

Az azonos napra eső születésnapok előfordulásának valószínűsége, várható értéke és szórása

MÓRICZ FERENC

1. Bevezetés

Az ún. *skatulya elv* lényegét egyszerűen a következőképpen lehet megfogalmazni: 12 skatulyába 12-nél több gyufaszál csak úgy helyezhető el, hogy legalább egy skatulyába egynél több gyufaszál kerül. Általánosabban kifejezve azt mondhatjuk, hogy ha n -nél több tárgyat n osztályba sorolunk be, akkor legalább egy osztályba egynél több tárgy jut; itt $n \geq 2$ tetszőleges természetes szám lehet.

Ezen elv alapján nyilvánvaló, hogy ha $n > 12$, akkor egy n személyből álló csoportban mindig van legalább két olyan ember, akik ugyanabban a hónapban születtek. Ha eltekintünk a szökőév február 29-ik napjától, akkor $n > 365$ esetén egy n személyből álló csoportban mindig van legalább két olyan ember, akik születésnapja az évnek ugyanazon napjára esik.

Ha viszont $2 \leq n \leq 365$; akkor már csak annak a valószínűségét tudjuk megadni, hogy egy n személyből álló csoportban van két olyan ember, akik az évnek ugyanazon napján születtek. Vagy még konkrétabban a következő kérdést tehetjük fel. Hány véletlenszerűen kiválasztott embernek kell egy helyiségben összejönnie ahhoz, hogy $\frac{1}{2}$ -nél nagyobb valószínűséggel állíthassuk, hogy az adott helyiségben legalább két olyan ember van, akik ugyanazon a napon ünneplik születésnapjukat? A meglepő válasz a következő részben az lesz, hogy ez a valószínűség már 23 ember esetén meghaladja az $\frac{1}{2}$ -et; amit úgy is kifejezhetünk, hogy ekkor a fentebbi esemény bekövetkezésének esélye nagyobb, mint 50 %. Ezen válasz megindoklásához csupán jólismert kombinatorikai megfontolásra lesz szükségünk.

Írásunk további részében is csak egyszerű valószínűségszámítási ismeretekre lesz szükségünk, amelyeket már a középiskolák matematika tagozatos osztályaiban is tanítanak: a binomiális és Poisson eloszlás, a véletlen változó, várható érték és szórás foglaimra. A binomiális eloszlásnak Poisson eloszlással történő approximációját részletesen fogjuk ismertetni, amely megközelítés a numerikus számítások lényeges egyszerűsítését teszi lehetővé. Továbbá, nemcsak ismertetjük, hanem alkalmazzuk is Csebisev híres egyenlőtlenségét.

Ismételten felhívjuk a figyelmet arra, hogy meggondolásainkban a szökőévektől eltekintünk; vagyis feltesszük, hogy az általunk tekintett év 365 napból áll. Például feltehetjük, hogy ha valaki február 29-én született, akkor születésnapját véletlenül kisorsoljuk az év többi 365 napja között. Azt is fontos feltételeznünk, hogy az n személyből álló csoport tagjainak a kiválasztása az ún. véletlen mintavétellel történik. Ez a feltétel azt jelenti, hogy a szóbanforgó n ember mindegyikének születésnapja az év bármelyik napjára egyenlő valószínűséggel esik; továbbá a csoportban levő emberek születésnapjai között semmiféle összefüggés nem áll fenn.

Cikkünkben az ún. eseményeket az ábécé nagy betűivel jelöljük:

$$A, A_2, A_3, A_4, \dots, A_n, A_n^{(i)}, A_{i,j}, B_n, B_{i,j,k}, C_1, C_2, C_{k,\ell};$$

valószínűségeiket pedig $P(A)$ -val, $P(A_2)$ -vel, $\dots$, $P(C_{k,\ell})$ -lel jelöljük.

2. Valószínűségek meghatározása kombinatorikai meggondolással

Jelöljük A_n -nel azt az eseményt, hogy egy n személyből álló csoportban van legalább két olyan ember, akik az évnek ugyanazon a napján születtek. Mivel $n > 365$ esetén A_n a bizonyos esemény, ezért ennek valószínűségére fennáll, hogy

$$P(A_n) = 1, \quad n > 365.$$

Ezért a következőkben feltesszük, hogy $2 \leq n \leq 365$.

Az A_n valószínűségének meghatározása végett először az A_n esemény *ellentétének* (amelyet szokás az A_n esemény komplementerének is nevezni) eseményt $\overline{A}_n$ -sal jelöljük. Az $\overline{A}_n$ azt az eseményt jelenti, hogy az n személyből álló csoportban nincs két olyan ember, akik az év ugyanazon napján születtek.

Feltevésünk szerint az emberek mindegyikének az általunk nem ismert születésnapja azonos valószínűséggel esik az év bármelyik napjára. Ezért az $\overline{A}_n$ esemény valószínűségének meghatározására a jólismert képletet használjuk:

$$P(\overline{A}_n) = \frac{\text{kedvező esetek száma}}{\text{lehetséges esetek száma}}.$$

Ennek alapján nyerjük az $n = 2, 3$ és 4 esetekben, hogy

$$\begin{aligned} P(\overline{A}_2) &= \frac{365 \cdot 364}{(365)^2} = \frac{364}{365}, \\ P(\overline{A}_3) &= \frac{365 \cdot 364 \cdot 363}{(365)^2} = \frac{364 \cdot 363}{(365)^2}, \\ P(\overline{A}_4) &= \frac{365 \cdot 364 \cdot 363 \cdot 362}{(365)^4} = \frac{364 \cdot 363 \cdot 362}{(365)^3}. \end{aligned}$$

Ebből már könnyen látszik, hogy az általános esetben a keresett valószínűség a következőképpen számítható ki:

$$P(\overline{A}_n) = \frac{364 \cdot 363 \cdot 362 \cdots (365 - n + 1)}{(365)^{n-1}}, \quad n = 2, 3, \dots$$

A szorzások és osztások elvégzésével ellenőrizhetjük (lásd az alábbi 1. Táblázatban), hogy

$$P(\overline{A}_{22}) > \frac{1}{2} > P(\overline{A}_{23}).$$

Így az ellentett esemény valószínűségére vonatkozó szabály szerint

$$P(A_{23}) = 1 - P(\overline{A}_{23}) > 1 - \frac{1}{2} = \frac{1}{2} > 1 - P(\overline{A}_{22}) = P(A_{22}).$$

Tehát az átlagember számára meglepő módon már 23 ember elegendő ahhoz, hogy $\frac{1}{2}$ -nél nagyobb valószínűséggel legyen közöttük legalább két olyan ember, akik az évnek ugyanazon a napján születtek.

A meglepetés oka bizonyára az lehet, hogy annak a valószínűsége, hogy két találmányra választott embernek ugyanarra a napra esik a születésnapja igen kicsi:

$$\frac{365 \cdot 1}{(365)^2} = \frac{1}{365}.$$

Ha azonban figyelembe vesszük azt a tényt, hogy egy 23 emberből álló csoportból

$$\binom{23}{2} = \frac{23 \cdot 22}{1 \cdot 2} = 253$$

különböző párt tudunk kiválasztani, akkor már nem látszik annyira meglepőnek az, hogy a 23 személyből álló csoportban $\frac{1}{2}$ -nél nagyobb valószínűséggel található két olyan ember, akiknek születésnapja az évnek ugyanazon napjára esik.

