Valószínűségszámítás
2. rész
Nevezetes diszkrét eloszlások
—
FOGALMAK ÉS KIDOLGOZOTT PÉLDÁK

Vetier András

2019. december 3.

Tartalomjegyzék

Tisztelt Hallgatók!

Kérem, hogy a könyvben talált hibákat a **vetier@math.bme.hu** email címen jelezzék nekem. A levél tárgya legyen: **Hibát találtam** . Egy hatékony módszer a hibák rögzítésére, ha valaki számítógépen olvassa a könyvet:

- nyom egy PRINT SCREEN-t
- behívja a PAINT programot
- nyom egy PASTE (ctrl-V) utasítást
- pirossal bekarikázza a hibát, esetleg valamit oda ír
- elmenti a JPG fájlt, a fájl neve legyen a hiba helyének az oldalszáma, vagy a nevében legyen benne az oldalszám
- a JPG fájlokat csatolt fájlként elküldi a fent megadott címre

Természetesen minden más módszerrel küldött hibajelzést is köszönök.

2019. február 5.

Vetier András

1. Nevezetes eloszlások

1.1. Egyenletes eloszlások

1.1.1. Egyenletes eloszlás egydimenzióban

Mikor használjuk: Ha adott n darab szám, és ezek mindegyike ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x) \;=\; \frac{1}{n} \qquad \text{minden lehetséges } x \text{ -re}$$

Példa: Ilyen valószínűségi változóhoz jutunk, amikor szabályos dobókockával dobunk, és

$$X \;=\; \text{a kockán felül lévő szám}$$

$$p(x) \;=\; \frac{1}{6} \qquad (\ x = 1, 2, 3, 4, 5, 6\ )$$

1.1.2. Egyenletes eloszlás kétdimenzióban

Mikor használjuk: Ha adott n darab pont a síkon (emlékeztetőül: egy pont a síkon egy számpárt jelent), és ezek a pontok mindegyike ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x,y) \;=\; \frac{1}{n} \quad ; \qquad \text{minden lehetséges } (x,y) \text{ pontra}$$

Példa: Ilyen valószínűségi változóhoz jutunk, amikor egy piros és egy zöld dobókockával dobunk, és

$$X \;=\; \text{a piros kockán felül lévő szám}$$

$$Y \;=\; \text{a zöld kockán felül lévő szám}$$

$$p(x,y) \;=\; \frac{1}{36} \qquad (\ x = 1, 2, 3, 4, 5, 6; \quad y = 1, 2, 3, 4, 5, 6\ )$$

1.1.3. Egyenletes eloszlás r -dimenzióban

Mikor használjuk: Ha adott n darab pont az r -dimenziós térben (emlékeztetőül: egy r -dimenziós pont egy r elemű sorozatot jelent), és mindegyik pont ugyanakkora eséllyel bukkan fel.

Súlyfüggvény:

$$p(x_1, x_2, \dots x_r) \;=\; \frac{1}{n} \qquad \text{minden lehetséges } (x_1, x_2, \dots x_r) \text{ pontra}$$

Példa: Ilyen 3 -dimenziós valószínűségi változóhoz jutunk, amikor egy piros, egy zöld és egy kék dobókockával dobunk, és

$$X \;=\; \text{a piros kockán felül lévő szám}$$

$$Y \;=\; \text{a zöld kockán felül lévő szám}$$

$$Z \;=\; \text{a kék kockán felül lévő szám}$$

$$p(x, y, z) \;=\; \frac{1}{216} \qquad (\; x = 1, 2, 3, 4, 5, 6; \quad y = 1, 2, 3, 4, 5, 6; \quad z = 1, 2, 3, 4, 5, 6 \;)$$

1.2. Hipergeometrikus eloszlások

1.2.1. Hipergeometrikus eloszlás

Súlyfüggvény:

$$p(x) \;=\; \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \quad \text{ha} \quad x \geq 0, \quad x \geq n - N + K, \quad x \leq n, \quad x \leq K$$

Mikor használjuk: Ha N darab golyó van egy ládában, közülük K darab piros, a többi $N - K$ darab fehér, és visszatevés **nélkül** húzunk n -szer, akkor az

$$X \;=\; \text{ahányszor pirosat húzunk}$$

valószínűségi változó ilyen eloszlást követ.

Paraméterek jelentése:

$$n \;=\; \text{a húzások száma}$$

$$K \;=\; \text{a piros golyók száma a dobozban a húzások megkezdése előtt}$$

$$N \;=\; \text{az összes golyók száma a dobozban a húzások megkezdése előtt}$$

Ha valamiért fontos, akkor az eloszlás neve előtt a paraméterek jelét vagy numerikus értékeiket az itt megadott sorrendben adjuk meg.

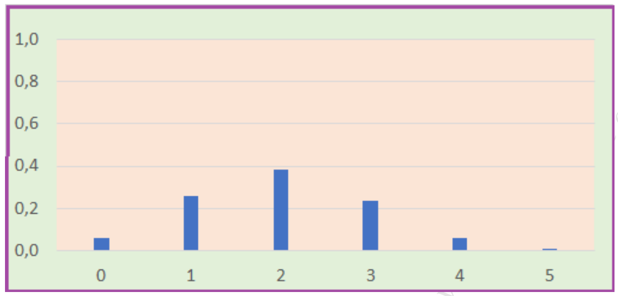
A súlyfüggvény képletének levezetése: Az N golyó közül n darabot kihúzni $\binom{N}{n}$ -féleképpen lehet, ennyi az összes kombinációk száma. Az $X = x$ esemény azt jelenti, hogy a kihúzott n golyó között x darab piros és $n - x$

darab fehér van. Ezért azt kell meghatároznunk, hogy hány olyan kombináció van, amiben x darab piros és $n-x$ darab fehér golyó van. Az x darab piros golyó a K darab piros közül $\binom{K}{x}$ féle képen kerülhet ki. Az $n-x$ fehér golyó az $N-K$ darab fehér közül $\binom{N-K}{n-x}$ -féleképpen kerülhet ki. Bármely piros kombinációt bármelyik fehér kombinációval össze lehet párosítani. Ezért az x darab piros és $n-x$ darab fehér eseményre nézve kedvező kombinációk száma a két szám szorzata:

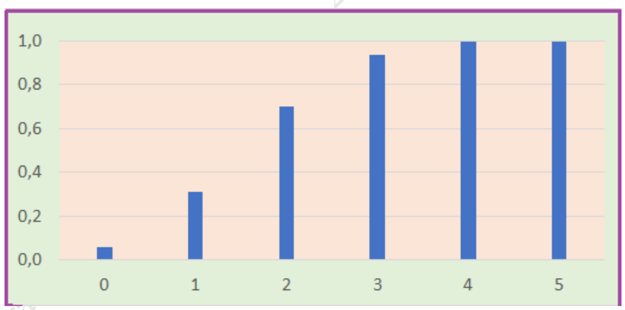
$$\binom{K}{x}\binom{N-K}{n-x}$$

Ezért a keresett súlyfüggvény érték:

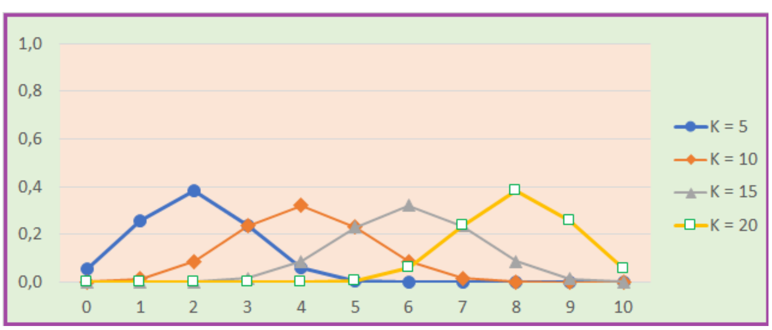
$$p(x) \;=\; \mathrm{P}(\,X = x\,) \;=\; \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$



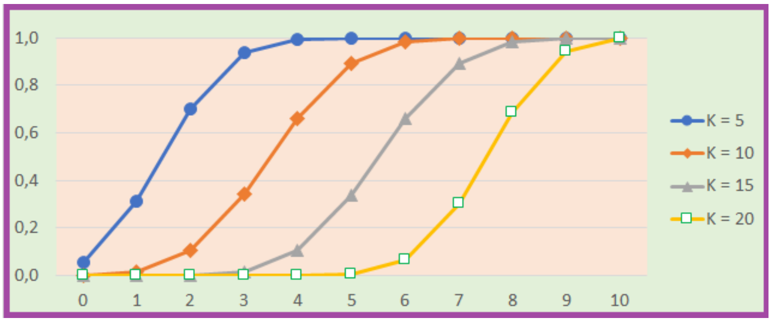
1. ábra. *Hipergeometrikus eloszlás súlyfüggvénye;* $n=10;\ K=5\ ;\ N=25$



2. ábra. *Hipergeometrikus eloszlás eloszlásfüggvénye;* $n=10;\ K=5\ ;\ N=25$



3. ábra. *Hipergeometrikus eloszlások súlyfüggvényei;* $n=10;\ K=5;\ 10;\ 15\ 20;\ N=25$



4. ábra. *Hipergeometrikus eloszlások eloszlásfüggvényei;* $n = 10;\ K = 5;\ 10;\ ;\ 20;\ N = 25$

Példa: Ha az ötös lottón egy szelvénnyel játszom, és X jelöli a találataim számát, akkor a 2, 3, 4, illetve 5 találat valószínűsége:

$$\mathrm{P}(\ X = 2\ ) = \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} \qquad \mathrm{P}(\ X = 3\ ) = \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \qquad \mathrm{P}(\ X = 4\ ) = \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}} \qquad \mathrm{P}(\ X = 5\ ) = \frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}}$$

azaz

$$p(x) = \frac{\binom{5}{x}\binom{85}{5-x}}{\binom{90}{5}} \qquad \text{ha} \quad 0 \le x \le 5$$

Excel-függvények:

$p(x)$ = `HYPGEOM.DIST(x;n;K;N;FALSE)` = `HIPERGEOM.ELOSZLÁS(x;n;K;N;HAMIS)`

$F(x)$ = `HYPGEOM.DIST(x;n;K;N;TRUE)` = `HIPERGEOM.ELOSZLÁS(x;n;K;N;IGAZ)`

1. Példa: Hány szarvas él az erdőben? Egy erdőben ismeretlen számú szarvas él. A számuk becslése céljából 60 szarvast piros festékkel megjelölünk, majd néhány hét eltelte után megszámoljuk, hogy 40 véletlenszerűen választott szarvas között hány megjelöltre bukkanunk. Tegyük fel, hogy 15 -re. Mit mondhatunk ezek alapján a szarvasok ismeretlen N számáról?

Megoldás: Átfogalmazzuk a feladatot erdőben szabadon élő szarvasok helyett dobozba zárt golyókra. Igaz, így kevésbé izgalmas a probléma, de könnyebben elképzelhető.

1. Példa módosítása: Hány golyó van a dobozban? Egy dobozban ismeretlen számú fehér golyó van, melyek közül 60 -at pirosra festünk, majd jól összekeverjük a golyókat, és 40 -szer húzunk visszatevés nélkül. Tegyük fel, hogy a kihúzott golyók között 15 piros akad. Mit mondhatunk ezek alapján a golyók ismeretlen N számáról?

Gyors megoldás: Az összes golyók között a pirosak aránya $60 : N$. A kihúzott golyók között a pirosak aránya $15 : 40$. Feltételezve, hogy ez a két arány körülbelül egyenlő, a

$$60 : N \approx 15 : 40$$

közelítő egyenletet kapjuk, amiből az jön ki, hogy

$$N \approx 60 * \frac{40}{15} = 160$$

Alaposabb megoldás: Ha X -szel jelöljük azt a valószínűségi változót, hogy hányszor húzunk piros golyót, akkor X hipergeometrikus eloszlást követ ismeretlen N , $K = 60$ és $n = 40$ paraméterekkel. Az ismeretlen N paraméter néhány értéke mellett – a hipergeometrikus eloszlás képlete alapján – kiszámoltuk a $\mathrm{P}(\ 15\ \text{piros}\ )$ valószínűséget, és az értékeket táblázatba foglaltuk:

N	P(15 piros)
100	0.00
120	0.02
140	0.11
160	0.15
180	0.12
200	0.08
220	0.04
240	0.02
260	0.01
280	0.01
300	0.00

Táblázat: *A P(15 piros) valószínűség N függvényében*

Vegyük észre, hogy $N = 160$ esetén a valószínűség értéke 0.15 , de $N \leq 100$ vagy $N \geq 300$ esetén – két tizedesre kerekítve – 0.00 , ami igazából azt jelenti, hogy a a valószínűség értéke 0.005 -nél is kisebb. Ezért úgy gondolhatjuk, hogy a golyók (szarvasok) száma 160 körül van. A táblázatban szereplő valószínűségekből arra következtetünk, hogy a golyók (szarvasok) száma 100 és 300 között van.

Az, hogy a golyók (szarvasok) száma 100 és 300 között van, a mi tippünk. Egyáltalán nem biztos, hogy ez így is van. Lehet, hogy a golyók (szarvasok) száma 100 -nál kevesebb vagy 300 -nál több, és ezért jól kinevetnek minket a szarvasok, és azt mondják, hogy "Bibibí, rosszul tippeltetek, mi többen vagy kevesebben vagyunk annál, mint ahogy azt ti tippeltétek". De megnyugtató érzést ad nekünk, hogy annak az esélye, hogy kinevetnek, kicsúfolnak minket a szarvasok, garantáltan kisebb, mint 0.005 .

