\Phi_{n-1}(\delta\sqrt{n}/Z_n) = 0,975.$$

A táblázatából kapjuk, hogy

$$\frac{\delta\sqrt{n}}{Z_n} = \Phi_{n-1}^{-1}(0,975)$$

vagyis

$$\delta = \frac{Z_n}{\sqrt{n}}\Phi_{n-1}^{-1}(0,975),$$

és a konfidenciaintervallum:

$$\left[\overline{X}_n - \delta, \overline{X}_n + \delta\right].$$

Beírva az adatainkat, $n = 5$, $\overline{X}_5 = 40,6$, $Z_5 = 3,29$, és $\Phi_4^{-1}(0,975) = 2,776$ (itt a 4. sort figyeltük), azaz $\delta = 3,29 \cdot 2,776/\sqrt{5} = 4,08$. Tehát a konfindenciaintervallum: [36,52, 44,68].

Ez persze nagyobb, mint az előbb, azaz kevésbé vagyunk biztosak a becslésben, de ezen nem lepődünk meg, hiszen kevesebbet tudunk, nem tudjuk a szórást. □

3.2.2. Egy véletlen változó értékeit megfigyelve a következő statisztikai mintát kapjuk: $6,5$, $7,3$, $5,4$, $6,5$, $2,1$. (a) Tegyük fel, hogy a háttéreloszlás normális, ismert $\sigma = 2$ szórással és ismeretlen μ várható értékkel. Teszteljük 5%-os szignifikanciaszinten azt a nullhipotézist, hogy az elméleti várható érték 8.

(b) Oldjuk meg a feladatot ismeretlen szórás esetén is.

Megoldás. Az (a) résznél egy u-próbát kell végrehajtanunk, hiszen normális eloszlás várható értékét teszteljük ismert szórás mellett. A fenti minta alapján becsült várható érték, a mintaátlag $\overline{X}_5 = 5,56$. A kérdés, hogy elhisszük-e azt a feltevést, hogy a valódi várható érték 8? Persze gyanús a dolog, de ha a valódi várható érték tényleg 8 lenne, akkor sem várnánk, hogy a mintaátlag egy 5 elemű minta esetén éppen 8.

A nullhipotézis az amit felteszünk. Jelen esetben az, hogy a valódi várható érték $\mu_0 = 8$. Az alternatív-, vagy ellenhipotézis, hogy $\mu_0 \neq 8$. A próbastatisztikát már láttuk a konfidenciaintervallum konstruálásánál is,

$$u = \sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} = -2,72.$$

Ha a nullhipotézis igaz, akkor ez az u statisztika standard normális eloszlású. Ezt is láttuk a konfidenciaintervallumnál. Tehát azt kell eldönteni, hogy elhisszük-e egy standard normális eloszlású véletlen változóról, hogy értéke

$-2,72$. Ehhez választunk egy α szignifikanciaszintet, ez most 0,05. Keressük azt az u_α küszöbértéket, amelyre teljesül, hogy $\mathbf{P}(|Z| > u_\alpha) = \alpha$, ahol Z standard normális. A normális eloszlás szimmetriája miatt

$$\mathbf{P}(|Z| > u_\alpha) = \mathbf{P}(Z > u_\alpha) + \mathbf{P}(Z < -u_\alpha) = 2(1 - \Phi(u_\alpha)),$$

tehát $u_\alpha = \Phi^{-1}(1 - \alpha/2)$. Esetünkben $u_{0,05} = \Phi^{-1}(0,975) = 1,96$. Mivel $|u| = 2,72 > 1,96$, ezért ha igaz lenne a nullhipotézis, akkor a bekövetkezett esemény valószínűsége 0,05-nél kisebb, vagyis egy valószínűtlen esemény következett be. Ekkor 5%-os szignifikanciaszinten *elvetjük* a nullhipotézist. Azt mondjuk, hogy az elfogadási tartomány a $[-1,96,\ 1,,96]$ intervallum. Ha a próbastatisztika értéke ide esik, elfogadjuk a nullhipotézist, különben elvetjük.