Igen meglepő az a tény is, hogy a $P(\overline{A}_n)$ valószínűség $n = 80$ esetén a 10^{-4} érték alá csökken, amelynek következtében

$$P(A_{80}) = 1 - P(\overline{A}_{80}) > 1 - \frac{1}{10\,000} = \frac{9\,999}{10\,000}.$$

Ezen valószínűség alapján 99,99 % biztonsággal állíthatjuk, hogy tetszőleges 80 főből álló csoportban legalább két embernek ugyanarra a napra esik a születésnapja. Az 1. Táblázatban még további érdekes valószínűség értékeket is feltüntettünk.

n	$P(\overline{A}_n)$	$P(A_n)$	az utóbbi valószínűség %-ban kifejezve
22	0,5243	0,4757	
23	0,4927	0,5073	> 50%
31	0,2695	0,7305	
32	0,2467	0,7533	> 75%
46	0,0517	0,9483	
47	0,0452	0,9548	> 95%
56	0,0117	0,9883	
57	0,0099	0,9901	> 99%
69	0,00103	0,99897	
70	0,00084	0,99916	> 99,9%
79	0,00011	0,99989	
80	0,00008	0,99992	> 99,99%
82	0,000053	0,999947	
83	0,000041	0,999959	> 99,995%

1. Táblázat

Az 1. Táblázat tanulmányozása során a következő *paradoxonra* (látszólag kép-
telenségre) hívjuk fel a figyelmet. Egy 69 főből álló csoportban $P(A_{69}) = 0,99847$,
míg ezt a 69 főt három, egyenként 23 főből álló csoportba szétosztva, mindegyik
esetében $P(A_{23}) = 0,5073$.

A jelenség pontosabb megértése céljából jelöljük $A_{23}^{(i)}$ -vel annak a valószínű-
ségét, hogy az i -edik 23 főből álló csoportban legalább két embernek az év azonos
napjára esik a születésnapja, ahol $i = 1, 2, 3$ lehet. A teljesen független események
valószínűségére vonatkozó szabály szerint (lásd a 3. rész elején) számolunk:

$$\begin{aligned}
 P\left(\begin{array}{l} \text{a három egyenként 23 főből álló csoport egyikében sincs két ember,} \\ \text{akiknek születésnapja az év azonos napjára esik} \end{array}\right) = \\
 = P\left(\overline{A_{23}^{(1)}} \cdot \overline{A_{23}^{(2)}} \cdot \overline{A_{23}^{(3)}}\right) = P\left(\overline{A_{23}^{(1)}}\right) P\left(\overline{A_{23}^{(2)}}\right) P\left(\overline{A_{23}^{(3)}}\right) = \\
 = (0,4927)^3 = 0,119604 \dots;
 \end{aligned}$$

ahonnan az ellentett esemény valószínűségére vonatkozó szabály szerint

$$P\left(\begin{array}{l} \text{a három egyenként 23 főből álló csoport legalább egyikében van két} \\ \text{ember, akiknek születésnapja az év azonos napjára esik} \end{array}\right) = \\ = 1 - (0,4927)^3 < 1 - 0,11960 = 0,88040.$$

Ez utóbbi valószínűség szignifikánsan (jelentősen, szembeszökően) kisebb, mint $P(A_{69}) = 0,99897$.

A két utóbbi valószínűség eltérését az okozza, hogy a 69 főből álló csoportból

$$\binom{69}{2} = \frac{69 \cdot 68}{2} = 2346$$

különböző pár választható ki; míg egy 23 főből álló csoportból

$$\binom{23}{2} = \frac{23 \cdot 22}{2} = 253,$$

és ezért három egyenként 23 főből álló csoportból csak

$$3 \binom{23}{2} = 759$$

különböző pár választható ki. Tehát a 69 főből álló csoportból több, mint háromszor annyi különböző pár választható ki, mint a három egyenként 23 főből álló csoportból együttesen.

3. A binomiális eloszlás approximációja Poisson eloszlással

Emlékeztetünk arra, hogy a binomiális eloszlás először az ún. *Bernoulli probléma* megoldásakor merült fel. Tekintsünk egy olyan kísérletet, amelynek két lehetséges kimenete van: egy A eseménynek és ennek ellentettjének (vagy komplementerének) a bekövetkezése (az utóbbit $\overline{A}$ -sal jelölve), amelyeknek valószínűségei

$$P(A) = p \text{ és } P(\overline{A}) = 1 - p, \text{ ahol } 0 < p < 1.$$

Az ilyen kísérletet *egyszerű alternatívának* szokás nevezni, ennek egymástól független többszöri megismétlését pedig *Bernoulli kísérletsorozatnak*. A következő kérdést Bernoulli Dániel vetette fel és oldotta meg. Mennyi annak a valószínűsége, hogy ha a kísérletet N -szer ismétljük meg (egymástól függetlenül), akkor az A esemény pontosan k -szor következik be, ahol $0 \leq k \leq N$. Jelöljük ezt a valószínűséget $p_k(N)$ -nel, ekkor

$$p_k(N) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k = 0, 1, 2, \dots, N.$$

Ugyanis, annak a valószínűsége, hogy az A esemény éppen az i_1 -edik, i_2 -edik, ..., i_k -edik kísérletekben következik be egyenlő a

$$p^k(1-p)^{N-k}$$

szorzattal; és N kísérletből pontosan $\binom{N}{k}$ számú ilyen k elemű kísérletsorozat választható ki. Itt

$$\begin{aligned}\binom{N}{k} &:= \frac{N!}{k!(N-k)!} = \frac{N(N-1)\dots(N-k+1)}{1 \cdot 2 \dots k}, \quad k = 1, 2, \dots, N-1; \\ \binom{N}{0} &= \binom{N}{N} := 1, \quad 0! := 1;\end{aligned}$$

a binomiális tételből jólismert *binomiális együtthatók*.

A *binomiális tételből* következik, hogy a

$$0 < p_k(N) < 1, \quad k = 0, 1, \dots, N,$$

számok ún. *valószínűségeloszlást* alkotnak, mivel

$$\begin{aligned}p_0(N) + p_1(N) + \dots + p_N(N) &= \\ &= (1-p)^N + \binom{N}{1}p(1-p)^{N-1} + \dots + \binom{N}{N-1}p^{N-1}(1-p) + p^N = \\ &= [(1-p) + p]^N = 1^N = 1.\end{aligned}$$

Az így nyert valószínűségeloszlást N -ed rendű, p paraméterű *binomiális eloszlásnak* nevezzük.

A binomiális eloszlásnak a Poisson eloszlással történő *approximációjához* (másszóval: közelítéséhez) úgy jutunk, hogy N -et ∞ -be, p -t pedig 0-hoz tartatjuk olyan módon, hogy

$$Np = \lambda, \quad \text{ahol } \lambda > 0 \text{ állandó.}$$

Egyszerű átalakítással kapjuk, hogy

$$\begin{aligned}\binom{N}{k}p^k(1-p)^{N-k} &= \frac{N(N-1)(N-2)\dots(N-k+1)}{k!}p^k(1-p)^{N-k} = \\ &= \frac{1\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{k-1}{N}\right)(Np)^k(1-p)^N}{k!(1-p)^k} = \\ &= \frac{\lambda^k}{k!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{k-1}{N}\right)\frac{\left(1-\frac{\lambda}{N}\right)^N}{\left(1-\frac{\lambda}{N}\right)^k}.\end{aligned}$$

Ismeretes az e szám definíciója:

$$e := \lim_{N \rightarrow \infty} \left(1 + \frac{1}{N}\right)^N.$$

Ebből egyszerűen levezethető, hogy

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = e^{-1}$$

és általában

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\lambda}{N}\right)^N = e^{-\lambda}.$$

Mivel k rögzített, ezért

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{k-1}{N}\right) \bigg/ \left(1 - \frac{\lambda}{N}\right)^k = 1.$$

Így a fentiekből következik, hogy

$$\lim_{N \rightarrow \infty} \binom{N}{k} p^k (1-p)^{N-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Az exponenciális függvény jólismert sorfejtése szerint

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = e^0 = 1.$$

Tehát az így kapott

$$p_k(\lambda) := \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

valószínűségek szintén valószínűségeloszlást alkotnak, amelyet *Poisson eloszlásnak* nevezünk, a λ számot pedig a Poisson eloszlás paraméterének.