Egy másik, kevésbé szégyenlős ember nagyobb rizikót is felvállalhat. Abból a tényből, hogy 15 piros golyó van a 40 kihúzott között, szűkebb határt is mondhat a golyók (szarvasok) számára. Például azt, hogy 120 és 240 között van a golyók (szarvasok) száma. Az, hogy a golyók (szarvasok) száma 120 és 120 között van, ez az ő tippje. Egyáltalán nem biztos, hogy igaza van. Lehet, hogy a golyók (szarvasok) száma 120 -nál kevesebb vagy 240 -nél több, és ezért jól kinevetik őt a szarvasok, és azt mondják, hogy "Bibibí, rosszul tippeltél, mi többen vagy kevesebben vagyunk annál, mint ahogy azt te tippelted". Ezt az embert megnyugtatja az a tény, hogy annak az esélye, hogy kinevetik és kicsúfolják a szarvasok, garantáltan kisebb, mint 0.025 . Annak a valószínűsége, hogy ezt a kevésbé szégyenlős embert kinevetik a szarvasok, nagyobb, mint annak a valószínűsége, hogy minket kinevetnek. Ez az ára annak, hogy ő szűkebb határokkal tippel, mint mi.

Tehát azt látjuk, hogy ilyen vagy olyan alsó és felső határokkal tippelve a golyók (szarvasok) számára, annak az esélye, hogy kinevetnek és kicsúfolnak minket a szarvasok, legfeljebb ennyi vagy annyi:

- a 100 – 300 -as tipp esetén legfeljebb 0.005
- a 120 – 240 -es tipp esetén legfeljebb 0.025

Azt a nyilvánvaló tényt, hogy szűkebb határok esetén a szarvasoknak nagyobb esélyük lehet a nevetésre, csúfolásra, mint tágabb határok esetén, a valószínűségszámítás elmélete alapján számszerű összefüggéssel tudtuk finomítani. Ezt a számszerű összefüggést a fentebb megadott táblázatból olvastuk ki (a táblázat által megengededett pontosság erejéig). A valószínűségszámítás elmélete – többek között - ilyesmire tanít meg minket.

2. Példa: Hányan szeretnek bridzsezni a városban? Városunban szeretnék egy felnőtt bridzs klubot szervezni. Gondoltam, felmérem, hogy a városban élő kb. 10 ezer felnőtt lakos közül kb. hányan fognak a dolog iránt érdeklődni. Ezért 100 véletlenszerűen választott felnőttnek feltettem a kérdést: szeretnek-e bridzsezni? A megkérdezettek közül 42 -en feleltek igennel, a többiek nemet mondtak. Mit mondhatunk ezek alapján a bridzset kedvelő felnőtttek ismeretlen K számáról?

Gyors megoldás: Az összes felnőtt között a bridzset kedvelő felnőttek aránya $K : 10\,000$. A megkérdezett felnőttek között a bridzset kedvelő felnőttek aránya $42 : 100$. Feltételezve, hogy ez a két arány körülbelül egyenlő, a

$$K : 10\,000 \;\approx\; 42 : 100$$

közelítő egyenletet kapjuk, amiből az jön ki, hogy

$$K \;\approx\; 10\,000 * \frac{42}{100} \;=\; 4\,200$$

Alaposabb megoldás: Ha X -szel jelöljük azt a valószínűségi változót, hogy a megkérdezett felnőttek közül hányan felelnek igennel, akkor X hipergeometrikus eloszlást követ $N = 10\,000$, ismeretlen K és $n = 100$ paraméterekkel. Az ismeretlen K paraméter néhány értéke mellett – a hipergeometrikus eloszlás képlete alapján – kiszámoltuk a P(42 igen) valószínűséget, és az értékeket táblázatba foglaltuk:

K	P(42 igen)
3 000	0.00
3 400	0.02
3 800	0.06
4 200	0.08
4 600	0.06
5 000	0.02
5 400	0.00

Táblázat: *A P(42 igen) valószínűség K függvényében*

Vegyük észre, hogy $K = 4200$ esetén a valószínűség értéke 0.08 , de $K \leq 3\,000$ vagy $K \geq 5\,400$ esetén – két tizedesre kerekítve – 0.00 , ami igazából azt jelenti, hogy a valószínűség értéke 0.005 -nél is kisebb. Ezért úgy gondolhatjuk, hogy a bridzset kedvelő felnőttek száma 4200 körül van. A táblázatban szereplő valószínűségekből arra következtetünk, hogy a bridzset kedvelő felnőttek száma $3\,000$ és $5\,400$ között van.

Megjegyzés: Ha a felnőtt emberek közül történő véletlen választás módszere nem garantálja, hogy minden embert legfeljebb egyszer kérdezünk meg, tehát ismétlődés is előfordulhat, akkor a P(42 igen) valószínűséget a binomiális eloszlás képlete alapján kell kiszámolni. Könnyen ellenőrizhető, hogy két tizedesre kerekítve ilyenkor is ugyanazokat a valószínűségeket kapjuk, mint amiket a hipergeometrikus eloszlással kaptunk a fentebbi táblázatban. A binomiális eloszlás és a hipergeometrikus eloszlás közelségén nem szabad meglepődni, mert ha N és K n -hez képest nagy, akkor alig számít (szinte nem is számít), hogy a választás során megengedjük-e az ismétlődést vagy nem.

az N , K , n paraméterű hipergeometrikus eloszlást jól közelíti az n , p paraméterű binomiális eloszlás, ahol $p = K/N$.

1.2.2. Polihipergeometrikus eloszlás

Súlyfüggvény:

$$p(x,y) \;=\; \frac{\binom{K_1}{x}\binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y}}{\binom{N}{n}}$$

ha

$$0 \leq x \leq \min(n, K_1),$$

$$0 \leq y \leq \min(n, K_2),$$
$$0 \leq n - x - y \leq \min(n, N - K_1 - K_2) \leq n$$

Mikor használjuk: Ha egy dobozban N golyó van, melyek közül K_1 darab piros, K_2 darab zöld, $N - K_1 - K_2$ darab fehér, és n -szer húzunk a dobozból **visszatevés nélkül**, akkor az a kétdimenziós (X, Y) valószínűségi változó, melynek koordinátái

$$X = \text{ahányszor piros golyót húzunk}$$
$$Y = \text{ahányszor zöld golyót húzunk}$$

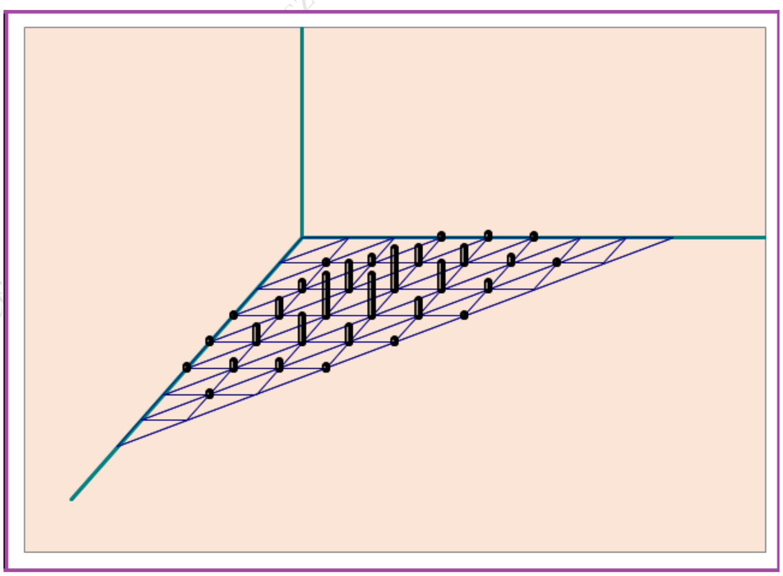
ilyen eloszlást követ.

A súlyfüggvény képletének levezetése: Az N golyó közül n darabot kihúzni $\binom{N}{n}$ -féleképpen lehet, ennyi az összes kombinációk száma. Az $X = x, Y = y$ esemény azt jelenti, hogy a kihúzott n golyó között x darab piros, x darab zöld és $n - x - y$ darab fehér van. Ezért azt kell meghatároznunk, hogy hány olyan kombináció van, amiben x darab piros, y darab zöld és $n - x - y$ darab fehér golyó van. Az x darab piros golyó az K_1 darab piros közül $\binom{K_1}{x}$ -féleképpen kerülhet ki. Az y darab zöld golyó az K_2 darab zöld közül $\binom{K_2}{y}$ -féleképpen kerülhet ki. Az $n - x - y$ fehér golyó az $N - K_1 - K_2$ darab fehér közül $\binom{N-K_1-K_2}{n-x-y}$ -féleképpen kerülhet ki. Bármely piros kombinációt bármelyik zöld és bármelyik fehér kombinációval össze lehet rakni. Ezért az x darab piros, y darab zöld és $n - x$ darab fehér eseményre nézve kedvező kombinációk száma e három szám szorzata:

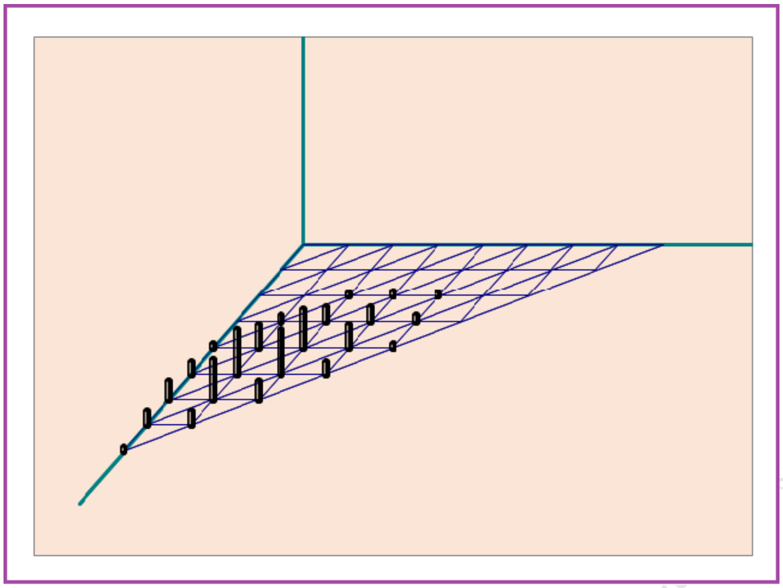
$$\binom{K_1}{x}\binom{K_2}{y}\binom{N - K_1 - K_2}{n - x - y}$$

Ezért a keresett súlyfüggvény érték:

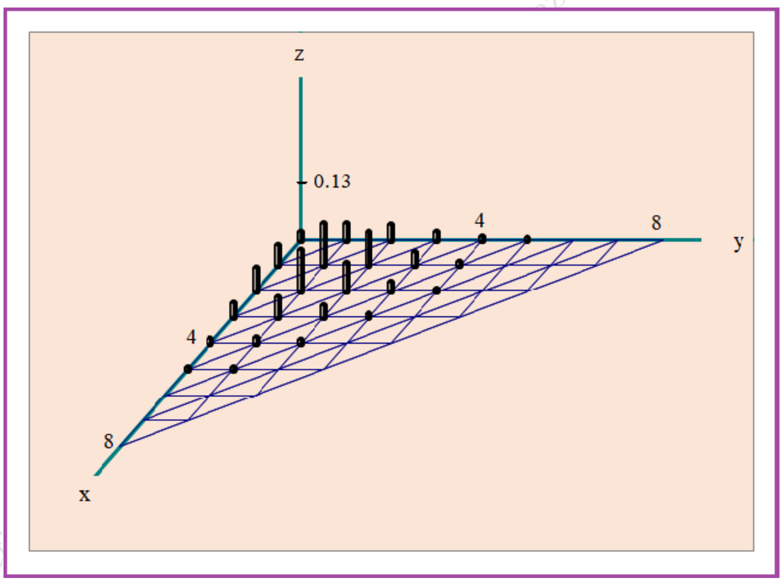
$$p(x, y) = \mathrm{P}(\, X = x, Y = y \,) = \frac{\binom{K_1}{x}\binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y}}{\binom{N}{n}}$$



5. ábra. *Polihipergeometrikus eloszlás súlyfüggvénye;* $n = 8;\ K_1 = 10,\ K_2 = 10;\ N = 30$



6. ábra. *Polihipergeometrikus eloszlás súlyfüggvénye;* $n = 8;\ K_1 = 30,\ K_2 = 10;\ N = 50$



7. ábra. *Polihipergeometrikus eloszlás súlyfüggvénye;* $n = 8;\ K_1 = 30,\ K_2 = 10;\ N = 50$

Megjegyzés: Nyilvánvaló, hogy ha a dobozból való húzásokat visszatevéssel végezzük, akkor más eloszlást kapunk. Néhány oldallal később ki fog derülni, hogy visszatevéses húzás esetén az (X, Y) valószínűségi változó polinomiális eloszlást követ $p_1 = \frac{K_1}{N}$, $p_2 = \frac{K_2}{N}$ paraméterekkel.

Vetület eloszlások:

Teljesen nyilvánvaló, hogy az

$$X = \text{ahányszor piros golyót húzunk}$$

valószínűségi változó hipergeometrikus eloszlást követ n , K_1 , N paraméterekkel, vagyis egy polihipergeometrikus eloszlásnak a vízszines tengelyre vetett vetülete (és természtetsen a függőleges tengelyre vetett vetülete is) hipergeometrikus eloszlás. Ez a tény a

$$p_1(x) \;=\; \sum_y \; p(x,y)$$

összegzési szabállyal is kijön, hiszen

$$\sum_y \frac{\binom{K_1}{x}\binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y}}{\binom{N}{n}} \;=\; \frac{\binom{K_1}{x}}{\binom{N}{n}} \cdot \sum_y \binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y} \;=$$

$$= \; \frac{\binom{K_1}{x}}{\binom{N}{n}} \cdot \binom{N-K_1}{n-x} \;=\; \frac{\binom{K_1}{x}\binom{N-K_1}{n-x}}{\binom{N}{n}}$$

A bizonyítás második lépésében felhasználtuk a

$$\sum_y \binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y} \;=\; \binom{N-K_1}{n-x}$$

azonosságot.