(b) Ismeretlen szórás esetén azt is becsüljük. Minden úgy megy, mint a konfidenciaintervallum esetén. Most a próbastatisztika

$$t = \sqrt{n}\frac{\overline{X}_n - \mu_0}{Z_n} = -2,67.$$

Itt persze Z_n a korrigált empirikus szórás. Ha a nullhipotézis igaz, akkor t véletlen változó Student-eloszlást követ $n-1$ szabadsági fokkal. Az ehhez tartozó kritikus érték az a t_α, melyre $\mathbf{P}(|T| > u_\alpha) = \alpha$, ahol T Student-eloszlású $n-1$ szabadsági fokkal. A Student-eloszlás szimmetriája miatt

$$\mathbf{P}(|T| > t_\alpha) = \mathbf{P}(T > t_\alpha) + \mathbf{P}(T < -t_\alpha) = 2(1 - \Phi_{n-1}(t_\alpha)),$$

tehát $t_\alpha = \Phi_{n-1}^{-1}(1 - \alpha/2)$. Esetünkben $t_{0,05} = \Phi_4^{-1}(0,975) = 2,776$. Most $|t| \leq 2,776$, azaz 5%-os szignifikanciaszinten elfogadjuk a nullhipotézist. $\square$

3.2.3. A Cambridge-i Egyetem kutatói a sztochasztika kurzusok vérnyomásra gyakorolt hatását vizsgálták. Öt fizikus hallgató vérnyomását közvetlenül sztochasztika kurzus után megmérve a szisztolés értékekre 115, 130, 120, 135, 120 értékeket kapták, míg a kontrollcsoportban 125, 120, 115, 120, 125 értékek adódtak. A vérnyomásértékek mindkét mintában normális eloszlást követnek μ_1 és μ_2 várható értékekkel, és megegyező σ szórással. Adjunk 90% megbízhatósági szintű konfidenciaintervallumot a $\mu_1 - \mu_2$ különbségre!

3.2.4. A Liverpool 2020. február 18. - március 11. között lejátszott 6 mérkőzésén a Liverpool labdabirtoklása (%-ban kifejezve) 71, 68, 68, 58, 70, 72. Tegyük fel, hogy a labdabirtoklás %-os értéke normális eloszlást követ ismeretlen μ várható értékkel és ismert $\sigma = 5$ szórással. Teszteljük 5%-os szignifikanciaszinten azt a nullhipotézist, hogy a $\mu = 65$.

3.2.5. Régészek radiokarbonos kormeghatározással szeretnék meghatározni egy lelőhely korát. Ismert, hogy a radiokarbonos módszert az egyazon ásatáson talált különböző leleteken alkalmazva nem pontosan ugyanazt a kort fogjuk megkapni minden lelet esetében, hanem a kapott korok (közelítőleg) normális eloszlást követnek, melynek elméleti várható értéke a lelőhely igazi kora.

(a) A radiokarbonos módszert hét leleten alkalmazva a következő korokat kapjuk: 1180, 1220, 1230, 1250, 1270, 1290 és 1340 év. Adjunk becslést a lelőhely korára, valamint írjunk fel egy 95% megbízhatósági szintű konfidenciaintervallumot erre a korra. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy a lelőhely kora 1220 év.

(b) Egy másik, közeli ásatásról 6 leletet vetnek alá kormeghatározásnak. A mintaátlag 1100 évnek, a korrigált empirikus szórás 50 évnek adódik. (Feltehető, hogy a két lelőhelyről származó minták esetében azonos a radiokarbonos módszerrel kapott korok elméleti szórása.) Teszteljük 10%-os szignifikancia szinten azt a nullhipotézist, hogy a két lelőhely egyidős, tehát azonos az elméleti várható érték. Adjunk meg egy 90% megbízhatósági szintű konfidenciaintervallumot a lelőhelyek kora közötti különbségre.

(c) Ellenőrizzük le a (b) feladatrész feltevését, teszteljük le 10%-os szignifikanciaszinten azt a nullhipotézist, hogy a két lelőhelyen azonos az elméleti szórás.