Fentebb azt bizonyítottuk be, hogy ha $N \rightarrow \infty$ és $p \rightarrow 0$ olyan módon, hogy $Np = \lambda$ állandó, akkor az N -ed rendű és p paraméterű binomiális eloszlás k -adik tagja a λ paraméterű Poisson eloszlás k -adik tagjához konvergál minden rögzített $k = 0, 1, 2, \dots$ esetén. Gyakorlatilag ez azt jelenti, hogy ha N „elég nagy” és p „elég kicsi”, miközben $Np = \lambda$ állandó, akkor az N -ed rendű és p paraméterű binomiális eloszlás tagjai „jól közelíthetők” (approximálhatók) az Np paraméterű Poisson eloszlás megfelelő tagjaival:

$$p_k(N) \approx p_k(\lambda), \quad \text{ahol } \lambda = Np \text{ és } k = 0, 1, 2, \dots \text{ rögzített;}$$

vagy részletesen kiírva:

$$\binom{N}{k} p^k (1-p)^{N-k} \approx \frac{(Np)^k}{k!} e^{-Np}.$$

Megjegyezzük, hogy az $Np = \lambda > 0$ egyenlőség helyett elegendő lenne csak a

$$\lim_{N \rightarrow \infty} Np = \lambda > 0$$

határérték létezését feltételeznünk ahhoz, hogy a Poisson eloszlást a binomiális eloszlás határértékeként megkapjuk.

Példaként megemlítjük, hogy adott időintervallum, mondjuk a $(0, t)$, $t > 0$, időintervallum alatt elbomlott atomok száma egy radioaktív anyagból Poisson eloszlást követ. Ugyancsak Poisson eloszlású egy telefonközpontba adott időtartam alatt befutó hívások száma, vagy pedig egy nyári éjszakán adott időtartam alatt észlelhető meteor (a köznyelv szerint csillag) hullások száma.

4. Valószínűségek közelítő meghatározása Poisson eloszlással való approximációval

Visszatérünk arra a feladatra, amelyet az 1. részben kombinatorikai módszerrel már megoldottunk: Mi annak a valószínűsége, hogy egy n személyből álló csoportban van legalább két olyan ember, akik az év azonos napján ünneplik születésnapjukat. Ezt az eseményt ott A_n -nel jelöltük. A továbbiakban legyen $n \geq 4$.

Most a Poisson eloszlással való approximációval adunk jó közelítő értéket a $P(A_n)$ valószínűségre. Ezen célból az n személyből álló csoport tagjait 1-től n -ig terjedő számokkal címkézzük meg. Legyen $1 \leq i < j \leq n$ és jelöljük $A_{i,j}$ -vel azt az eseményt, hogy az i -edik és j -edik embernek az év ugyanazon napjára esik a születésnapja. Nyilvánvaló, hogy

$$P(A_{i,j}) = \frac{365 \cdot 1}{(365)^2} = \frac{1}{365}, \quad 1 \leq i < j \leq n.$$

Egyszerűen belátható, hogy az így definiált $\binom{n}{2}$ számú esemény páronként független, mivel

$$P(A_{i,j} \cdot A_{k,\ell}) = P(A_{i,j})P(A_{k,\ell}), \quad \text{ha } (i,j) \neq (k,\ell).$$

Két esemény $A_{i,j} \cdot A_{k,\ell}$ szorzatán együttes bekövetkezésüket értjük; az $(i,j) \neq (k,\ell)$ reláción pedig azt értjük, hogy vagy $i \neq k$ (míg j és ℓ tetszőleges), vagy $j = \ell$ (míg i és k tetszőleges). Például nyilvánvaló, hogy

$$\begin{aligned} P(A_{1,2} \cdot A_{3,4}) &= \frac{365 \cdot 1 \cdot 365 \cdot 1}{(365)^4} = \frac{1}{(365)^2} = \\ &= \frac{1}{365} \cdot \frac{1}{365} = P(A_{1,2})P(A_{3,4}). \end{aligned}$$

Hasonló módon láthatjuk be, hogy az $A_{1,2}$ és $A_{1,3}$ események is függetlenek; mivel

$$P(A_{1,2} \cdot A_{1,3}) = \frac{365 \cdot 1 \cdot 1}{(365)^3} = \frac{1}{(365)^2} = P(A_{1,2})P(A_{1,3}).$$

Viszont az $\{A_{i,j} : 1 \leq i < j \leq n\}$ események összességükben már nem függetlenek, amit úgy fejezünk ki, hogy az $\{A_{i,j}\}$ események *nem teljesen* függetlenek; mivel például

$$\begin{aligned} P(A_{1,2} \cdot A_{1,3} \cdot A_{2,3}) &= P(A_{1,2} \cdot A_{1,3}) = P(A_{1,2})P(A_{1,3}) = \\ &= \frac{1}{365} \cdot \frac{1}{365} \neq \frac{1}{365} \cdot \frac{1}{365} \cdot \frac{1}{365} = P(A_{1,2})P(A_{1,3})P(A_{2,3}). \end{aligned}$$

Azonban az $\{A_{i,j} : 1 \leq i < j \leq n\}$ események függősége „meglehetősen gyenge” (de ezen kijelentésnek a precíz kifejtése meghaladja jelen írásunk kereteit). Ennek következtében mégis állíthatjuk, hogy ha $N := \binom{n}{2}$ „elég nagy” és $p := \frac{1}{365}$ már „elég kicsinek” tekinthető, akkor az N -ed rendű és p paraméterű binomiális eloszlásnak „elég jó” közelítése lesz a

$$\lambda := Np = \binom{n}{2} P(A_{i,j}) = \frac{n(n-1)}{2} \frac{1}{365} = \frac{n(n-1)}{730}$$

paraméterű Poisson eloszlás.

Így az alábbi közelítések „jóknak” mondhatók:

$$\begin{aligned} P(\overline{A}_n) &= P\left(\begin{array}{l} \text{az } n \text{ személyből álló csoportban nincs két olyan ember,} \\ \text{akik ugyanazon a napon ünneplik születésnapjukat} \end{array}\right) \approx \\ &\approx \binom{N}{0} p^0 (1-p)^N \approx \frac{\lambda^0}{0!} e^{-\lambda} = \exp\left\{-\frac{n(n-1)}{730}\right\}. \end{aligned}$$

Az utóbbi közelítésre támaszkodva határozzuk meg azt a legkisebb n természetes számot, amelyre a $P(\overline{A}_n)$ valószínűség fenti közelítése $< \frac{1}{2}$. Nyilván az

$$\exp\left\{-\frac{n(n-1)}{730}\right\} := e^{-n(n-1)/730} < \frac{1}{2}$$

egyenlőtlenség azzal ekvivalens, hogy

$$\exp\left\{\frac{n(n-1)}{730}\right\} > 2.$$

Mindkét oldal természetes e alapú logaritmusát véve nyerjük, hogy

$$n(n-1) > 730 \ln 2 \approx 505,9974\dots,$$

amely egyenlőtlenség $n = 23$ esetén teljesül, de $n = 22$ esetén még nem teljesül. Tehát a

$$P(A_{23}) = 1 - P(\overline{A}_{23}) \approx 1 - \exp \left\{ -\frac{23 \cdot 22}{730} \right\}$$

közelítés jobboldala már nagyobb $\frac{1}{2}$ -nél. Ugyanezt az $n = 23$ értéket kaptuk a kombinatorikai módszerrel is az 1. részben.