Feltételes eloszlások:

- Y feltételes eloszlása az $X = x$ feltétel mellett olyan hipergeometrikus eloszlás, mintha $(n-x)$ -szer húznánk visszatevés nélkül egy olyan dobozból, ami $N - K_1$ golyót tartalmaz, melyek közül K_2 zöld és a többi fehér színű, és mi azt néznénk, hogy hányszor húzunk zöldet. Az állítást heurisztikus gondolatmenet is kiadja, és – természetesen – osztással, egyszerűsítéssel is kijön:

$$p_{2|1}(\,y\,|\,x\,) \;=\; \mathsf{P}(\,Y = y \,|\, X = x\,) \;=\; \frac{\mathsf{P}(\,X = x\,,\, Y = y\,)}{\mathsf{P}(\,X = x\,)} \;=$$

$$= \; \frac{\dfrac{\binom{K_1}{x}\binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y}}{\binom{N}{n}}}{\dfrac{\binom{K_1}{x}\binom{N-K_1}{n-x}}{\binom{N}{n}}} \;=\; \frac{\binom{K_2}{y}\binom{N-K_1-K_2}{n-x-y}}{\binom{N-K_1}{n-x}}$$

  Hasonlóan kiadódik az is, hogy

- X feltételes eloszlása az $Y = y$ feltétel mellett olyan hipergeometrikus eloszlás, mintha $(n-y)$ -szor húznánk visszatevés nélkül egy olyan dobozból, ami $N - K_2$ golyót tartalmaz, melyek közül K_1 piros és a többi fehér színű, és mi azt néznénk, hogy hányszor húzunk pirosat.

A kétdimenziós súlyfüggvény előállítása Excelben:

A kétdimenziós súlyfüggvényt – mint tudjuk – a vetület eloszlás és a feltételes eloszlás súlyfüggvényeiből szorzatként állíthatjuk elő a

$$p(x,y) \;=\; p_1(\,x\,) \,\cdot\, p_{2|1}(\,y\,|\,x\,)$$
$$p(x,y) \;=\; p_2(\,y\,) \,\cdot\, p_{1|2}(\,x\,|\,y\,)$$

szorzási szabályok alapján. Ezért – fentebbiek figyelembe vételével – a kétdimenziós polinomiális eloszlás $p(x,y)$ súlyfüggvényét Excellel az alábbi képletek segítségével tudjuk előállítani:

$p(x,y) \;=\;$ `HYPGEOM.DIST(` x ; n ; K_1 ; N ; `FALSE` `)` $\cdot$ `HYPGEOM.DIST(` y ; $n{-}x$; K_2 ; $N{-}K_1$; `FALSE` `)`

$p(x,y) \;=\;$ `HYPGEOM.DIST(` y ; n ; K_2 ; N ; `FALSE` `)` $\cdot$ `HYPGEOM.DIST(` x ; $n{-}y$; K_1 ; $N{-}K_2$; `FALSE` `)`

1.2.3. Polihipergeometrikus eloszlás, r -dimenziós *(Extra tananyag)*

Súlyfüggvény:

$$p(x_1, x_2, \ldots, x_r) \;=\; \frac{\binom{K_1}{x_1}\binom{K_2}{x_2}\ldots\binom{K_r}{x_r}\binom{N-K_1-K_2-\ldots-K_r}{n-x_1-x_2-\ldots-x_r}}{\binom{N}{n}}$$

$$\text{ha} \qquad 0 \le x_i \;\le\; \min(n, K_i) \qquad (1 \le i \le r) \qquad \text{és}$$

$$0 \le n - x_1 - x_2 - \ldots - x_r \;\le\; \min(n, N - K_1 - K_2 - \ldots - K_r)$$

Mikor használjuk: Ha egy dobozban N golyó van, melyek közül K_1 darab piros, K_2 darab zöld, és így tovább K_r darab lila golyó van, továbbá $N - K_1 - K_2 - \ldots K_r$ darab fehér, és n -szer húzunk a dobozból **visszatevés nélkül**, akkor az az n -dimenziós valószínűségi változó, melynek koordinátái

$$X_1 \;=\; \text{ahányszor piros golyót húzunk}$$

$$X_2 \;=\; \text{ahányszor zöld golyót húzunk}$$

$$\vdots$$

$$X_r \;=\; \text{ahányszor lila golyót húzunk}$$

ilyen eloszlást követ.

Megjegyzés: Nyilvánvaló, hogy ha a dobozból való húzásokat visszatevéssel végezzük, akkor az $(X_1, X_2, \ldots, X_r)$ valószínűségi változó polinomiális eloszlást követ

$$p_1 \;=\; \frac{K_1}{N} \qquad p_2 \;=\; \frac{K_2}{N} \qquad \ldots \qquad p_r \;=\; \frac{K_r}{N}$$

paraméterekkel.

1.3. Binomiális eloszlás és társai

1.3.1. Binomiális eloszlás

Súlyfüggvény:

$$p(x) \;=\; \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{ha} \quad x = 0, 1, 2, \ldots, n$$

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Egy p valószínűségű eseményre n kísérletet végzünk. Az alábbi valószínűségi változó követ ilyen eloszlást:

$$X \;=\; \text{ahányszor bekövetkezik az esemény az } n \text{ kísérlet során}$$

Paraméterek jelentése:

$$n \;=\; \text{a kísérletek száma}$$

$$p \;=\; \text{az esemény valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** n darab független, külön-külön p valószínűségű esemény kapcsán az alábbi valószínűségi változó követ ilyen eloszlást:

$$X = \text{ahány esemény bekövetkezik az } n \text{ esemény közül}$$

Paraméterek jelentése:

$$n = \text{ az események száma}$$

$$p = \text{ az események közös valószínűség értéke}$$

3. **Golyók húzása dobozból:** Ha N darab golyó van egy ládában, közülük K darab piros, a többi $N - K$ darab fehér, és n -szer húzunk **visszatevéssel**, akkor a valószínűségi változó ilyen eloszlást követ:

$$X = \text{ahányszor pirosat húzunk}$$

Paraméterek jelentése:

$$n = \text{ a húzások száma}$$

$$p = K/N = \text{ piros húzásának valószínűsége minden egyes húzásnál}$$

A súlyfüggvény képletének levezetése: A levezetést egy példán keresztül mutajuk be. E célból tegyük fel, hogy egy 10 tagú társaságból mindenki, a többiektől függetlenül, p valószínűséggel megy el vasárnap kirándulni. Jelölje X a kirándulók számát. Most azt kérdezzük, hogy mennyi az X súlyfügvényének az értéke az $x = 3$ helyen, azaz mennyi annak az eseménynek a valószínűsége, hogy $X = 3$, vagyis hogy pontosan 3 ember megy el kirándulni:

$$p(3) = \mathrm{P}(\ X = 3\ ) = \ ?$$

Válasz: Az, hogy 3 ember megy el kirándulni, bekövetkezhet például úgy, hogy a társaság 3 legöregebb tagja elmegy, a többi 7 pedig nem. Ennek a valószínűsége, a független eseményekre vonatkozó szorzási szabály miatt nyilván:

$$\begin{aligned} p\,p\,p\ (1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p)\,(1-p) &= \\ &= p^3\,(1-p)^7 \end{aligned}$$

Az, hogy 3 ember megy el kirándulni, természetesen másképpen is lehetséges: a társaság akármelyik 3 tagja elmehet, a többi 7 pedig nem. Akármelyik 3 tag esetében annak a valószínűsége, hogy ők elmennek, a többi 7 pedig nem, ugyancsak

$$p^3\,(1-p)^7$$

Ezért, ha nem fontos számunkra, hogy ki az a 3 ember, aki elmegy, csak az fontos számunkra, hogy pontosan 3 ember menjen el (márpedig az $X = 3$ esemény esetében ez a helyzet), akkor ennek az eseménynek a valószínűségét úgy kapjuk meg, hogy a $p^3\,(1-p)^7$ közös valószínűség értéket megszorozzuk a 3 elemű kombinációk számával, vagyis $\binom{10}{3}$ -mal:

$$p(3) = \mathrm{P}(\ X = 3\ ) = \binom{10}{3}\,p^3\,(1-p)^7$$

Világos, hogy ha a társaság nem 10 tagú, hanem n tagú, akkor annak a valószínűsége, hogy az n ember közül pontosan x ember megy el kirándulni:

$$p(x) = \mathrm{P}(\ X = x\ ) = \binom{n}{x} p^x (1-p)^{n-x}$$

Példa: Ha dobókockával 20 -szor dobok, és a dobott hatosok számát X jelöli, akkor a súlyfüggvény:

$$p(x) = \binom{20}{x} p^x (1-p)^{20-x} \qquad (x = 0, 1, \ldots, 20)$$

A pontosan 3 hatos valószínűsége:

$$\mathrm{P}(\ X = 3\ ) \ = \ \binom{20}{3}(1/6)^3(5/6)^{17}$$

Megjegyzés: A binomiális eloszlást a **visszatevéses**, a hipergeometrikus eloszlást a **visszatevés nélküli** húzások esetén kell használni.

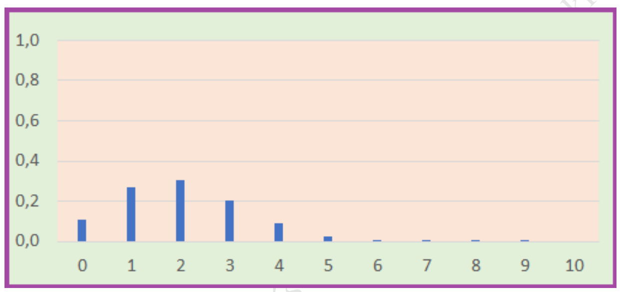
Excel-függvények:

$$p(x) \ = \ \texttt{BINOM.DIST(x;n;p;FALSE)} \ = \ \texttt{BINOM.ELOSZLÁS(x;n;p;HAMIS)}$$
$$F(x) \ = \ \texttt{BINOM.DIST(x;n;p;TRUE)} \ = \ \texttt{BINOM.ELOSZLÁS(x;n;p;IGAZ)}$$

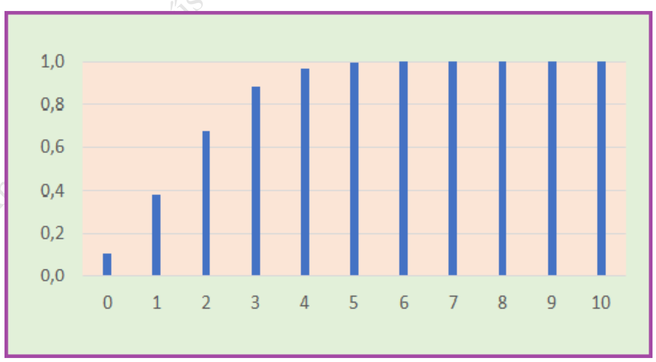
Megjegyzés: Egyes Excel verziókban az eloszlások nevében a . (pont) elhagyható, – például –

`BINOM.DIST` helyett `BINOMDIST`

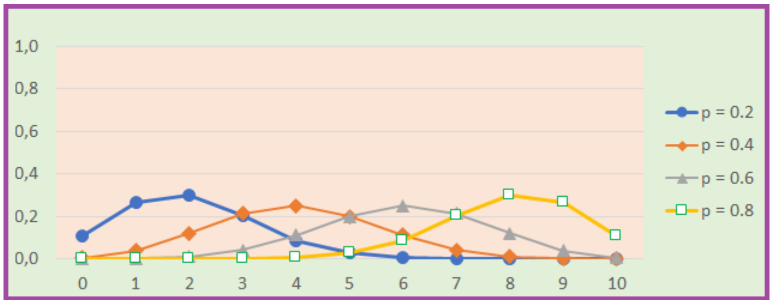
írható.



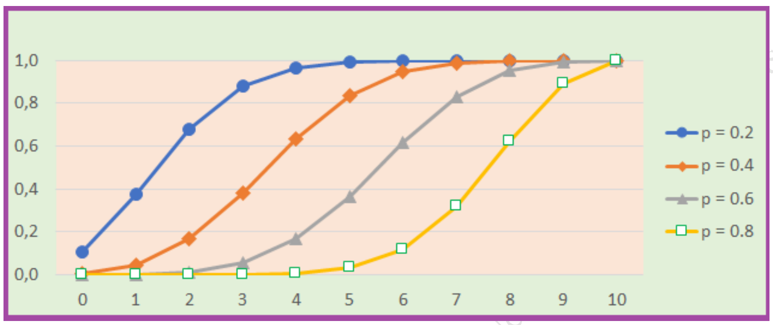
8. ábra. *Binomiális eloszlás súlyfüggvénye;* $n = 10;\ p = 0.2$



9. ábra. *Binomiális eloszlás eloszlásfüggvénye;* $n = 10;\ p = 0.2$



10. ábra. *Binomiális eloszlások súlyfüggvényei;* $n = 10;\ p = 0.2,\ 0.4,\ 0.6,\ 0.8$



11. ábra. *Binomiális eloszlások eloszlásfüggvényei;* $n = 10;\ p = 0.2,\ 0.4,\ 0.6,\ 0.8$

Megjegyzés: Hipergeometrikus eloszlás $\approx$ binomiális eloszlás. Kis ügyeskedéssel be lehet látni, hogy akármilyen x pozitív egész szám esetén az

$$K \to \infty\,, \quad N \to \infty\,, \quad K/N \to p$$

feltételek mellett

$$\lim \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

Ez azt jelenti, hogy ha x -hez képest K is és N is nagy, és a K/N arány közelítőleg p , akkor a hipergeometrikus eloszlás x -ik tagja közelítőleg egyenlő a binomiális eloszlás x -ik tagjával:

$$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x}$$

Az egzakt levezetést az "ügyeskedésekkel" az Olvasóra bízzuk. Cserébe egy heurisztikus magyarázatot adunk színes golyók segítségével:

Heurisztikus magyarázat: Tegyünk N golyót egy dobozba úgy, hogy közülül K piros, a többi fehér. Húzzunk a dobozból n -szer visszatevés nélkül. Annak az esélye, hogy pontosan x -szer húzunk pirosat – mint tudjuk – a hipergeometrikus eloszlás képlete szerint

$$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Most tegyünk N golyót egy másik dobozba ugyanúgy, mint az előbb, de most visszatevéssel húzzunk a dobozból n -szer. Annak a esélye, hogy pontosan x -szer húzunk pirosat – mint tudjuk – a binomiális eloszlás képlete szerint

$$\binom{n}{x} \left(\frac{K}{N}\right)^x \left(1 - \frac{K}{N}\right)^{n-x}$$

Ha x -hez képest K is és N is nagy, akkor a kihúzott piros golyók számára nézve nincs sok hatása annak, hogy visszatevés nélkül vagy visszatevéssel húzúnk, ezért az x darab piros golyó esélye így is úgy is kb ugyanannyi:

$$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\left(\frac{K}{N}\right)^{x}\left(1-\frac{K}{N}\right)^{n-x}$$

Mivel feltettük, hogy az K/N arány közelítőleg p , a jobboldali kifejezést kicserélhetjük így:

$$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}p^{x}(1-p)^{n-x}$$

Ezek után talán kedve támad az Olvasónak, hogy az egzakt levezetést "kiügyeskedje". Hajrá!