Megoldás. Az (a) részben nincs semmi új. A mintaelemszám $n = 7$, a mintaátlag $\overline{X}_7 = 1254$, a korrigált empirikus szórás $Z_7 = 52$, és $\Phi_6^{-1}(0,975) = 2,36$, így $\delta = \Phi_6^{-1}(0,975) \cdot Z_7/\sqrt{7} = 46$. Tehát a becslésünk a lelőhely korára 1254, a 95%-os konfidenciaintervallum pedig $[1208, 1300]$.

Mivel nem tudjuk az elméleti szórást t próbát használunk. A t-statisztika értéke

$$t = \sqrt{7}\frac{1254 - 1220}{Z_7} = 1,73.$$

Az elfogadási tartomány [−2,36, 2,36], tehát elfogadjuk a nullhipotézist.

A (b) részben két mintát kell összehasonlítani. Feltettük, hogy a szórások megegyeznek. Az új minta elemszáma $m = 6$, mintaátlaga $\overline{Y}_6 = 1100$, és korrigált empirikus szórása $W_6 = 50$. A szórásnégyzet mindkét mintában ugyanannyi, de ismeretlen. Ezért a szórásnégyzetet becsülni kell a különkülön vett empirikus szórásnégyzetek megfelelő súlyozott összegével:

$$D^* = \left[\left((n-1)Z_n^2 + (m-1)W_m^2\right)\frac{n+m}{nm(n+m-2)}\right]^{1/2} = 28,4$$

Azt teszteljük, hogy a várható értékek egyenlők (azaz $\Delta = 0$ a képletgyűjteményben), így a statisztikánk

$$t' = \frac{\overline{X}_7 - \overline{Y}_6}{D^*} = \frac{1254 - 1100}{28,4} = 5,4.$$

Ha a nullhipotézis igaz, azaz a két várható érték megegyezik, akkor a statisztika eloszlása Student $n + m - 2 = 11$ szabadsági fokkal. A megfelelő kritikus érték $\Phi_{11}^{-1}(0,95) = 1,8$, ami kisebb, mint az 5,4, ezért elvetjük a nullhipotézist.

A konfidenciaintervallum a várható értékek különbségére

$$\left[\overline{X}_7 - \overline{Y}_6 - \Phi_{11}^{-1}(0,95) \cdot D^*, \overline{X}_7 - \overline{Y}_6 + \Phi_{11}^{-1}(0,95) \cdot D^*\right]$$
$$= [103, 205].$$

Látjuk, hogy a 0 nincs benne az intervallumban, ezért vetettük el a nullhipotézist.

A (c) részben a szórások egyenlőségét kell tesztelni. Ilyen nem volt előadáson. Ehhez a korrigált empirikus szórásnégyzetek hányadosát kell nézegetni. Ha ez közel 1, akkor elhisszük, hogy a két szórás azonos, ha kicsi vagy nagy, akkor nem hisszük el. Pontosabban, ha az elméleti szórások egyenlők, akkor az

$$F = \frac{Z_n}{W_m} \tag{F}$$

próbastatisztika F eloszlású $n - 1$ és $m - 1$ szabadsági fokkal. Most vigyázni kell, mert az F eloszlás nem szimmetrikus a 0-ra (persze, hiszen ő pozitív). Ezért rögzített α szignifikanciaszinthez vesszük az $f_1 = F_{n-1,m-1}^{-1}(\alpha/2)$, és $f_2 = F_{n-1,m-1}^{-1}(1 - \alpha/2)$ értékeket. Akkor vetjük el a nullhipotézist, ha az F statisztika értéke kicsi, azaz $F < f_1$, vagy ha nagy, azaz $F > f_2$. Itt egy kicsit egyszerűsíthetjük a dolgot a következőképpen. Kiszámoljuk a Z_n, $V_m^*(Y)$ értékeket, és *a nagyobbat osztjuk a kisebbikkel* az (F) képletben. Ekkor figyelni kell, hogy mi van a számlálóban, hiszen ha $Z_n > W_m$ akkor $F_{n-1,m-1}$ az eloszlás, különben $F_{m-1,n-1}$. Ekkor csak az f_2 kritikus értéket kell meghatározni, és ha $F > f_2$ akkor elvetem a nullhipotézist.