A binomiális eloszlás Poisson eloszlással való approximációja alkalmas arra, hogy bonyolultabb események valószínűségét is „jó közelítéssel” tudjuk meghatározni. Például ilyen feladat a következő. Mi a valószínűsége annak az eseménynek, hogy n személyből álló csoportban nincs 3 olyan ember, akik születésnapja az évnek ugyanazon napjára esik.

Ennek megválaszolása céljából, legyen $1 \leq i < j < k \leq n$ és jelöljük $B_{i,j,k}$ -val azt az eseményt, hogy az i -vel, j -vel és k -val jelölt emberek az év azonos napján ünneplik születésnapjukat. Nyilvánvaló, hogy

$$P(B_{i,j,k}) = \frac{365 \cdot 1 \cdot 1}{(365)^3} = \frac{1}{(365)^2}, \quad 1 \leq i < j < k \leq n.$$

Most tehát az $N := \binom{n}{3}$ -ad rendű és $p := \frac{1}{(365)^2}$ paraméterű binomiális eloszlás lép fel, amelyet a

$$\lambda := Np = \binom{n}{3} \frac{1}{(365)^2} = \frac{n(n-1)(n-2)}{3!(365)^2} = \frac{n(n-1)(n-2)}{799\,350}$$

paraméterű Poisson eloszlással approximálhatunk feltéve, hogy N „elég nagy” és p „elég kicsinek” tekinthető.

Jelöljük B_n -nel azt az eseményt, hogy az n személyből álló csoportban van legalább 3 olyan ember, akiknek az év azonos napjára esik a születésnapjuk. A fentiek szerint az ellentett $\overline{B}_n$ esemény valószínűségére az alábbi approximáció adódik:

$$\begin{aligned} P(\overline{B}_n) &= P \left(\begin{array}{l} \text{a csoportban nincs három olyan ember, akik} \\ \text{születésnapja az év azonos napjára esik} \end{array} \right) \approx \\ &\approx \binom{N}{0} p^0 (1-p)^N \approx \frac{\lambda^0}{0!} e^{-\lambda} = \exp \left\{ -\frac{n(n-1)(n-2)}{799\,350} \right\}. \end{aligned}$$

Ezen egyenlőtlenség jobboldala akkor lesz $\frac{1}{2}$ -nél kisebb, ha

$$n(n-1)(n-2) > (799\,350) \ln 2 \approx 554\,067,19 \dots$$

Mivel

$$84 \cdot 83 \cdot 82 = 571\,740 \text{ és } 83 \cdot 82 \cdot 81 = 551\,286,$$

ezért $n = 84$ a legkisebb olyan egészszám, amelyre a $P(\overline{B}_n)$ valószínűség fenti approximációja kisebb, mint $\frac{1}{2}$. Ezért legalább $\frac{1}{2}$ a közelítő valószínűsége annak, hogy egy 84 személyből álló csoportban legalább 3 embernek az év azonos napjára esik a születésnapja. Megjegyezzük, hogy a lényegesen bonyolultabb kombinatorikai megfontolások is ugyanerre az eredményre vezetnek.

További két érdekes példán mutatjuk még be a Poisson eloszlással való approximáció hatékonyságát.

Két példa

1. Magyarországon két évente kb. 80 000 házasságkötés történik. Jelöljük C_1 -gyel azt az eseményt, hogy ezen házasságot kötő párok közül legalább egy olyan pár található, ahol az ifjú férj és feleség egyaránt az év egy adott napján, mondjuk április 30-án született. Nyilván

$$P(C_1) = \frac{1 \cdot 1}{(365)^2} = \frac{1}{133\,225}.$$

Most az $N := 80\,000$ -ed rendű és $p := \frac{1}{(365)^2}$ paraméterű binomiális eloszlást a

$$\lambda := Np = \frac{80\,000}{(365)^2} = 0,600\,487\,896\,4\dots$$

paraméterű Poisson eloszlással approximáljuk (azon feltevés mellett, hogy N „élég nagynak” és p „élég kicsinek” tekinthető). Tehát írhatjuk, hogy

$$\begin{aligned} P(\overline{C}_1) &= P\left(\begin{array}{l} \text{nincs olyan ifjú házaspár, amikor a férj és feleség} \\ \text{születésnapja egyaránt április 30-ára esik} \end{array}\right) \approx \\ &\approx \binom{N}{0} p^0 (1-p)^N = \left(1 - \frac{1}{(365)^2}\right)^{80\,000} \approx \\ &\approx \frac{\lambda^0}{0!} e^{-\lambda} = \exp\left\{-\frac{80\,000}{(365)^2}\right\} = \exp\left\{-\frac{80\,000}{133\,225}\right\} = \\ &= \exp(-0,600\,487\,896\,4\dots). \end{aligned}$$

Ebből következik, hogy

$$P(C_1) = 1 - P(\overline{C}_1) \approx 1 - e^{-0,6000} = 0,45118 \approx 0,4512.$$

Tehát 45% körül van annak az esélye, hogy 80 000 házaspár esetén a férj és a feleség egyaránt az évnek egy megadott napján született.

2. Szegeden évente kb. 800 házasságot kötnek. Jelöljük C_2 -vel azt az eseményt, hogy ezen házasságot kötő párok között legalább egy olyan pár található, ahol a

férj és feleség születésnapja *ez év azonos* (de nem egy meghatározott) *napjára* esik. Nyilván

$$P(C_2) = \frac{365 \cdot 1}{(365)^2} = \frac{1}{365}.$$

Most az $N := 800$ és $p := \frac{1}{365}$ paraméterű binomiális eloszlást a

$$\lambda := Np = \frac{800}{365} = 2,191\,780\,6\dots$$

paraméterű Poisson eloszlással approximáljuk (azon feltevés mellett, hogy N „élég nagynak” és p „élég kicsinek” tekinthető). Tehát írhatjuk, hogy

$$\begin{aligned} P(\overline{C}_2) &= P\left(\begin{array}{l} \text{nincs olyan ifjú házaspár, amikor a férj és feleség} \\ \text{születésnapja az év azonos napjára esik} \end{array}\right) \approx \\ &\approx \binom{N}{0} p^0 (1-p)^N = \left(1 - \frac{1}{365}\right)^{800} \approx \frac{\lambda^0}{0!} e^{-\lambda} = \\ &= \exp\left\{-\frac{800}{365}\right\} = \exp(-2,191\,780\,6\dots). \end{aligned}$$

Ebből következik, hogy

$$P(C_2) = 1 - P(\overline{C}_2) \approx 1 - e^{-2,1918} = 0,88828\dots \approx 0,8883.$$

Tehát elég magas, körülbelül 89 % annak az esélye, hogy 800 házaspár esetén a férj és feleség az év azonos napján született.

5. Várható érték és szórás definíciója és alapvető tulajdonságai

Egy X *véletlen* (azt is mondhatjuk, hogy valószínűségi) *változón* olyan mennyiséget értünk, amelynek értéke a véletlentől függ, és amelyre a véletlentől való függést az alábbi módok egyikén ismerjük:

(i) Ha X csak *véges vagy megszámlálhatóan végtelen sok* (azaz sorozatba rendezhető) $x_0, x_1, x_2, \dots$ értéket vehet fel, akkor ismerjük a

$$p_k := P(X = x_k), \quad k = 0, 1, 2, \dots$$

valószínűségeket. Az ilyen X -et *diszkrét* véletlen változónak nevezzük.

(ii) Ha X nem diszkrét értékeket vesz fel (az ilyen X -et *folytonos eloszlású* véletlen változónak is szokás nevezni), akkor ismerjük az

$$F(x) := P(X \leq x), \quad -\infty < x < \infty,$$

ún. *eloszlásfüggvényt*.