1. Feladat: Vajon mindenki le tud ülni? 400 hallgató mindegyike egymástól függetlenül 0.6 valószínűséggel jár órára. A teremben 250 db szék van. Mi a valószínűsége annak, hogy aki elmegy órára, mind le tud ülni egy-egy székre?

Megoldás: Először vegyük észre, hogy az órát látogató diákok X száma valószínűségi változó, mely – a diákok habitusa miatt – binomiális eloszlást követ $n = 400$ és $p = 0.6$ paraméterekkel. Az az esemény, hogy mindenkinek, aki elmegy órára, jut szék, azt jelenti, hogy $X \leq 250$. Ennek az eseménynek a P($X \leq 250$) valószínűségét az eloszlásfüggvénynek a 250 helyen felvett értéke adja meg, ami Excellel így számolható:

$$F(250) = \texttt{BINOM.DIST(250 ; 400 ; 0.6 ; TRUE)} =$$

$$= \texttt{BINOM.ELOSZLÁS(250 ; 400 ; 0.6 ; IGAZ)} = 0.86$$

Tehát 0.86 valószínűséggel csak 250 vagy kevesebb hallgató megy órára, ezért 250 széken remekül elférnek. 0.14 valószínűséggel 250 -nél több hallgató megy órára, ezért ilyenkor lesz, aki nem tud székre ülni.

2. Feladat: Biztos, hogy mindenki le tud ülni? Hány szék kell ahhoz, hogy biztosan (1 valószínűséggel, 100% biztonsággal) jusson szék mindenkinek, aki elmegy órára?

Megoldás: Ahhoz, hogy biztosan jusson szék mindenkinek, aki elmegy órára, nyilván 400 vagy több székre van szükség.

3. Feladat: Spóroljunk a székekkel! Hány szék kell ahhoz, hogy legalább 0.99 legyen a valószínűsége annak (99% legyen a biztonsága annak), hogy aki elmegy órára, mind le tud ülni egy-egy székre?

Megoldás: Ahhoz, hogy legalább 0.99 legyen a valószínűsége annak, hogy mindenkinek, aki elmegy órára, jusson szék, az kell, hogy a székek x száma eleget tegyen az alábbi egyenlőtlenségnek:

$$F(x) = \mathrm{P}(\ X \leq x\ ) \geq 0.99$$

Ezért az

$$F(x) = \texttt{BINOM.DIST(x ; 400 ; 0.6 ; TRUE)} =$$

$$= \texttt{BINOM.ELOSZLÁS(x ; 400 ; 0.6 ; IGAZ)}$$

függvényre készítettünk egy táblázatot:

x	$F(x)$
260	0.9824
261	0.9864
262	0.9897
263	0.9922
264	0.9942
265	0.9957
266	0.9968
267	0.9977
268	0.9984
269	0.9988
270	0.9992

Táblázat: *Numerikus értékek* $F(x)$*-re*

amiből kiolvashatjuk, hogy x értéke, vagyis a székek száma legalább 263, akkor $F(x) = \mathrm{P}(\ X \le x\ ) \ \ge\ 0.99$.

1. Megjegyzés: Vegyük észre, hogy ha a 100 % biztonság helyett megelégszünk 99 % biztonsággal, akkor a székeknek több mint $1/3$ -át megspórolhatjuk! Ha valaki a 99% biztonsággal még nem elégedett, akkor tegyen be a 263 székhez még 7 -et, és a biztonság – a táblázat utolsó sora szerint – megnő 99.9 %-ra.

2. Megjegyzés: Ilyen jellegű feladatok adódnak, amikor valamilyen rendszer, például egy telefonközpont, kapacitását kell megtervezni úgy, hogy a rendszer nagy valószínűséggel el tudja látni a véletlenszerűen felmerülő igényeket.

4. Feladat: A gubanc valószínűsége. Tegyük fel, hogy egy repülőgépen 200 ülés van az utasok számára, de – bízván abban, hogy néhány utas betegség, forgalmi dugó, stb. miatt lemarad a járatról – az extra profit reményében 203 jegyet adnak el. Feltéve, hogy minden utas a többitől függetlenül 0.05 valószínűséggel marad le a járatról, számoljuk ki annak a valószínűségét, hogy a beszálláskor "gubanc támad", mert 200 -nál több utas jelenik meg?

Megoldás: A járathoz időben odaérkező utasok száma nyilván binomiális eloszlást követ $n = 203$ és $p = 0.95$ paraméterekkel. Ezért annak a valószínűsége, hogy 200 -nál több utas jelenik meg a beszálláskor, vagyis előáll a "gubanc", egyenlő

$$1 - \texttt{BINOM.DIST(200 ; 203 ; 0.95 ; TRUE)} =$$

$$= 1 - \texttt{BINOM.ELOSZLÁS(200 ; 203 ; 0.95 ; IGAZ)} = 0.0021$$

5. Feladat: Hány extra repülőjegyet adjunk el? Az előző feladatban kapott valószínűség értéke elég kicsi ahhoz, hogy a repülőtársaság felvállalja a gubanc következményeit. Nézzük meg, hogyan függ a gubanc valószínűsége az eladott extra jegyek számától! Gondoljuk meg, hogy a p paraméter függvényében hány extra jegyet adjon el a társaság, ha nem szeretne túl gyakran gubancot!

Megoldás: Excelben egyszerűen megszerkeszthető egy olyan táblázat, ami a gubanc valószínűségét mutatja p és az extra jegyek számának függvényében:

		p									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
	1	0.1326	0.0172	0.0022	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.3992	0.0865	0.0154	0.0025	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.6685	0.2265	0.0555	0.0113	0.0021	0.0003	0.0001	0.0000	0.0000	0.0000
extra	4	0.8507	0.4159	0.1368	0.0354	0.0078	0.0015	0.0003	0.0000	0.0000	0.0000
jegyek	5	0.9437	0.6091	0.2613	0.0844	0.0224	0.0051	0.0010	0.0002	0.0000	0.0000
száma	6	0.9818	0.7675	0.4144	0.1647	0.0522	0.0140	0.0033	0.0007	0.0001	0.0000
	7	0.9948	0.8763	0.5721	0.2751	0.1036	0.0323	0.0086	0.0021	0.0004	0.0001
	8	0.9987	0.9407	0.7120	0.4056	0.1794	0.0647	0.0198	0.0053	0.0013	0.0003
	9	0.9997	0.9741	0.8211	0.5414	0.2781	0.1152	0.0400	0.0120	0.0032	0.0008
	10	0.9999	0.9897	0.8971	0.6675	0.3926	0.1856	0.0728	0.0244	0.0072	0.0019

Táblázat: *A gubanc valószínűsége* p *és az extra jegyek számának függvényében*

A táblázatból kiolvashatjuk, hogy ha a gubanc valószínűségét 0.005 alatt akarjuk tartani, akkor

$p = 0.01$	esetén	0
$p = 0.02$	esetén	0
$p = 0.03$	esetén	1
$p = 0.04$	esetén	2
$p = 0.05$	esetén	3
$p = 0.06$	esetén	4
$p = 0.07$	esetén	6
$p = 0.08$	esetén	7
$p = 0.09$	esetén	9
$p = 0.10$	esetén	10

Táblázat: *Eladható extra jegyek száma* p *függvényében*

darab extra jegy adható el.

Feladat: Amikor a tesztek különböző eredményeket adnak! Tegyük fel, hogy egy bizonyos ritka betegségben az embereknek csupán 1 ezreléke szenved. A betegséget egy olyan vizsgálattal lehet kimutatni, ami sajnos mindkét irányban tévedhet: beteg emberek esetén csak 0.8 a valószínűsége annak, hogy a vizsgálat jelzi a betegséget, egészséges emberek esetében pedig csak 0.9 a valószínűsége annak, hogy az egészséges embert egészségesnek jelzi. Barátomat nemrég – egymástól függetlenül – n intézetben is vizsgálták A vizsgálatok közül k esetben betegnek jelezték, $n-k$ pedig egészségesnek. Ilyen feltételek mellett mennyi a valószínűsége annak, hogy barátom beteg?

Megoldás: Feltételezve, hogy a vizsgálatok függetlenek egymástól, igazak az alábbiak:

$$\text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)} = \binom{n}{k} 0.8^k (1-0.8)^{n-k} = \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)}$$

$$\text{P(betegnek diagnosztizálják } k \text{ -szor | egészséges)} = \binom{n}{k} 0.1^k (1-0.1)^{n-k} = \text{BINOM.DIST(k ; n ; 0.1 ; FALSE)}$$

Ezeket felhasználva – barátom esélyei így festenek:

P(beteg | betegnek diagnosztizálják k -szor) $=$

$$= \frac{\text{P(beteg és betegnek diagnosztizálják } k \text{ -szor)}}{\text{P(betegnek diagnosztizálják } k \text{ -szor)}}$$

$$= \frac{\text{P(beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)}}{\text{P(beteg és betegnek diagnosztizálják } k \text{ -szor)} + \text{P(egészséges és betegnek diagnosztizálják } k \text{ -szor)}}$$

$$= \frac{\text{P(beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)}}{\text{P(beteg)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | beteg)} + \text{P(egészséges)} \cdot \text{P(betegnek diagnosztizálják } k \text{ -szor | egészséges)}}$$

$$= \frac{0.001 \cdot \binom{n}{x} 0.8^x (1-0.8)^{n-x}}{0.001 \cdot \binom{n}{x} 0.8^x (1-0.8)^{n-x} + 0.999 \cdot \binom{n}{x} 0.1^x (1-0.1)^{n-x}} =$$

$$= \frac{0.001 \cdot \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)}}{0.001 \cdot \text{BINOM.DIST(k ; n ; 0.8 ; FALSE)} + 0.999 \cdot \text{BINOM.DIST(k ; n ; 0.1 ; FALSE)}}$$

Az alábbi táblázat a P(beteg | betegnek diagnosztizálják k -szor) feltételes valószínűség numerikus értékeit tartalmazza a $0 < k \leq n \leq 5$ esetekre:

	5	4	3	2	1	k
1					0.008	
2				0.060	0.002	
3			0.339	0.014	0.000	
4		0.804	0.102	0.003	0.000	
5	0.970	0.477	0.025	0.001	0.000	
n						

Táblázat: *P(beteg | n vizsgálat során k -szor diagnosztizálják betegnek)*

Ez a másik a táblázat pedig az $n = 10$ és $k = 9, 8, 7, 6, 5, 4$ esetekről szól:

	9	8	7	6	5	4	k
$n = 10$	1.000	0.999	0.958	0.390	0.017	0.000	

Táblázat: *A $k = 9, 8, 7, 6, 5, 4$ esetek*

Érdemes elgondolkodni a numerikus értékek jelentésén!

Megjegyzés: A binomiális elolszlás $n = 1$ mellett vett speciális esete lesz az indikátor eloszlás, többdimenziós általánosítása pedig a polinomiális eloszlás.

1.3.2. Indikátor eloszlás

Súlyfüggvény:

$$p(x) = \begin{cases} p & \text{ha } x = 1 \\ 1 - p & \text{ha } x = 0 \end{cases}$$

Mikor használjuk: Egy eseménnyel kapcsolatban bizonyos gondolatmenetben hasznos, ha az esemény bekövetkezését egy kétértékű valószínűségi változóval kódoljuk: az 1 az esemény bekövetkezését, a 0 az esemény be nem következését jelenti:

$$X \;=\; \begin{cases} 1 & \text{ha az esemény bekövetkezik} \\ 0 & \text{ha az esemény nem következik be} \end{cases}$$

Ezt a valószínűségi változót az esemény **indikátorá**nak nevezzük. Ha például több eseménnyel kapcsolatban azt nézzük, hogy közülük hány következik be, akkor ez a valószínűségi változó az egyes események indikátorainak az összege. A binomiális eloszlással kapcsolatos számításoknál hasznos, hogy indikátorok összegeként binomiális eloszlású valószínűségi változót tudunk előállítani.

Paraméter jelentése:

$$p \;=\; \text{az esemény valószínűsége}$$

1.3.3. Binomiális eloszlás számsorozaton

Ha az $a_0 \,,\, a_1 \,,\, a_2 \,,\, \ldots \,,\, a_n$ sorozat egy számtani sorozatot alkot, akkor

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

képlettel definiált eloszlás neve: **binomiális eloszlás a** $a_0 \,,\, a_1 \,,\, a_2 \,,\, \ldots \,,\, a_n$ számok **halmazán**.