Beírva az adatokat, $Z_7 = 52^2$, és $W_6 = 50^2$, azaz

$$F = \frac{52^2}{50^2} = 1,08.$$

Persze látjuk mindjárt, hogy ezek nagyon egyenlőek, ezért úgyis elfogadjuk a nullhipotézist. Most $7 - 1 = 6$ és $6 - 1 = 5$ a két paraméter (*számít a sorrend!*), ezért a kritikus érték $f = F_{6,5}^{-1}(0,95) = 4,95$. Ez jóval nagyobb, mint a mi F statisztikánk, ezért elfogadjuk a nullhipotézist, azaz a szórások egyenlőségét. □

3.2.6. Egy játék fizikai terhelhetőségére elvégzett tesztek kg-ban a következő eredményeket adták: 40, 45, 40, 42, 36, 41, 43. A 2019. január 1. óta hatályos EU-s ISO-31415 szabvány szerint a gyermekjátékoknak terhelhetősége legalább 40 kg kell legyen.

(a) Tegyük fel, hogy a terhelhetőség normális eloszlást követ ismeretlen μ várható értékkel és ismert $\sigma = 3.1$ szórással. Teszteljük 1%-os szignifikanciaszinten azt a nullhipotézist, hogy a $\mu = 40$.

(b) Oldjuk meg a feladatot ismeretlen szórás esetén.

3.2.7. Az FC Barcelona 2019. április 16. - május 1. között lejátszott 5 mérkőzésén a Barcelona labdabirtoklása (%-ban kifejezve) 47, 62, 77, 50, 66.

(a) Adjuk meg a mintához tartozó mintaátlagot, a tapasztalati szórásnégyzetet és az empirikus eloszlásfüggvényt!

(b) Tegyük fel, hogy a labdabirtoklás %-os értéke normális eloszlást követ ismeretlen μ várható értékkel és ismert $\sigma = 10$ szórással. Teszteljük 5%-os szignifikanciaszinten azt a nullhipotézist, hogy a $\mu = 60$.

3.2.8. Az FC Barcelona 2019. április 16. - május 1. között lejátszott 5 mérkőzésén a Barcelona rúgott góljainak száma 3, 1, 2, 2, 3.

(a) Adjuk meg a mintához tartozó mintaátlagot, a tapasztalati szórásnégyzetet és az empirikus eloszlásfüggvényt!

(b) Tegyük fel, hogy a lőtt gólok száma Poisson-eloszlást követ, ismeretlen λ paraméterrel. Adjunk ML és momentumbecslést λ értékére!

3.2.9. A hároméves Máté reggelente véletlen időpontban ébred. Az elmúlt 5 napban az ébredésekre 6:20, 6:55, 6:30, 7:15, és 6:40 adódott.

(a) Adjuk meg a mintához tartozó mintaátlagot, a tapasztalati szórásnégyzetet és az empirikus eloszlásfüggvényt!

(b) Tegyük fel, hogy Máté ébredésének időpontja normális eloszlást követ ismeretlen μ várható értékkel és $\sigma = 20$ szórással. Adjunk 95%-os konfidenciaintervallumot az ébredés várható értékére!

3.2.10. A '80-as években egy klinikai kísérlet keretei között azt vizsgálták, hogy a nagy- dózisú kálciumbevitelnek van-e vérnyomáscsökkentő hatása. A kísérlet időtartama alatt 10 alany kálciumtablettákat szedett, míg 11 másik ember, a kontroll csoport, placebot kapott. A 12 hetes kísérlet végén a kísérleti alanyok vérnyomása 100, 114, 105, 112, 115, 116, 106, 102, 125 és 104 Hgmm volt, míg a kontroll csoportban mért vérnyomásértékek 124, 97, 113, 105, 95, 119, 114, 114, 121, 118 és 133 Hgmm voltak. Feltehető, hogy a vérnyomásértékek mindkét csoportban normális eloszlást követnek.