Írásunk további részében csak diszkrét véletlen változókkal foglalkozunk. Ezért a következőkben csak a diszkrét véletlen változók várható értékének és szórásának definícióját ismertetjük. Ha az X véletlen változó lehetséges értékei $x_0, x_1, x_2, \dots$ és ezeket rendre a $p_0, p_1, p_2, \dots$ valószínűségekkel veszi fel, ahol

$$p_k > 0 \text{ és } \sum_k p_k = 1,$$

akkor az X várható értékét az

$$E(X) := \sum_k p_k x_k$$

összeggel definiáljuk, a következő kiegészítéssel. Ha X végtelen sok x_k értéket vesz fel, akkor megköveteljük, hogy a definícióban szereplő végtelen sor abszolút konvergencia legyen, azaz

$$\sum_{k=0}^{\infty} p_k |x_k| < \infty.$$

Ha ez a feltétel nem teljesül, akkor az X várható értékét nem értelmezzük. Ezen feltételre azért van szükség, mert az X véletlen változó lehetséges értékeinek sorozatba történő elrendezése önkényes. Márpedig *Riemann* német matematikus egyik *tétele* szerint, ha egy végtelen sor csak konvergencia, de nem abszolút konvergencia, akkor a tagok alkalmas átrendezésével tetszőleges értéket állíthatunk elő, sőt a végtelen sort akár divergencia is tehetjük. Az $E(X)$ várható értéket tehát csak abban az esetben értelmezzük, ha $\sum p_k |x_k| < \infty$; különben azt mondjuk, hogy X -nek nem létezik várható értéke.

Megjegyezzük, hogy a várható érték jelölésére használt „ E ” betű az angol „expectation” szó kezdőbetűje, amelynek magyar jelentése: elvárás, elvárható eredmény.

Röviden összefoglaljuk a várható érték legfontosabb tulajdonságait:

(i) Ha $X \equiv c$ = állandó, akkor $E(X) = c$.

(ii) Ha X várható értéke létezik és c tetszőleges állandó, akkor cX várható értéke is létezik, és

$$E(cX) = cE(X).$$

(iii) Ha X_1 és X_2 várható értéke létezik, akkor $X_1 + X_2$ összegük várható értéke is létezik, és

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

(iv) Ha $c_1, c_2, \dots, c_n$ állandók és $X_1, X_2, \dots, X_n$ olyan véletlen változók, amelyeknek létezik a várható értékük, akkor a $\sum_{k=1}^n c_k X_k$ várható értéke is létezik, és

$$E\left(\sum_{k=1}^n c_k X_k\right) = \sum_{k=1}^n c_k E(X_k).$$

(v) Ha X_1 és X_2 független véletlen változók, amelyeknek létezik a várható értékük, akkor az $X_1 X_2$ szorzat várható értéke is létezik, és

$$E(X_1 X_2) = E(X_1)E(X_2).$$

Megemlítjük, hogy ha az X véletlen változónak létezik a várható értéke, akkor $E(X)$ az a szám, amely körül az X ún. empirikus közepei ingadoznak. Pontosabban szólva, ha $X, X_1, X_2, X_3, \dots$ független, azonos eloszlású véletlen változók és az $E(X)$ várható érték létezik (ekkor az $E(X_k)$ várható értékek minden k -ra léteznek és az $E(X)$ -szel egyenlők), akkor

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow E(X), \quad n \rightarrow \infty.$$

Az ilyen típusú tételeket a *nagy számok törvényeinek* nevezik a valószínűségszámításban.

Az X véletlen változó szórását a

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

képlettel definiáljuk, feltéve, hogy a négyzetgyök jel alatti várható érték létezik. Alább, a (iii)-vel jelzett részben kiderül, hogy az $(X - E(X))^2$ véletlen változónak akkor és csak akkor létezik a várható értéke, ha X^2 -nek létezik az $E(X^2)$ várható értéke, és ez a feltétel egyben $E(X)$ létezését is biztosítja.

Szemléletesen szólva, a szórás annak mértéke, hogy az X véletlen változó értékei négyzetes eltérésben számolva milyen mértékben ingadoznak (szóródnak) az $E(X)$ várható érték körül

Megjegyezzük, hogy a szórás jelölésére használt „ D ” betű az angol „deviation” szó kezdőbetűje, amelynek magyar jelentése: (valamitől való) eltérés, elhajlás; míg egyesek az angol „dispersion” szót is említik, amelynek magyar jelentése: (szét)szóródás.

Röviden összefoglaljuk a szórás alapvető tulajdonságait:

(i) Ha $X \equiv c$ =állandó, akkor $D(X) = 0$.

(ii) Ha X szórása létezik és c tetszőleges állandó, akkor cX szórása is létezik, és

$$D(cX) = |c|D(X).$$

(iii) Ha X szórása létezik, akkor

$$\begin{aligned} D^2(X) &= E[(X - E(X))^2] = E[X^2 - 2XE(X) + E^2(X)] = \\ &= E(X^2) - 2E(X)E(X) + E^2(X) = E(X^2) - E^2(X). \end{aligned}$$

Az így levezetett

$$D^2(X) = E(X^2) - E^2(X)$$

formula speciális esete az

$$E[(X - c)^2] = D^2(X) + (E(X) - c)^2$$

ún. *Steiner formulának*, ahol c tetszőleges állandó. Mivel a $D^2(X)$ szórásnégyzet mindig nemnegatív, ezért a fentebb levezetett formulából egyből következik, hogy

$$E^2(X) \leq E(X^2).$$

Speciálisan ez azt is jelenti, hogy ha $E(X^2)$ létezik, akkor $E(X)$ is létezik (amint ezt már korábban említettük).

(iv) Ha X_1 és X_2 független véletlen változók, amelyeknek létezik a szórásuk, akkor összegük szórása is létezik, és

$$D^2(X_1 + X_2) = D^2(X_1) + D^2(X_2).$$

Ugyanis, hogy ha X_1 és X_2 független valószínűségi változók, amelyeknek létezik a szórásuk, akkor

$$E[(X_1 - E(X_1))(X_2 - E(X_2))] = E[X_1 - E(X_1)]E[X_2 - E(X_2)] = 0,$$

és ennek következtében az is fennáll, hogy

$$\begin{aligned} D^2(X_1 + X_2) &= E[(X_1 + X_2 - E(X_1 + X_2))^2] = \\ &= E[(X_1 - E(X_1) + X_2 - E(X_2))^2] = \\ &= E[(X_1 - E(X_1))^2] - 2E[(X_1 - E(X_1))(X_2 - E(X_2))] + E[(X_2 - E(X_2))^2] = \\ &= E[(X_1 - E(X_1))^2] + E[(X_2 - E(X_2))^2] = \\ &= D^2(X_1) + D^2(X_2). \end{aligned}$$

(v) Ha $X_1, X_2, \dots, X_n$ páronként független véletlen változók, amelyeknek létezik a szórásuk, akkor összegük szórása is létezik, és

$$D^2\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n D^2(X_k).$$

Megjegyezzük, hogy a $D^2(X)$ neve *szórásnégyzet* (angolul: „variance”; míg a $D(X)$ szórás angol neve: „standard deviation”).

A fentiek szerint, ha X diszkrét véletlen változó, amely az $x_0, x_1, x_2, \dots$ értékeket rendre a $p_0, p_1, p_2, \dots$ valószínűségekkel veszi fel, akkor a $D^2(X)$ szórásnégyzetet a

$$D^2(X) = \sum_k p_k x_k^2 - \left(\sum_k p_k x_k\right)^2$$

képlettel tudjuk kiszámítani, feltéve, hogy

$$\sum_k p_k x_k^2 < \infty;$$

különben X -nek nem létezik szórása. A $D(X)$ szórás pedig $D^2(X)$ -nek a nemnegatív négyzetgyöke.