1. Példa. Ha egy eseménnyel kapcsolatban n kísérletet végzünk, akkor az esemény relatív gyakorisága – nyilvánvalóan – binomiális eloszlást követ a

$$\frac{0}{n} \,,\, \frac{1}{n} \,,\, \frac{2}{n} ,\, \ldots \,,\, \frac{k}{n} \,,\, \ldots \,,\, \frac{n}{n}$$

számok halmazán.

2. Példa. Ha egy n házaspárból álló társaságból minden házaspár a többitől függetlenül p valószínűséggel megy el egy összejövetelre, akkor az összejövetelen résztvevő *emberek* száma – nyilvánvalóan – binomiális eloszlást követ

$$0 \,,\, 2 \,,\, 4 \,,\, \ldots \,,\, 2n$$

számok halmazán.

3. Példa. Ha n háromgyerekes család mindegyike a többitől függetlenül p valószínűséggel vesz részt egy rendezvényen, akkor az résztvevő emberek száma – nyilvánvalóan – binomiális eloszlást követ a

$$0 \,,\, 5 \,,\, 10 \,,\, \ldots \,,\, 5n$$

számok halmazán.

1.3.4. Polinomiális eloszlás

Súlyfüggvény :

$$p(x,y) \;=\; \frac{n!}{x!\; y!\; (n-x-y)!}\;\; (p_1)^x\; (p_2)^y\;\; (1-p_1-p_2)^{n-x-y} \;=$$

$$=\; \binom{n}{x} \cdot \binom{n-x}{y}\;\; (p_1)^x\; (p_2)^y\;\; (1-p_1-p_2)^{n-x-y}$$

ha

$$0 \le x \le n$$

$$0 \le y \le n$$

$$0 \le n-x-y \le n$$

Mikor használjuk: Ha A és B egymást kizáró események, és n kísérletet végzünk velük kapcsolatban, akkor az a kétdimenziós valószínűségi változó, melynek koordinátái

$X \;=\;$ ahányszor bekövekezik az A esemény

$Y \;=\;$ ahányszor bekövekezik a B esemény

ilyen eloszlást követ.

Paraméterek jelentése:

$n \;=\;$ ahány kísérletet végzünk

$p_1 \;=\;$ az A esemény valószínűsége

$p_2 \;=\;$ a B esemény valószínűsége

A súlyfüggvény képletének levezetése: A levezetést egy példán keresztül mutajuk be. E célból tegyük fel, hogy egy 10 tagú társaságból mindenki, a többiektől függetlenül, vasárnap p_1 valószínűséggel megy el kirándulni, p_2 valószínűséggel pedig moziba megy. Jelölje X a kirándulók számát, Y pedig a moziba menők számát. Most azt kérdezzük, hogy mennyi az (X,Y) súlyfügvényének az értéke az $x=3$, $y=2$ helyen, azaz mennyi annak az eseménynek a valószínűsége, hogy $X=3$ és $Y=2$, vagyis hogy pontosan 3 ember megy el kirándulni, és pontosan 2 megy moziba:

$$p(3,2) \;=\; \mathrm{P}(\, X=3\,,\, Y=2\,) \;=\; ?$$

Válasz: Az, hogy 3 ember megy el kirándulni, és 2 megy moziba, bekövetkezhet például úgy, hogy a társaság 3 legöregebb tagja kirándulni megy, a 2 legfiatalabb moziba megy, a többi 5 pedig sem nem kirándul, sem nem megy moziba. Ennek a valószínűsége, a független eseményekre vonatkozó szorzási szabály miatt nyilván:

$$p_1\, p_1\, p_1\;\; p_2\, p_2\;\; (1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2)\,(1-p_1-p_2) \;=$$

$$=\; p_1^3\, p_2^2\, (1-p_1-p_2)^5$$

Az, hogy 3 ember megy el kirándulni, és 2 pedig moziba természetesen másképpen is lehetséges: a társaság akármelyik 3 tagja mehet kirándulni, akármelyik másik 2 tagja mehet moziba, és a többi 5 pedig se ide se oda nem megy. Akármelyik 3 kirándulós és 2 mozizós ember esetében annak a valószínűsége, hogy ők mennek kirándulni, illetve moziba, és a többi 5 pedig se ide se oda, ugyancsak

$$p_1^3\, p_2^2\, (1-p_1-p_2)^5$$

Ezért, ha nem fontos számunkra, hogy ki az a 3 ember, aki kirándulni megy, és ki az a 2 ember, aki moziba megy, csak az fontos számunkra, hogy pontosan 3 ember menjen kirándulni és pontosan 2 pedig moziba (márpedig az

$$X = 3 \quad \text{és} \quad Y = 2$$

esemény esetében ez a helyzet), akkor ennek az eseménynek a valószínűségét úgy kapjuk meg, hogy a

$$p_1^3\, p_2^2\, (1 - p_1 - p_2)^5$$

közös valószínűség értéket megszorozzuk a 3 kiránduló és a 2 mozizó kiválasztási lehetőségeinek számával, vagyis

$$\binom{10}{3} \binom{7}{2}$$

-vel:

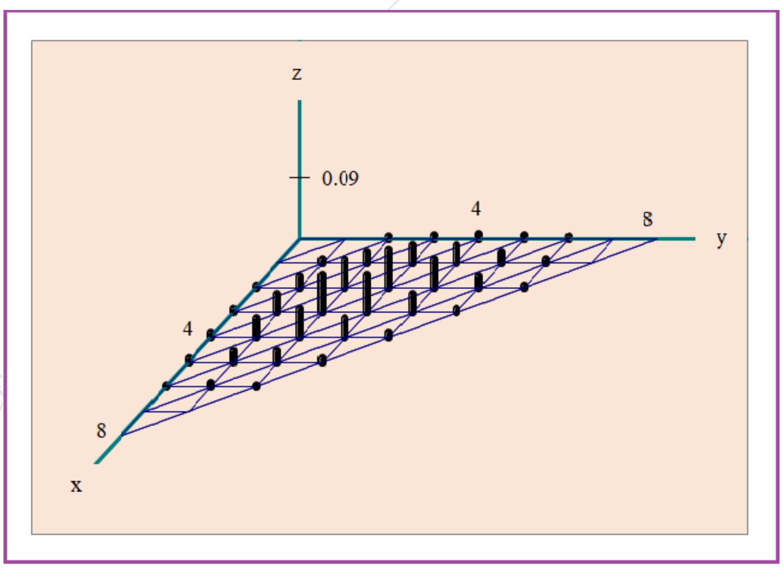
$$p(3,2) \;=\; \mathrm{P}(\, X = 3\,,\, Y = 2\,) \;=\; \binom{10}{3} \binom{7}{2}\, p_1^3\, p_2^2\, (1 - p_1 - p_2)^5$$

A binomiális együtthatókat faktoriálisokkal kifejtve, egyszerűsítés után az eredményt így is írhatjuk:

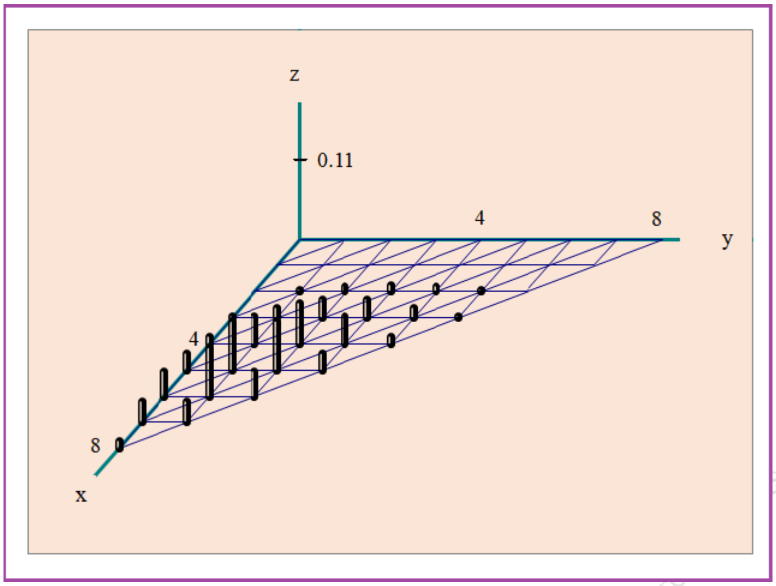
$$p(3,2) \;=\; \mathrm{P}(\, X = 3\,,\, Y = 2\,) \;=\; \frac{10!}{3!\;2!\;5!}\, p_1^3\, p_2^2\, (1 - p_1 - p_2)^5$$

Világos, hogy ha a társaság nem 10 tagú, hanem n tagú, akkor annak a valószínűsége, hogy az n ember közül pontosan x ember megy el kirándulni, és y ember mozizni:

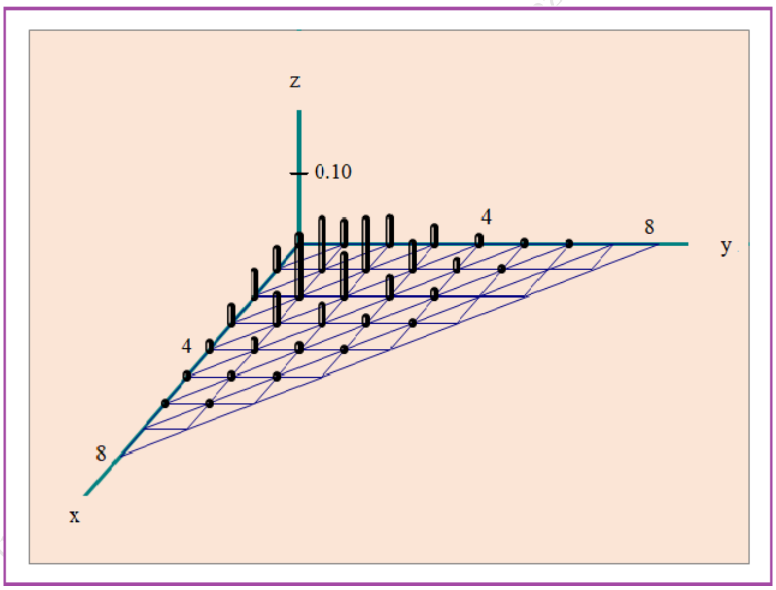
$$p(x,y) \;=\; \mathrm{P}(\, X = x\,,\, Y = y\,) \;=\; \frac{n!}{x!\;y!\;(n - x - y)!}\, p_1^x\, p_2^y\, (1 - p_1 - p_2)^{n-x-y}$$



12. ábra. *Polinomiális eloszlás súlyfüggvénye;* $n = 8;\; p_1 = 0.33,\; p_2 = 0.33$



13. ábra. *Polinomiális eloszlás súlyfüggvénye;* $n = 8;\ p_1 = 0.6,\ p_2 = 0.2$



14. ábra. *Polinomiális eloszlás súlyfüggvénye;* $n = 8;\ p_1 = 0.2,\ p_2 = 0.2$

Példa: Ha szabályos dobókockával 20 -szor dobunk, és

$$X \;=\; \text{ahányszor 1 -est dobunk}$$

$$Y \;=\; \text{ahányszor 2 -est vagy 3 -ast dobunk}$$

akkor az (X, Y) valószínűségi változó kétdimenziós polinomiális eloszlást követ $n = 20$, $p_1 = \frac{1}{6}$, $p_2 = \frac{2}{6}$ paraméterekkel. A súlyfüggvény képlete:

$$p(x, y) \;=\; \frac{20!}{x!\; y!\; (20 - x - y)!} \; \left(\frac{1}{6}\right)^{x} \left(\frac{2}{6}\right)^{y} \left(1 - \frac{1}{6} - \frac{2}{6}\right)^{(20-x-y)}$$

Például az $X = 3$ és $Y = 5$ esemény valószínűsége:

$$\mathrm{P}(\ X = 3\ ,\ Y = 5\ )\ =\ \frac{20!}{3!\ 5!\ (20-3-5)!}\ \left(\frac{1}{6}\right)^3\ \left(\frac{2}{6}\right)^5\ \left(1-\frac{1}{6}-\frac{2}{6}\right)^{(20-3-5)}$$

Megjegyzés: Vegyük észre, hogy az itt bevezetett polinomiális eloszlás a binomiális eloszlás általánosítása egydimenzióról kétdimenzióra.

Vetület eloszlások:

Teljesen nyilvánvaló, hogy az

$$X\ =\ \text{ahányszor bekövekezik az } A \text{ esemény}$$

valószínűségi változó binomiális eloszlást követ n és p_1 paraméterekkel, vagyis egy polinomiális eloszlásnak a vízszintes tengelyre vetett vetülete (és természetesen a függőleges tengelyre vetett vetülete is) binomiális eloszlás. Ez a tény a

$$p_1(x)\ =\ \sum_y\ p(x,y)$$

összegzési szabállyal is kijön, hiszen

$$\sum_y \binom{n}{x}\cdot\binom{n-x}{y}\ \cdot\ (p_1)^x\ \cdot\ (p_2)^y\ (1-p_1-p_2)^{n-x-y}\ =$$

$$=\ \binom{n}{x}\ \cdot\ (p_1)^x\ \cdot\ \sum_y \binom{n-x}{y}\ (p_2)^y\ (1-p_1-p_2)^{n-x-y}\ =$$

$$=\ \binom{n}{x}\ \cdot\ p_1^x\ \cdot\ (1-p_1)^{n-x}$$

Az első lépésben kiemelést végeztünk, a másodikban – az algebrában tanult "binomiális tétel"-re támaszkodva – felhasználtuk, hogy

$$\sum_y \binom{n-x}{y}\ (p_2)^y\ (1-p_1-p_2)^{n-x-y}\ =\ (1-p_1)^{n-x}$$

azonosságot.