(a) Tegyük fel, hogy a kíséreti és a kontroll csoportban azonos a vérnyomásértékek elméleti szórása. Teszteljük le 5%-os szignifikancia szinten azt a nullhipotézist, hogy a két csoportban azonos a vérnyomásértékek elméleti várható értéke. Érdemes bevezetni a gyógyászatban a nagydózisú kálciumkezelést, mint a ma- gas vérnyomás ellenszerét?

(b) Teszteljük le azt a feltevést, hogy a két csoportban azonos a vérnyomásértékek elméleti szórása.

([8])

3.2.11. Ismert, hogy a kakukkok más madarak fészkeibe rakják a tojásukat. 1940-ben Edgar Chance angol ornitológus azt vizsgálta, hogy a kakukktojások mérete függ-e attól, hogy a kakukk milyen fajtájú madár fészkébe csempészi bele a tojását. Megmért 16 illetve 15 kakukktojást, melyeket vörösbegyek illetve ökörszemek fészkében talált. A vörösbegyfészkekben talált tojások átlagos hosszúsága 22,4 mm volt, míg ugyanez az érték az ökörszemfészkekben talál tojásoknál 21,2 mm volt. A korrigált empirikus szórás a két minta esetében 0,94 mm illetve 0,68 mm volt. Feltehető, hogy a kakukktojások hossza mindkét fészekben normális eloszlást követ.

(a) Tegyük fel, hogy a két fészektípus esetében azonos a kakukktojások hosszá- nak az elméleti szórása. Teszteljük le 10%-os szignifikancia szinten azt a null- hipotézist, hogy a kakukktojások hosszának az elméleti várható értéke azonos a vörösbegyek és az ökörszemek esetében. Adjunk 90% megbízhatósági szintű konfidencia intervallumot a várható értékek különbségére.

(b) Teszteljük le az a. pontban alkalmazott feltevésünket is, tehát azt, hogy a két fészektípus esetében megegyezik a kakukktojások hosszának a szórása. A szignifikancia szint legyen 10%.

([8])

3.2.12. Feldobunk egy nem feltétlenül szabályos dobókockát 100 alkalommal. A dobá- sok során 15 egyest, 15 kettest, 15 hármast, 15 négyest, 20 ötöst és 20 hatost kaptunk. A minta alapján adjunk pontbecslést az egyes értékek dobásának a valószínűségére.

(a) Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy a dobókocka szabályos. Teszteljük külön azt a nullhipotézhist, hogy a hatosdobás valószínűsége 1/6.

(b) Ugyanezt a dobókockát most 1000 alkalommal dobjuk fel, melyből 150 egyest, 150 kettest, 150 hármast, 150 négyest, 200 ötöst és 200 hatost kapunk. Oldjuk meg az (a) feladatrészt ezzel a módosítással!

([8])

Hivatkozások

[1] Alon, Spencer: *The probability method.*

[2] Bognár Jánosné, Mogyoródi József, Prékopa András, Rényi Alfréd, Szász Domokos: *Valószínűségszámítási feladatgyűjtemény.* Negyedik kiadás, Typotex 2001.

[3] Branko Curgus, Robert I. Jewett, An unexpected limit of expected values, *Expo. Math.* 25 (2007) 1–20.

[4] William Feller: Bevezetés a valószínűségszámításba és alkalmazásaiba I.

[5] Hajnal Péter: *Elemi kombinatorikai feladatok.* Polygon, 1997.

[6] Nagy-György Judit, Osztényiné Krauczi Éva, Székely László: *Valószínűségszámítás és statisztika példatár.* Polygon, 2007.

[7] Rüschendorf: On the distributional transform, Sklar's theorem, and the empirical copula process, *J. Statist. Plann. Inference* 139, 11, 3921–3927, 2009.

[8] Viharos László: Sztochasztika alapjai gyakorló feladatok. `http://www.math.u-szeged.hu/~viharos/vizsgak/sztochasztika_gyakorlat_fizikusoknak.html`.