Legyen A egy tetszőleges esemény. Az

$$I_A := \begin{cases} 1, & \text{ha } A \text{ bekövetkezik,} \\ 0, & \text{ha } A \text{ nem következik be (azaz, ha } \bar{A} \text{ következik be);} \end{cases}$$

képlettel definiált véletlen változót az A esemény *indikátorának* nevezzük. Ennek várható értéke

$$E(I_A) = 1 \cdot P(A) + 0 \cdot P(\bar{A}) = P(A),$$

és mivel

$$E(I_A^2) = E(I_A) = P(A),$$

ezért

$$D(I_A) = \sqrt{P(A) - P^2(A)} = \sqrt{P(A)(1 - P(A))}.$$

Tanulságos kiszámítanunk az N -ed rendű és p paraméterű binomiális eloszlású X véletlen változó várható értékét:

$$\begin{aligned} E(X) &= \sum_{k=0}^N p_k x_k = \sum_{k=0}^N k \binom{N}{k} p^k (1-p)^{N-k} = \\ &= Np \sum_{k=1}^N \binom{N-1}{k-1} p^{k-1} (1-p)^{N-k} = \\ &= Np \sum_{\ell=0}^{N-1} \binom{N-1}{\ell} p^{\ell} (1-p)^{N-1-\ell} = Np[p + (1-p)]^{N-1} = Np. \end{aligned}$$

Hasonlóképpen számolva nyerjük, hogy

$$\begin{aligned}
 E(X^2) &= \sum_{k=0}^N k^2 \binom{N}{k} p^k (1-p)^{N-k} = Np \sum_{k=1}^N k \binom{N-1}{k-1} p^{k-1} (1-p)^{N-k} = \\
 &= Np \sum_{k=1}^N [(k-1) + 1] \binom{N-1}{k-1} p^{k-1} (1-p)^{N-k} = \\
 &= Np[(N-1)p \sum_{k=2}^N \binom{N-2}{k-2} p^{k-2} (1-p)^{N-k} + \\
 &\quad + \sum_{k=1}^N \binom{N-1}{k-1} p^{k-1} (1-p)^{N-k}] = Np[(N-1)p + 1].
 \end{aligned}$$

Tehát a fenti binomiális eloszlású X véletlen változó szórása

$$D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{Np - Np^2} = \sqrt{Np(1-p)}.$$

A λ paraméterű Poisson eloszlású X véletlen változó várható értéke

$$\begin{aligned}
 E(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \\
 &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.
 \end{aligned}$$

Hasonlóképpen kapjuk, hogy

$$\begin{aligned}
 E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \\
 &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} [(k-1) + 1] \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \left(\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) = \\
 &= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda + 1),
 \end{aligned}$$

és így szórása

$$D(X) = \sqrt{\lambda(\lambda + 1) - \lambda^2} = \sqrt{\lambda}.$$

6. Az azonos napra eső születésnapokkal kapcsolatos véletlen változók várható értéke és szórása

Ezen előzmények után tekintsünk egy 80 személyből álló csoportot (azért választottuk a csoport létszámot ismét 80-nak, mivel az 1. Táblázat szerint éppen 80 ember esetén állapíthatjuk először 99,99 % biztonsággal, hogy a csoportban legalább két embernek az év azonos napjára esik a születésnapja). A következő feladat megoldását tűzzük ki célul. Határozzuk meg az év azon napjainak várható értékét (számát), amelyek mindegyikére a fenti csoportból *pontosan* k embernek esik a születésnapja. Speciálisan, $k = 0$ esetén ez az év azon napjainak várható értékét jelenti, amelyre a csoportból senkinek sem esik születésnapja.

Jelöljük X_k -val az év azon napjainak a számát, amelyek mindegyikére a 80 főből álló csoportból pontosan k embernek esik a születésnapja; és jelöljük $C_{k,\ell}$ -el azt az eseményt, hogy az év ℓ -edik napjára pontosan k embernek esik a születésnapja. Itt $k = 0, 1, 2, \dots, 80$ lehet (de mi csak a $k = 0, 1, 2, 3, 4$ eseteket vizsgáljuk); míg $\ell = 1, 2, 3, \dots, 365$ lehet. Nyilván

$$X_k = \sum_{\ell=1}^{365} I_{C_{k,\ell}} \text{ és } E(X_k) = \sum_{\ell=1}^{365} E(I_{C_{k,\ell}}).$$

Mivel minden ℓ -re,

$$P(C_{k,\ell}) = \binom{80}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{80-k},$$

ezért

$$E(X_k) = 365 \binom{80}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{80-k}.$$

Például, $k = 2$ esetén a számítást az alábbi módon rendeztük el:

$$\begin{aligned} E(X_k) &= 365 \binom{80}{2} \left(\frac{1}{365}\right)^2 \left(\frac{364}{365}\right)^{78} = \frac{80 \cdot 79}{2} \frac{(364)^{78}}{(365)^{79}} = \\ &= \left(\frac{364}{365}\right)^{81} \frac{365}{364} \cdot \frac{365}{364} \cdot \frac{80 \cdot 79}{2 \cdot 364} = 6,9897 \end{aligned}$$

négy tizedesjegyre kerekítve. A 81-edik hatvány kiszámítására az

$$\left(\left(\left(\left(\frac{364}{365} \right)^3 \right)^3 \right)^3 \right)^3 = 0,800\,737\,291\,307\,157\,054\,769\,545\,826\,605$$

érték adódott, amely 30 tizedesjegyre pontos. A 2. Táblázatban tüntettük fel a számítások eredményét a $k = 0, 1, 2, 3$ és 4 esetekre.

k	$P(C_{k,\ell})$	$E(X_k)$
0	0,8029	293,0702
1	0,1765	64,4143
2	0,0191	6,9897
3	0,0014	0,4993
4	0,0001	0,0264
Összesen	1,0000	364,9999

2. Táblázat

k	$P(C_{k,\ell})$	$E(X_k)$
0	0,8032	293,1602
1	0,1760	64,2543
2	0,0193	7,0416
3	0,0014	0,5145
4	0,0001	0,0282
Összesen	1,0000	364,9988

3. Táblázat

Ha az előző számításban előforduló $N = 80$ és $p = 1/365$ paraméterű binomiális eloszlást a

$$\lambda := Np = \frac{80}{365}$$

paraméterű Poisson eloszlással approximáljuk (azon feltevés mellett, hogy $N = 80$ „elég nagy” és $p = \frac{1}{365}$ pedig „elég kicsinek” tekinthető) akkor az alábbi közelítéseket kapjuk:

$$P(C_{k,\ell}) \approx \frac{\lambda^k}{k!} e^{-\lambda} = \frac{1}{k!} \left(\frac{80}{365} \right)^k \exp \left(-\frac{80}{365} \right)$$

és ennek megfelelően írhatjuk, hogy

$$E(X_k) = \sum_{k=1}^{365} P(C_{k,\ell}) \approx \frac{365}{k!} \left(\frac{80}{365} \right)^k \exp \left(-\frac{80}{365} \right), \quad k = 0, 1, 2, 3 \text{ és } 4.$$

A Poisson approximációval kapott közelítő értékeket a 3. Táblázatban tüntettük fel; amelyek egészen jó megegyezést mutatnak a 2. Táblázatbeli értékekkel.