Feltételes eloszlások:

- Y feltételes eloszlása az $X = x$ feltétel mellett binomiális eloszlás $n - x$ és $\frac{p_2}{1-p_1}$ paraméterekkel. Az állítást heurisztikus gondolatmenet is kiadja, és – természetesen – osztással, egyszerűsítésekkel is kijön:

$$p_{2|1}(\,y\,|\,x\,)\ =\ \mathrm{P}(\,Y = y\,|\,X = x\,)\ =\ \frac{\mathrm{P}(\,X = x\,,\,Y = y\,)}{\mathrm{P}(\ X = x)}\ =$$

$$=\ \frac{\binom{n}{x}\cdot\binom{n-x}{y}\ \ (p_1)^x\ (p_2)^y\ \ (1-p_1-p_2)^{n-x-y}}{\binom{n}{x}p_1^x(1-p_1)^{n-x}}\ =$$

$$=\ \binom{n-x}{y}\left(\frac{p_2}{1-p_1}\right)^y\left(1-\frac{p_2}{1-p_1}\right)^{n-x-y}$$

Hasonlóan kiadódik az is, hogy

- X feltételes eloszlása az $Y = y$ feltétel mellett binomiális eloszlás $n - y$ és $\frac{p_1}{1-p_2}$ paraméterekkel.

A kétdimenziós súlyfüggvény előállítása Excelben:

A kétdimenziós súlyfüggvényt – mint tudjuk – a vetület eloszlás és a feltételes eloszlás súlyfüggvényeiből szorzatként állíthatjuk elő a

$$p(x,y) \;=\; p_1(\,x\,) \;\cdot\; p_{2|1}(\,y\,|\,x\,)$$

$$p(x,y) \;=\; p_2(\,y\,) \;\cdot\; p_{1|2}(\,x\,|\,y\,)$$

szorzási szabályok alapján. Ezért – fentebbiek figyelembe vételével – a kétdimenziós polinomiális eloszlás $p(x,y)$ súlyfüggvényét Excellel az alábbi képletek segítségével tudjuk előállítani:

$$p(x,y) \;=\; \texttt{BINOM.DIST}(\,x\,;\,n\,;\,p_1\,;\,\texttt{FALSE}\,) \cdot \texttt{BINOM.DIST}(\,y\,;\,n-x\,;\,\frac{p_2}{1-p_1}\,;\,\texttt{FALSE}\,)$$

$$p(x,y) \;=\; \texttt{BINOM.DIST}(\,y\,;\,n\,;\,p_2\,;\,\texttt{FALSE}\,) \cdot \texttt{BINOM.DIST}(\,x\,;\,n-y\,;\,\frac{p_1}{1-p_2}\,;\,\texttt{FALSE}\,)$$

1.3.5. Polinomiális eloszlás, r -dimenziós *(Extra tananyag)*

Súlyfüggvény:

$$p(x_1, x_2, \ldots, x_r) \;=$$

$$= \frac{n!}{x_1!\; x_2! \ldots x_r!\; (n - x_1 - x_2 - \ldots - x_r)!}\; (p_1)^{x_1}\, (p_2)^{x_2} \;\ldots (p_r)^{x_r}\; (1 - p_1 - p_2 - \ldots - p_r)^{(n - x_1 - x_2 - \ldots - x_r)}$$

ha

$$0 \le x_i \le n \qquad (1 \le i \le r)$$

$$0 \le n - x_1 - x_2 - \ldots - x_r \le n$$

Mikor használjuk: Ha $A_1, A_2, \ldots, A_r$ egymást kizáró események, és n kísérletet végzünk velük kapcsolatban, akkor az az r -dimenziós valószínűségi változó, melynek koordinátái

X_1 = ahányszor bekövetkezik az A_1 esemény

X_2 = ahányszor bekövetkezik az A_2 esemény

$\vdots$

X_r = ahányszor bekövetkezik az A_r esemény

Paraméterek jelentése:

p_1 = az A_1 esemény valószínűsége

p_2 = az A_2 esemény valószínűsége

$\vdots$

p_r = az A_r esemény valószínűsége

Példa: Ha szabályos dobókockával 20 -szor dobunk, és

$$
\begin{aligned}
X_1 &= \text{ahányszor 1 -est dobunk} \\
X_2 &= \text{ahányszor 2 -est vagy 3 -ast dobunk} \\
X_3 &= \text{ahányszor 4 -est dobunk}
\end{aligned}
$$

akkor az (X_1, X_2, X_3) valószínűségi változó 3-dimenziós polinomiális eloszlást követ

$$n = 20 , \qquad p_1 = \frac{1}{6} , \qquad p_2 = \frac{2}{6} , \qquad p_3 = \frac{1}{6}$$

paraméterekkel.

Megjegyzés: Vegyük észre, hogy az itt bevezetett polinomiális eloszlás a binomiális eloszlás általánosítása egydimenzióról r -dimenzióra.

1.4. Különböző valószínűségű események közül hány következik be? *(Extra tananyag)*

Tegyük fel, hogy $A_1, A_2, \ldots, A_n$ független események. Valószínűségeiket jelöljük $p_1, p_2, \ldots, p_n$ -nel. Érdekelhet minket, hogy az események közül hány következik be. Ezt a valószínűségi változó jelöljük X_n -nel:

$$X_n = \text{ahány bekövetkezik az } A_1, A_2, \ldots, A_n \text{ események közül}$$

Ennek a valószínűségi változónak az eloszlását ismerjük abban a speciális esetben, amikor a $p_1, p_2, \ldots, p_n$ valószínűségek egyenlőek: binomiális eloszlás n és p paraméterekkel, ahol a paraméter p a valószínűségek közös értékét jelenti. Most azzal az esettel foglalkozunk, amikor a $p_1, p_2, \ldots, p_n$ valószínűségek nem mind egyenlőek.

Mindjárt látni fogjuk, hogy ebben az általánosabb esetben X_n eloszlásának tagjaira egyszerű rekurzív képletet adható. A rekurzív képlet tényleg egyszerű lesz: csak összeadásokat és szorzásokat tartalmaz, ami lehetővé teszi, hogy kis n értékekre kézi számolással vagy kalkulátorral, nagyobb n értékekre pedig számítógép segítségével meghatározhassuk az eloszlás tagjait.

A rekurzív képlet levezetéshez szükség van az

$$X_k = \text{ahány esemény bekövetkezik } A_1 \; A_2, \ldots \; A_k \text{ közül}$$

valószínűségi változókra is $(k = 1, 2, \ldots, n)$. Előre bocsátjuk, hogy

- az $X_k = k$ esemény azt jelenti, hogy az $A_1 \; A_2, \ldots \; A_k$ események mind bekövetkeznek,
- az $X_k = 0$ esemény azt jelenti, hogy az $A_1 \; A_2, \ldots \; A_k$ események egyike sem következik be,
- az $X_k = i$ esemény azt jelenti, hogy az $A_1 \; A_2, \ldots \; A_k$ események közül pontosan i darab következik be.

A $\mathrm{P}(X_k = k)$ valószínűség értéke nyilvánvaló:

$$
\begin{aligned}
&\mathrm{P}(X_k = k) = \\
&= \mathrm{P}(\text{az } A_1 \; A_2, \ldots \; A_k \text{ események mind bekövetkeznek}) = p_1 \cdot p_2 \cdot \ldots \cdot p_k
\end{aligned}
$$

Elnevezhetnénk ezt a a szabályt *az események bekövetkezésével kapcsolatos szorzási szabály*nak, de a rövidség kedvéért inkább *első szorzási szabály*nak nevezzük.

A $\mathrm{P}(X_k = 0)$ valószínűség értéke is nyilvánvaló:

$\mathrm{P}(X_k = 0) \; =$

$= \; \mathrm{P}(\text{az } A_1 \, A_2, \ldots \, A_k \text{ események közül egyik sem következik be}) \; = \; (1 - p_1) \cdot (1 - p_2) \cdot \ldots \cdot (1 - p_k)$

Elnevezhetnénk ezt a a szabályt *az események be nem következésével kapcsolatos szorzási szabály*nak, de a rövidség kedvéért inkább *második szorzási szabály*nak nevezzük.

A $\mathrm{P}(X_k = i)$ valószínűségre pedig $0 < i < k$ esetén nyilván igaz az alábbi rekurzív összefüggés:

$\mathrm{P}(X_k = i) \; =$

$= \; \mathrm{P}(X_{k-1} = i - 1) \cdot \mathrm{P}(\; A_k \text{ bekövetkezik}) \; + \; \mathrm{P}(X_{k-1} = i) \cdot \mathrm{P}(\; A_k \text{ nem következik be}) \; =$

$= \; \mathrm{P}(X_{k-1} = i - 1) \cdot p_k \; + \; \mathrm{P}(X_{k-1} = i) \cdot (1 - p_k)$

hiszen az, hogy az $A_1, A_2, \ldots, A_k$ események közül pontosan i következik be, az alábbi kétféle, egymást kizáró módon lehetséges:

- vagy úgy, hogy A_1 , A_2 , ... , A_{k-1} közül pontosan $i - 1$ következik be, és A_k is bekövetkezik
- vagy úgy, hogy A_1 , A_2 , ... , A_{k-1} közül pontosan i következik be, és A_k nem következik be

Ezt a szabályt egyszerűen *rekurzív szabály*nak nevezzük.

1. Példa. Bizonyára lesz, akinek segít, ha most egy konkrét numerikus példát mutatunk, és elemzünk. E célból tegyük fel, hogy az A_1 , A_2 , A_3 , A_4 események függetlenek, és valószínűségeik:

$$p_1 \; = \; 0.1 \qquad p_2 \; = \; 0.2 \qquad p_3 \; = \; 0.3 \qquad p_4 \; = \; 0,4$$

Az A_1 , A_2 , A_3 , A_4 eseményekkel kapcsolatban értelmezett X_1 , X_2 , X_3 , X_4 valószínűségi változók eloszlásait táblázatba foglaltuk. A táblázat bal oldalán feltüntettük az események és a komplementereik valószínűségeit:

		X_i lehetséges értékei	0	1	2	3	4
$\mathrm{P}(A_1) = 0.1$	$\mathrm{P}(\overline{A_1}) = 0.9$	X_1 **eloszlása:**	0.9	0.1			
$\mathrm{P}(A_2) = 0.2$	$\mathrm{P}(\overline{A_2}) = 0.8$	X_2 **eloszlása:**	0.72	0.26	0.02		
$\mathrm{P}(A_3) = 0.3$	$\mathrm{P}(\overline{A_3}) = 0.7$	X_3 **eloszlása:**	0.504	0.398	0.092	0.006	
$\mathrm{P}(A_4) = 0.4$	$\mathrm{P}(\overline{A_4}) = 0.6$	X_4 **eloszlása:**	0.3024	0.4404	0.2144	0.0404	0.0024

Táblázat: *Az X_1 , X_2 , X_3 , X_4 valószínűségi változók eloszlásai*

A fentebb kimondott szabályok ezeken a numerikus értékeken keresztül ellenőrizhetők. Például:

- $\mathrm{P}(\; X_2 = 2 \;) \; = \; 0.02 \; = \; 0.1 \cdot 0.2$ *(első szorzási szabály)*

- P($X_2 = 1$) $=\ 0.26\ =\ 0.9\cdot 0.2\ +\ 0.1\cdot 0.8$ *(rekurzív szabály)*
- P($X_2 = 0$) $=\ 0.72\ =\ 0.9\cdot 0.8$ *(második szorzási szabály)*

- P($X_3 = 3$) $=\ 0.006\ =\ 0.1\cdot 0.2\cdot 0.3$ *(első szorzási szabály)*
- P($X_3 = 1$) $=\ 0.398\ =\ 0.72\cdot 0.3\ +\ 0.26\cdot 0.7$ *(rekurzív szabály)*
- P($X_3 = 2$) $=\ 0.092\ =\ 0.26\cdot 0.3\ +\ 0.02\cdot 0.7$ *(rekurzív szabály)*
- P($X_3 = 0$) $=\ 0.504\ =\ 0.9\cdot 0.8\cdot 0.7$ *(második szorzási szabály)*

- P($X_4 = 4$) $=\ 0.0024\ =\ 0.1\cdot 0.2\cdot 0.3\cdot 0.4$ *(első szorzási szabály)*
- P($X_4 = 1$) $=\ 0.4404\ =\ 0.504\cdot 0.4\ +\ 0.398\cdot 0.6$ *(rekurzív szabály)*
- P($X_4 = 2$) $=\ 0.2144\ =\ 0.398\cdot 0.4\ +\ 0.092\cdot 0.6$ *(rekurzív szabály)*
- P($X_4 = 3$) $=\ 0.0404\ =\ 0.092\cdot 0.4\ +\ 0.006\cdot 0.6$ *(rekurzív szabály)*
- P($X_4 = 0$) $=\ 0.3024\ =\ 0.9\cdot 0.8\cdot 0.7\cdot 0.6$ *(második szorzási szabály)*

Ebben a példában 4 esemény kapcsán 4 lépésben jutottunk el az X_4 valószínűségi változó eloszlásához a táblázat első, fejléces sorát 4 további sor követi.

Ha – mondjuk – 400 esemény kapcsán foglalkozunk azzal a valószínűségi változóval, ami megmondja, hogy a 400 esemény közül hány következik be, akkor ennek a valószínűségi változónak az eloszlásához lényegében ugyanígy, csak hosszadalmasabb számolással (400 lépésben) juthatunk el. Ez a hosszadalmas számolás számítógéppel, például Excellel, egyáltalán nem hosszadalmas, csak – a táblázat felépítésénél – okosan kell lefordítani a rekurzív képletet helyes Excel képletté.

2. Példa: Diákok különböző szorgalommal. 400 hallgató mindegyike egymástól függetlenül jár órára, de nem egyforma szorgalommal:

- a legszorgalmasabb 0.800
- a második legszorgalmasabb 0.799
- a harmadik legszorgalmasabb 0.798
- és így tovább
- a legkevésbé szorgalmas 0.401

valószínűséggel megy el az órákra. Hány szék kell ahhoz, hogy 0.99 biztonsággal jusson szék mindenkinek, aki elmegy órára?