Megjegyezzük, hogy a $k = 5$ esetre a Poisson approximáció használatával az alábbi közelítéseket kaptuk:

$$P(C_{5,\ell}) = 0,000\,003\,0\dots \quad \text{és} \quad E(X_5) = 0,001\,102\,4\dots$$

Így annak az eseménynek a valószínűsége, hogy az év ℓ -edik napjára pontosan öt (vagy akár ennél több) ember születésnapja esik a 80 személyből álló csoportban, kisebb, mint $0,5 \cdot 10^{-5}$; tehát ennek bekövetkezésére gyakorlatilag már nem

számíthatunk. Ugyanígy azon napok számának várható értéke sem számottevő gyakorlatilag, amelyeken legalább 5 ember ünnepli születésnapját.

Ez utóbbi megállapításunkat támasztja alá az is, hogy például a 3. Táblázat harmadik oszlopát figyelembe véve, a 80 ember születésnapjai számának várható értéke a következőnek adódik:

$$\begin{aligned} 64, 2543 \cdot 1 + 7, 0416 \cdot 2 + 0, 5145 \cdot 3 + 0, 0282 \cdot 4 = \\ = 64, 2543 + 14, 0832 + 1, 5435 + 0, 1128 = 79, 9938 \approx 80. \end{aligned}$$

Természetesen ez csak egy kvalitatív ellenőrzést jelent, de a megegyezés mégis rendkívül figyelemre méltó. Hasonló megegyezést kapunk akkor is, ha a 2. Táblázat harmadik oszlopában levő értékekkel számolunk.

A következőkben az

$$X_k = \sum_{\ell=1}^{365} I_{C_{k,\ell}}, \quad k = 0, 1, 2, 3, 4,$$

véletlen változók szórásnégyzetét, illetve szórását fogjuk meghatározni a

$$D^2(X_k) := E[(X_k - E(X_k))^2] = E(X_k^2) - E^2(X_k)$$

Steiner képlettel. Fentebb már kiszámítottuk az $E(X_k)$ várható értékeket a $k = 0, 1, 2, 3, 4$ esetekben (lásd a 2. Táblázatot). Hátra van még az $E(X_k^2)$ várható értékek kiszámítása. A négyzetreemelés elvégzése után azt kapjuk, hogy

$$E(X_k^2) = \sum_{\ell=1}^{365} E(I_{C_{k,\ell}}^2) + \sum_{\substack{\ell=1 \\ \ell \neq \ell'}}^{365} \sum_{\ell'=1}^{365} E(I_{C_{k,\ell}} \cdot I_{C_{k,\ell'}}).$$

A jobb oldalon az első összeg egyszerűen számítható, mivel

$$E(I_{C_{k,\ell}}^2) = E(I_{C_{k,\ell}}) = P(C_{k,\ell}) = \binom{80}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{80-k}, \quad k = 0, 1, 2, 3, 4.$$

Viszont a jobboldali kettős összegben szereplő várható érték kiszámítása a következő kombinatorikai megfontolást igényli. A kedvező esetek összeszámlálása végett a 80 főből kiválasztunk $2k$ számút, majd ezek közül kiválasztunk k számút, akik az év ℓ -edik napján, míg a másik k számú fő az év ℓ' -edik napján született; a többi $(80 - k)$ fő pedig az év többi 363 napjának bármelyikén születhetett. Ezek alapján írhatjuk, hogy

$$\begin{aligned} E(I_{C_{k,\ell}} \cdot I_{C_{k,\ell'}}) &= P(C_{k,\ell} \cdot C_{k,\ell'}) = \binom{80}{2k} \binom{2k}{k} \frac{\overbrace{1 \cdots 1}^k \cdot \overbrace{1 \cdots 1}^k \cdot (363)^{80-2k}}{(365)^{80}} = \\ &= \binom{80}{2k} \binom{2k}{k} \left(\frac{1}{365}\right)^{2k} \left(\frac{363}{365}\right)^{80-2k}, \quad k = 0, 1, 2, 3, 4. \end{aligned}$$

Megjegyezzük, hogy az a kiválasztási mód is ugyanerre az eredményre vezet, amikor a 80 főből először csak k számút, majd a maradék $(80 - k)$ főből ismét k számút választunk ki, mivel

$$\binom{80}{k} \binom{80-k}{k} = \binom{80}{2k} \binom{2k}{k}.$$

Nyilvánvaló, hogy tetszőleges $\ell, \ell', \ell'', \ell'''$ esetén

$$P(C_{k,\ell}) = P(C_{k,\ell'}), \ell \neq \ell';$$

$$P(C_{k,\ell} \cdot C_{k,\ell'}) = P(C_{k,\ell''} \cdot C_{k,\ell'''}), \ell \neq \ell', \ell'' \neq \ell'''.$$

Eddigi meggondolásainkat összefoglalva kapjuk, hogy

$$\begin{aligned} E(X^2) &= 365P(C_{k,\ell}) + 365 \cdot 364P(C_{k,\ell} \cdot C_{k,\ell'}) = \\ &= 365 \binom{80}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{80-k} + \\ &+ 365 \cdot 364 \binom{80}{2k} \binom{2k}{k} \left(\frac{1}{365}\right)^{2k} \left(\frac{363}{365}\right)^{80-2k} =: \\ &=: F_k + G_k, \quad k = 0, 1, 2, 3, 4. \end{aligned}$$

Számításainkat a 4. Táblázatban foglaltuk össze.

k	F_k	G_k	$E(X_k^2)$	$E^2(X_k)$	$D^2(X_k)$	$D(X_k)$
0	293	85 604	85 897	85 890	7	2, 65
1	64	4 137	4 201	4 149	52	7, 21
2	6, 97	47, 39	54, 36	48, 85	5, 51	2, 35
3	0, 4993	0, 1138	0, 6131	0, 2493	0, 3638	0, 60
4	0, 0365	0, 0002	0, 0368	0, 0007	0, 0361	0, 19

4. Táblázat

Emlékeztetünk arra, hogy X_k -val az év azon napjainak a számát jelöltük, amelyek mindegyikére a 80 főből álló csoportból pontosan k embernek esik a születésnapja. A 4. Táblázat szerint a szórás a $k = 1$ esetben lényegesen nagyobb, mint a többi esetben. Az X_1 véletlen változó átlagos négyzetes eltérése (ingadozása) az $E(X_1) \approx 64$ nap várható érték körül $D(X_1) \approx 7$ nap. Meglepő, hogy az X_0 véletlen változó szórása $D(X_0) \approx 3$ nap lényegesen kisebb, mint az X_1 szórása. Az is érdekes továbbá, hogy az X_2 véletlen változó $D(X_2) \approx 2$ nap szórása alig kisebb, mint az X_0 szórása.

Felhívjuk az olvasó figyelmét arra, hogy a $C_{k,\ell}$ és $C_{k,\ell'}$ események nem függetlenek. Például, a $k = 0$ esetben

$$P(C_{0,\ell}) = \left(\frac{364}{365}\right)^{80},$$

míg

$$P(C_{0,\ell} \cdot C_{0,\ell'}) = \left(\frac{363}{365}\right)^{80}, \quad \ell \neq \ell'.$$

Tekintve, hogy

$$363 \cdot 365 < (364)^2,$$

ezért

$$P(C_{0,\ell} \cdot C_{0,\ell'}) < P(C_{0,\ell})P(C_{0,\ell'}) = P^2(C_{0,\ell}).$$

Ennek következtében nem tudtuk alkalmazni azt a 2. részben a szórásra vonatkozó (iv) alaptulajdonságot, amely szerint páronként független véletlen változók összegének szórásnégyzete egyenlő az egyes változók szórásnégyzeteinek összegével (természetesen feltéve, hogy a szóban forgó véletlen változók mindegyikének létezik a szórása; márpedig a fenti esetünkben ez a helyzet).