Megoldás: Excellel először elkészítettük – a rekurzív képlet segítségével, 400 lépésben – az órát látogató diákok száma súlyfüggvényének a táblázatát, majd abból összegzéssel az eloszlásfüggvény táblázatát. Ezeket a táblázatokat nagy méretük miatt a könyv sorai közé nem tesszük be. Az Olvassó megtekintheti őket a

`http://math.bme.hu/~vetier/2016_osz/A4_vill_2016_osz.html` címen a lap felső részében a mellékletek között található

`400 különböző diák` linken található Excel fájlban.

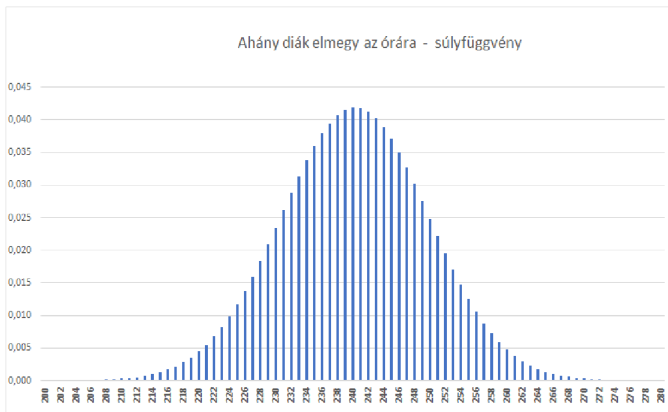
Ízelítőül a súlyfüggvény és az eloszlásfüggvény ábráit megmutajuk: lásd az "Ahány diák elmegy az órára– súlyfüggvény" és az "Ahány diák elmegy az órára – eloszlásfüggvény" feliratú ábrákat.

Az eloszlásfüggvény táblázatából idemásoltuk azt a részt, ahol az $F(x)$ átlépi a 0.99 értéket:

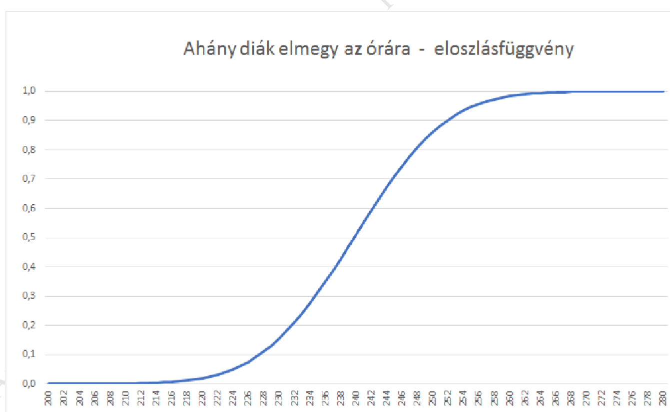
x	260	261	262	263
$F(x)$	0.984	0.988	0.991	0.993

Ahol az $F(x)$ *átlépi a* 0.99 *értéket*

Látjuk, hogy $x = 261$ esetén $F(x) = 0.988$, de $x = 262$ esetén $F(x) = 0.991$. Tehát 261 szék még nem, de 262 már elegendő!



15. ábra. *Ahány diák elmegy az órára– súlyfüggvény*



16. ábra. *Ahány diák elmegy az órára – eloszlásfüggvény*

1.5. Geometriai eloszlások és társaik

1.5.1. Geometriai eloszlás (optimista)

Súlyfüggvény:

$$p(x) = (1-p)^{x-1}p \qquad \text{ha} \quad x = 1, 2, \ldots$$

Megjegyzés: Vegyük észre, hogy a súlyfüggvény értékei mértani sorozatot alkotnak. A sorozat első tagja p , a kvóciense $q = 1 - p$.

Eloszlásfüggvény:

$$F(x) \ = \ 1 - (1-p)^x \quad (x = 1, 2, \ldots)$$

Jobboldali eloszlásfüggvény:

$$T(x) \ = \ q^{x-1} \quad (x = 1, 2, \ldots)$$

ahol $q \ = \ 1 - p$.

Megjegyzés: Vegyük észre, hogy a jobboldali eloszlásfüggvény nem más, mint egy q alapú exponenciális függvény, mely az $x = 1, 2, \ldots$ értékeken értelmezett, $q = 1 - p$.

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:**Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X \ = \ \text{ahányadik kísérletnél először következik be az esemény}$$

   Paraméter jelentése:

   $$p \ = \ \text{az esemény valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén

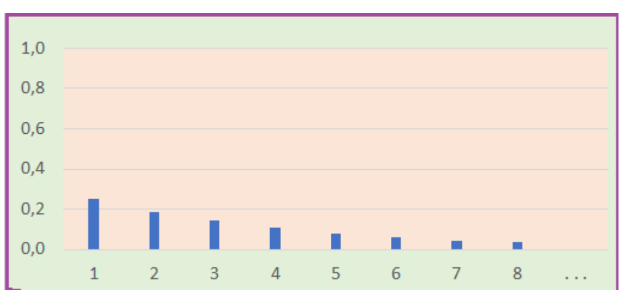
   $$X \ = \ \text{ahányadik esemény az első olyan, ami bekövetkezik}$$

   Paraméter jelentése:

   $$p \ = \ \text{az események közös valószínűség értéke}$$

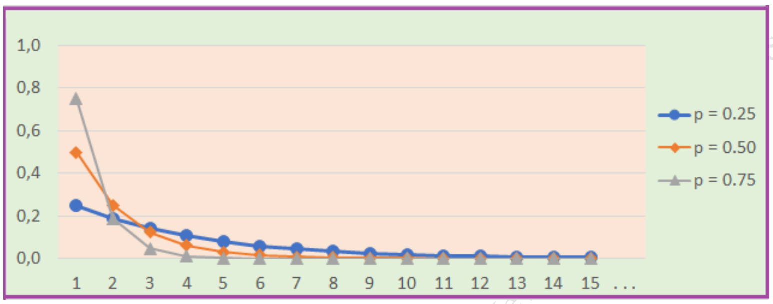
A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelöli annak a kísérletnek a sorszámát, amikor az A esemény először következik be, akkor az $X = x$ esemény azt jelenti, hogy az első $x - 1$ kísérlet során az A nem következett be, de az x -ik kísérletnél A bekövetkezett. Ezért – a független eseményekre vonatkozó szorzási szabály felhasználva – ezt kapjuk:

$$p(x) \ = \ \mathrm{P}\left(X = x\right) \ = \ (1-p)^{x-1}p$$

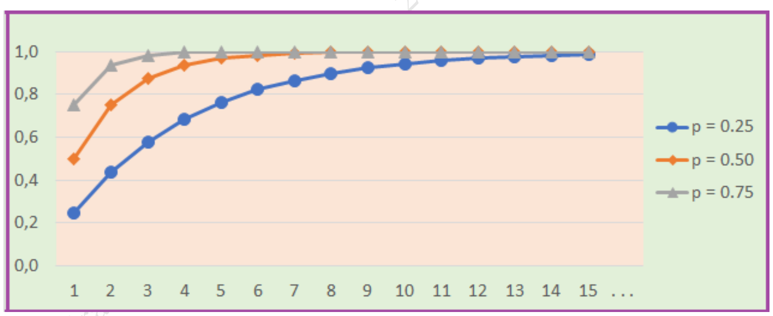


17. ábra. *Geometriai (optimista) eloszlás súlyfüggvénye;* $p = 0.25$

18. ábra. *Geometriai (optimista) eloszlás eloszlásfüggvénye;* $p = 0.25$



19. ábra. *Geometriai (optimista) eloszlások súlyfüggvényei;* $p = 0.25,\ 0.50,\ 0.75$



20. ábra. *Geometriai (optimista) eloszlások eloszlásfüggvényei;* $p = 0.25,\ 0.50,\ 0.75$

1. Feladat: Vadászat az első nyúlig. Vadász barátom hétköznapokon (hétfőtől péntekig) a rosszabbik puskáját használja. Ezzel a puskával minden lövése a többitől függetlenül csak 0.05 valószínűséggel teríti le a célba vett nyulat. Hétvégén (szombaton és vasárnap) a jobb puskájával lövöldöz. Ezzel a puskával minden lövése a többitől függetlenül 0.4 valószínűséggel teríti le a célba vett nyulat. Minden nap addig lövöldöz a nyulakra, míg sikerül egyet leterítenie.

Első kérdés: Azt mesélik, hogy egy bizonyos napon, melyről nem tudjuk, hogy a hét melyik napja volt, 7 lövést adott le a nyúlvadászaton, hogy teljesítse aznapi feladatát. Ez a bizonyos nap vajon hétköznap volt, vagy hétvége?

Második kérdés: Hogyan lehet következtetni a napi lövések számából arra, hogy a lövések napja vajon hétköznap volt, vagy hétvége?

Megoldás: Tekintsünk egy hétköznapot. A nyúl leterítéséhez szükséges lövések száma optimista geometriai eloszlást követ 0.05 paraméterrel. Ezért egy hétköznapon annak a valószínűsége, hogy k lövés dördül,

$$\mathrm{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) \ = \ 0.05\,(1-0.05)^{k-1}$$

Egy hétvégi napokon annak a valószínűsége, hogy k lövés dördül,

$$\text{P(} k \text{ lövés } | \text{ hétvégi nap)} \;=\; 0.4\,(1-0.4)^{k-1}$$

Mivel a napok $5/7$ -e hétköznap, $2/7$ -e hétvégi nap, annak a valószínűsége, hogy egy véletlenszerűen választott napon k lövés dördül,

$$\frac{5}{7} \cdot 0.05\,(1-0.05)^{k-1} \;+\; \frac{2}{7} \cdot 0.4\,(1-0.4)^{k-1}$$

Ha egy véletlenszerűen választott napon k lövés dördül, akkor annak a valószínűsége, hogy ez a nap hétköznap volt:

$$\text{P(hétköznap } | \text{ } k \text{ lövés)} \;=\; \frac{\frac{5}{7} \cdot 0.05\,(1-0.05)^{k-1}}{\frac{5}{7} \cdot 0.05\,(1-0.05)^{k-1} \;+\; \frac{2}{7} \cdot 0.4\,(1-0.4)^{k-1}}$$

Annak a valószínűsége pedig, hogy hétvégi nap volt:

$$\text{P(hétvégi nap } | \text{ } k \text{ lövés)} \;=\; \frac{\frac{2}{7} \cdot 0.4\,(1-0.4)^{k-1}}{\frac{5}{7} \cdot 0.05\,(1-0.05)^{k-1} \;+\; \frac{2}{7} \cdot 0.4\,(1-0.4)^{k-1}}$$

Excellel könnyű kiszámolni a képletek numerikus értékét. Íme a táblázat:

k	P(hétköznap | k lövés)	P(hétvégi nap | k lövés)
1	0.24	0.76
2	0.33	0.67
3	0.44	0.56
4	0.55	0.45
5	0.66	0.34
6	0.76	0.24
7	0.83	0.17
8	0.89	0.11
9	0.93	0.07
10	0.95	0.05

Táblázat: *Hétköznap és hétvégi nap?*

Válasz az első kérdésre: A táblázatból kiolvassuk, hogy $k = 7$ esetén a hétköznap valószínűsége 0.83 , a hétvégi nap valószínűsége 0.17

Válasz a második kérdésre: A táblázat azt is mutatja, hogy $k = 1, 2, 3$ esetén a hétvégi nap a valószínűbb, más lövészszám esetén a hétköznap a valószínűbb.

1.5.2. Geometriai eloszlás (pesszimista)

Súlyfüggvény:

$$p(x) \;=\; (1-p)^x p \qquad \text{ha} \quad x = 0, 1, 2, \ldots$$

Megjegyzés: Vegyük észre, hogy a súlyfüggvény értékei mértani sorozatot alkotnak. A sorozat első tagja p , a kvóciense $q = 1 - p$. Az optimista geometriai eloszláshoz képest ez az eloszlás annyiban más, hogy most a lehetséges értékek az $x = 0, 1, \ldots$ számok, míg az optimista esetben a lehetséges értékek az $x = 1, 2, \ldots$ számok. Szemléletesen úgy fogalmazhatunk, hogy a pesszimista geometriai eloszlás súlyfüggvényének az ábrája úgy származtatható az optimistából, hogy a súlyfügvény grafikonját egy egységgel balra toljuk. Az optimista geometriai eloszlás súlyfüggvényének az ábrája pedig úgy származtatható az pesszimistából, hogy a súlyfügvény grafikonját egy egységgel jobbra toljuk.

Eloszlásfüggvény:

$$F(x) \ = \ 1 - (1-p)^{x+1} \quad (x = 0, 1, 2, \ldots)$$

Jobboldali eloszlásfüggvény:

$$T(x) \ = \ q^x \quad (x = 0, 1, 2, \ldots)$$

ahol $q \ = \ 1 - p$.

Megjegyzés: Vegyük észre, hogy a jobboldali eloszlásfüggvény nem más, mint egy q alapú exponenciális függvény, mely a $x = 0, 1, 2, \ldots$ értékeken értelmezett, $q \ = \ 1 - p$.