Ezért tanulságos annak végiggondolása, hogy ha a $C_{k,\ell}$ és $C_{k,\ell'}$ eseményeket tévesen függetleneknek fogtuk volna fel, és a független véletlen változókra érvényes fenti tétel alapján a

$$D^2(X_k) = 365D^2(I_{C_{k,\ell}}) = 365P(C_{k,\ell})(1 - P(C_{k,\ell}))$$

képlettel számoltunk volna, akkor például a $k = 0$ esetben

$$D^2(X_0) = 365 \left(\frac{364}{365}\right)^{80} \left(1 - \left(\frac{364}{365}\right)^{80}\right) = 61,8883 \dots$$

szórásnégyzetet, és ebből a

$$D(X_0) = 7,87$$

szórást kaptuk volna, ami lényegesen eltér a 4. Táblázatbeli 2,65 értéktől.

7. Markov és Csebisev egyenlőtlenségei

Az 5. részben tárgyalt X_0, X_1, X_2, X_3 és X_4 véletlen változók értékei csak nemnegatív egészs számok lehetnek, és ott kiszámítottuk ezek várható értékét és szórását. Ezek alapján szeretnénk olyan $[a_k, b_k]$ intervallumokat megadni, hogy fennálljon a

$$P(a_k \leq X_k \leq b_k) > p, \quad k = 0, 1, 2, 3, 4,$$

egyenlőtlenség, ahol p adott valószínűség, például $p = 0,9$. Ezen célból bizonyítjuk be a fentebbi címben jelzett egyenlőtlenségeket.

Markov egyenlőtlensége. Ha X nemnegatív véletlen változó, amelynek létezik a várható értéke, akkor tetszőleges $c > 0$ szám esetén

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

Ezt bizonyítandó, legyenek az X lehetséges értékei $x_0, x_1, x_2, \dots$, amelyeket rendre a $p_0, p_1, p_2, \dots$ valószínűségekkel vesz fel. Egyszerű becsléssel adódik, hogy

$$E(X) := \sum_k p_k x_k \geq \sum_{k: x_k \geq c} p_k x_k \geq c \sum_{k: x_k \geq c} p_k = cP(X \geq c),$$

amelyből Markov egyenlőtlenségét c -vel való osztással kapjuk.

Csebisev egyenlőtlensége. Ha az X véletlen változónak létezik a szórása, akkor tetszőleges $c > 0$ szám esetén

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

Ezt úgy bizonyítjuk, hogy Markov egyenlőtlenségét az

$$Y := (X - E(X))^2$$

véletlen változóra és c helyett c^2 -re alkalmazzuk. A levezetés tehát a következő:

$$\begin{aligned} P(|X - E(X)| \geq c) &= P((X - E(X))^2 \geq c^2) = P(Y \geq c^2) \leq \frac{E(Y)}{c^2} = \\ &= \frac{E[(X - E(X))^2]}{c^2} = \frac{D^2(X)}{c^2}. \end{aligned}$$

Csebisev egyenlőtlenségének jelentőségét az adja, hogy ennek segítségével becsülni tudjuk az X véletlen változó $E(X)$ várható értéke körüli ingadozásokat a $D^2(X)$ szórásnégyzet ismeretében.

Példaképpen, tekintsük az 6. részben bevezetett X_0 véletlen változót, amellyel az év azon napjainak a számát jelöltük, amelyek egyikére sem esik a 80 főből álló csoport tagjainak születésnapja. A 2. és 4. Táblázatból véve az $E(X_0)$ és $D^2(X_0)$ értékeket, Csebisev egyenlőtlensége szerint fennáll a

$$P(|X_0 - 293,07| \geq c) \leq \frac{7}{c^2}, \quad c > 0,$$

egyenlőtlenség. A $c = 4$ választással kapjuk, hogy

$$P(|X_0 - 293,07| \geq 4) \leq \frac{7}{16} < \frac{1}{2}.$$

A komplementer eseményre áttérve írhatjuk, hogy

$$P(|X_0 - 293,07| < 4) \geq 1 - \frac{7}{16} = \frac{9}{16} > \frac{1}{2}.$$

Mivel az

$$|X_0 - 293,07| < 4$$

abszolút értéket tartalmazó egyenlőtlenség a

$$-4 < X_0 - 293,07 < 4,$$

$$289,07 < X_0 < 297,07$$

egyenlőtlenségpárokkal ekvivalens, és mivel X_0 értékei csak egészszámok lehetnek, ezért a

$$290 \leq X_0 \leq 297$$

egyenlőtlenségpár fennállásának a valószínűsége $\frac{1}{2}$ -nél nagyobb. Ezt úgyis kifejezhetjük, hogy $X_0 \in [290; 297]$.

A $c = 9$ választás esetén, ugyanezen megfontolással azt kapjuk, hogy $X_0 \in [283; 302]$ fennállásának valószínűsége

$$P(283 \leq X_0 \leq 302) = 1 - \frac{7}{81} = \frac{74}{81} = 0,9135 \dots > 0,9.$$

Az utóbb kapott eredményt a következőképpen értelmezhetjük. Ha az X_0 véletlen változó értékére 10 független megfigyelést végzünk; azaz, ha az ország különböző vidékein olyan 80 főből álló csoportokban vizsgáljuk meg X_0 értékét, amely csoportok között semmiféle kapcsolat nincs; akkor azt találjuk, hogy általában 10 csoport közül 9 esetben $X_0 \in [283; 302]$ és csak 1 esetben fordulhat elő az, hogy X_0 értéke nem esik bele ebbe az intervallumba.

A fentihez hasonló számításokat végeztük el az X_1, X_2, X_3 és X_4 változók esetében is. A kapott eredményeket az 5. Táblázatban foglaljuk össze. A táblázat második oszlopában tüntetjük fel a c_k szám értékét, amellyel Csebisev egyenlőtlenségét alkalmaztuk a X_k változó esetében; míg a táblázat harmadik oszlopában tüntetjük fel azt az $[a_k; b_k]$ intervallumot, amelybe az X_k változó értéke a negyedik oszlopban megadott valószínűséggel esik.

k	c_k	$[a_k; b_k]$	$P(X_k \in [a_k; b_k])$
0	9	$[283; 302]$	0,9135
1	23	$[42; 87]$	0,9018
2	8	$[2; 14]$	0,9139
3	2	$[0; 2]$	0,9090
4	0,7	0	0,9263

5. Táblázat

Az X_3 változó esetén Csebisev egyenlőtlensége a $c_3 = 2$ választással azt adja, hogy az

$$a_3 := -1,5007 < X_3 < 2,4993 =: b_3$$

esemény $p_3 = 0,9090$ valószínűséggel következik be. Tekintve, hogy X_3 értékei csak nemnegatív egészszámok lehetnek, ezért került a táblázatba az $[a_3; b_3] = [0; 2]$ intervallum.

Hasonlóképpen, az X_4 változó esetén Csebisev egyenlőtlensége a $c_4 = 0,7$ választással azt adja, hogy

$$a_4 := -0,6736 < X_4 < 0,7264 =: b_4.$$

Így ebben az esetben az $[a_4; b_4]$ intervallum a $[0; 0]$ intervallumot, azaz az egyetlen 0 számot adja.

Köszönetnyilvánítás. Ezúttal is szeretném köszönetemet kifejezni Krámlí András professzor társamnak cikkem írása során folytatott tartalmas beszélgetésekért és a kéziratral kapcsolatos hasznos észrevételeiért.

Irodalomjegyzék

- [1] Prékopa András, Valószínűségelmélet műszaki alkalmazásokkal, Műszaki Könyvkiadó, Budapest, 1962; második kiadás: 1972.
- [2] Rényi Alfréd, Valószínűségszámítás, Tankönyvkiadó, Budapest, 1954; második kiadás: 1968.