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X \ = \ \text{ahányszor nem következik be az esemény az első bekövetkezés előtt}$$

   Más terminológiát használva:

   $$X \ = \ \text{ahány kudarc van az első siker előtt}$$

   Paraméter jelentése:

   $$p \ = \ \text{a siker valószínűsége}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

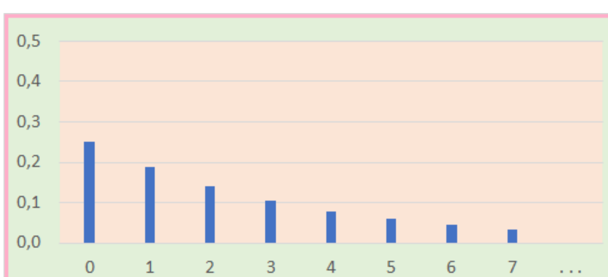
   $$X \ = \ \text{ahány kudarcos esemény van az első sikeres előtt}$$

   Paraméter jelentése:

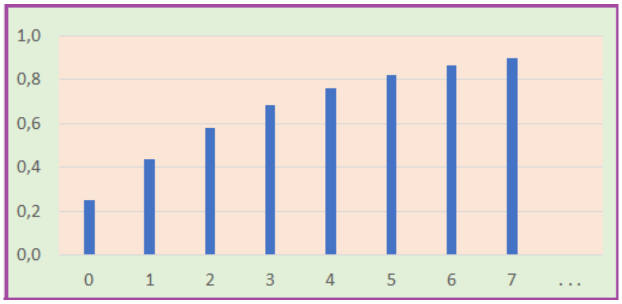
   $$p \ = \ \text{az események közös valószínűség értéke}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X azt jelölöli, hogy az A esemény első bekövetkezése előtt hányszor nem következett be, akkor az $X = x$ esemény azt jelenti, hogy az első x kísérlet során az A nem következett be, de az $(x+1)$ -ik kísérletnél A bekövetkezett. Ezért – a független eseményekre vonatkozó szorzási szabály felhasználva – ezt kapjuk:

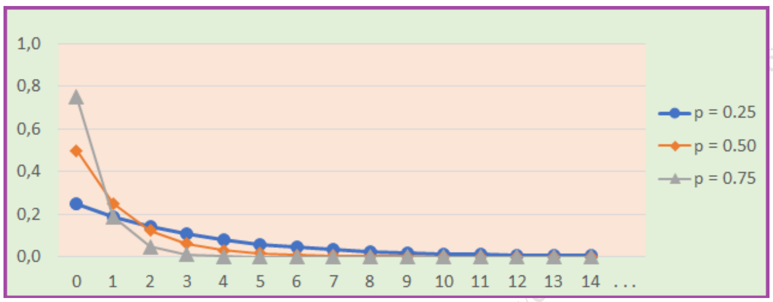
$$p(x) \ = \ \mathrm{P}\left(X = x\right) \ = \ (1-p)^x p$$



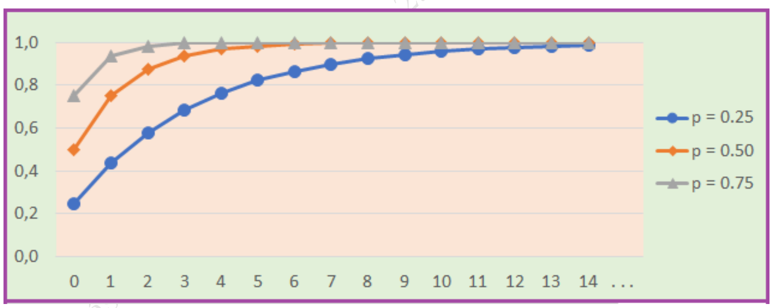
21. ábra. *Geometriai (pesszimista) eloszlás súlyfüggvénye;* $p = 0.25$



22. ábra. *Geometriai (pesszimista) eloszlás eloszlásfüggvénye;* $p = 0.25$



23. ábra. *Geometriai (pesszimista) eloszlások súlyfüggvényei;* $p = 0.25,\ 0.50,\ 0.75$



24. ábra. *Geometriai (pesszimista) eloszlások eloszlásfüggvényei;* $p = 0.25,\ 0.50,\ 0.75$

Megjegyzés: Az "optimista", "pesszimista" jelzők használatát az indokolja, hogy az optimista esetben **sikeres** kísérletre vadászunk, a pesszimista esetben pedig **a kudarcok** számát számoljuk.

1.5.3. Geometriai eloszlás az $\{a, a+1, a+2, \dots\}$ halmazon

Súlyfüggvény:

$$p(x) \;=\; (1-p)^{x-a}p \qquad \text{ha} \quad x = a, a+1, a+2, \dots$$

Megjegyzés: Vegyük észre, hogy a súlyfüggvény a -hoz, $a+1$ -hez, $a+2$ -höz, $\dots$ rendelt értékei mértani sorozatot alkotnak. A sorozat első tagja p , a kvóciense $q = 1-p$. A súlyfüggvény ábrája úgy származtatható a pesszimista geometriai eloszlás ábrájából, hogy a súlyfügvény grafikonját a egységgel jobbra toljuk.

Eloszlásfüggvény:

$$F(x) \;=\; 1-(1-p)^{x-a+1} \qquad (x=a,a+1,a+2,\dots)$$

Jobboldali eloszlásfüggvény:

$$T(x) \;=\; q^{x-a} \qquad (x=a,a+1,a+2,\dots)$$

ahol $q \;=\; 1-p$.

Mikor használjuk:

Ha X (pesszimista vagy optimista) geometriai eloszlású valószínűségi változó, akkor az $X \geq a$ feltétel melletti feltételes eloszlása ilyen eloszlás. A pesszimista esetre ezt így láthajuk be: akármilyen pozitív egész a és $x \geq a$ esetén

$$\mathrm{P}(\, X=x \mid X \geq a \,) \;=\; \frac{\mathrm{P}(\, X=x \,)}{\mathrm{P}(\, X \geq a \,)} \;=\; \frac{(1-p)^x p}{(1-p)^a} \;=\; (1-p)^{x-a} p$$

Optimista esetre pedig így: akármilyen pozitív egész a és $x \geq a$ esetén

$$\mathrm{P}(\, X=x \mid X \geq a \,) \;=\; \frac{\mathrm{P}(\, X=x \,)}{\mathrm{P}(\, X \geq a \,)} \;=\; \frac{(1-p)^{x-1} p}{(1-p)^{a-1}} \;=\; (1-p)^{x-a} p$$

1.5.4. Geometriai eloszlás jellemzése feltételes valószínűségekkel

A most következő gondolatok megértésének megsegítése érdekében előrebocsátjuk azt a triviális tényt, hogy ha X azt jelenti, hogy hányadik dobásra dobunk először 6 -ost, akkor a $\mathrm{P}(\, X \geq x+1 \mid X \geq x \,)$ feltételes valószínűség értéke minden x -re $\frac{5}{6}$ -dal egyenlő.

Vegyük észre, hogy ez általánosabban is igaz: ha X geometriai eloszlást követ, akkor a $\mathrm{P}(\, X \geq x+1 \mid X \geq x \,) = q$ feltételes valószínűség értéke nem függ x -től, mert minden x -re q-val egyenlő, ahol $q \;=\; 1-p$:

$$\mathrm{P}(\, X \geq x+1 \mid X \geq x \,) \;=\; \frac{\mathrm{P}(\, X \geq x+1 \,)}{\mathrm{P}(\, X \geq x \,)} \;=\; \frac{T(x+1)}{T(x)} \;=\; q$$

Szavakkal: ha tudjuk, hogy az X valószínűségi változó értéke nagyobb vagy egyenlő, mint x, akkor ilyen feltétel mellett annak a valószínűsége, hogy X értéke $x+1$ -nél is nagyobb vagy egyenlő, nem függ attól, hogy x mekkora.

Könnyű meggyőződni róla, hogy ez a tulajdonság jellemzi is a geometriai eloszlásokat: ha

$$\mathrm{P}(\, X=x \mid X \geq x \,) \;=\; p$$

fennáll minden $x \geq a$ esetén, akkor X eloszlása geometriai eloszlás.

1. Megjegyzés: A geometriai eloszlások bevezetésekor leszögeztük, hogy végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az

$$X \;=\; \text{ahányadik esemény az első olyan, ami bekövetkezik}$$

valószínűségi változó geometriai eloszlást követ.

Itt hívjuk fel a figyelmet, hogy ha az események függetlenek, de nem egyforma valószínűségűek, akkor az X valószínűségi változó nem követ geometriai eloszlást. Ha az események valószínűségeit $p_1, p_2, \dots$ jelöli, akkor

$$\mathrm{P}(\, X=x \,) \;=\; q_1 \dots q_{x-1}\, p_x$$

ahol $q_x = 1-p_x$ az x -ik esemény komplementerének valószínűségét jelöli $(x=1,2,\dots)$.

2. Megjegyzés: Érdemes megjegyezni, hogy a

$$\mathrm{P}(\ X = x\ )\ =\ q_1 \ldots q_{x-1}\, p_x$$

képlet érvényes akkor is, ha (az események nem függetlenek, viszont) p_x , illetve q_x az alábbi feltételes valószínűséget jelöli:

$$p_x \ =\ \mathrm{P}(\ \text{az } x\text{-ik esemény bekövetkezik} \mid \text{a korábbi események nem következtek be}\ )$$

$$q_x \ =\ \mathrm{P}(\ \text{az } x\text{-ik esemény nem következik be} \mid \text{a korábbi események nem következtek be}\ )$$

1.5.5. Negatív binomiális eloszlás (optimista)

Súlyfüggvény:

$$p(x)\ =\ \binom{x-1}{n-1} p^n (1-p)^{x-n} \quad \text{ha } x = n, n+1, n+2, \ldots$$

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X\ =\ \text{ahányadik kísérletnél } n \text{ -edszer következik be az esemény}$$

   Más terminlógiát használva:

   $$X\ =\ \text{ahányadik kísérletnél adódik az } n \text{ -ik sikeres esemény}$$

   Paraméterek jelentése:

   $$p\ =\ \text{az esemény valószínűsége}$$

   $$n\ =\ \text{ahányadik bekövetkezésre (sikerre) vadászunk}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X\ =\ \text{ahányadik esemény az } n \text{ -ik olyan, ami bekövetkezik}$$

   Más terminlógiát használva:

   $$X\ =\ \text{ahányadik az } n \text{ -ik sikeres esemény}$$

   Paraméterek jelentése:

   $$p\ =\ \text{az események közös valószínűség értéke}$$

   $$n\ =\ \text{ahányadik bekövetkezésre (sikerre) vadászunk}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelölöli annak a kísérletnek a sorszámát, amikor az A esemény n -edszer következik be, akkor az $X = x$ esemény azt jelenti, hogy az első $(x-1)$ kísérlet során az A esemény $(n-1)$ -szer következett be, és az A esemény az x -ik kísérletnél is bekövetkezik. Annak a valószínűsége, hogy az első $(x-1)$ kísérlet során az A esemény $(n-1)$ -szer következik be, a binomiális eloszlás képlete szerint

$$\binom{x-1}{n-1} p^{n-1} (1-p)^{(x-1)-(n-1)}\ =\ \binom{x-1}{n-1} p^{n-1} (1-p)^{x-n}$$

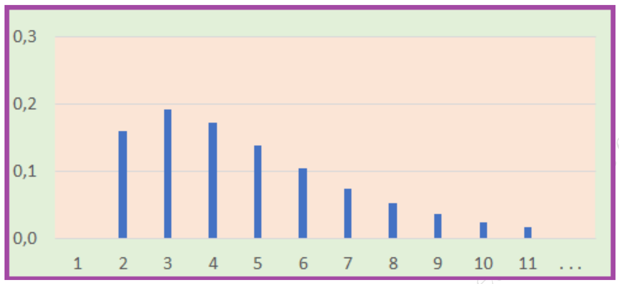
Mivel az x -ik kísérletnél az A esemény bekövetkezik, a keresett valószínűséget úgy kapjuk meg, hogy ezt a valószínűséget még megszorozzuk p -vel:

$$p(x) \;=\; \mathrm{P}\left(X = x\right) \;=\; \binom{x-1}{n-1} p^{n-1} (1-p)^{x-n)} \; p \;=\; \binom{x-1}{n-1} p^{n} (1-p)^{x-n}$$

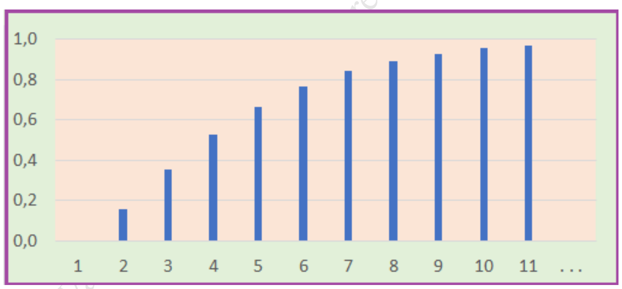
Excel-függvények:

$p(x)$ = `NEGBINOM.DIST(x-n;n;p;FALSE)` = `NEGBIN.ELOSZLÁS(x-n;n;p;HAMIS)`

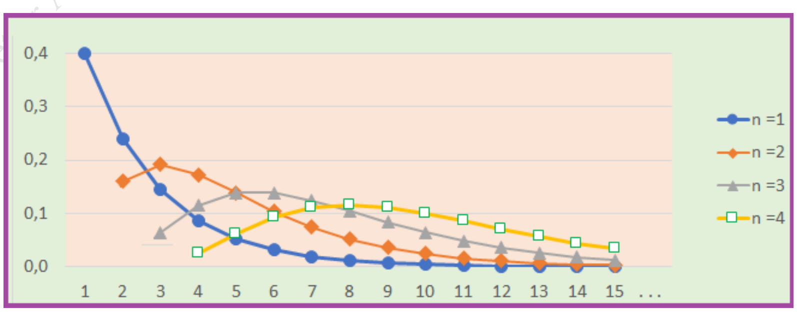
$F(x)$ = `NEGBINOM.DIST(x-n;n;p;TRUE)` = `NEGBIN.ELOSZLÁS(x-n;n;p;IGAZ)`



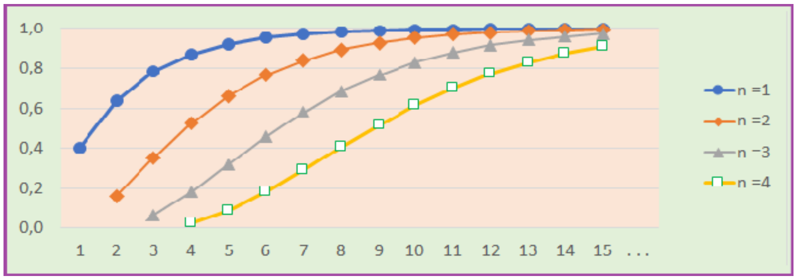
25. ábra. *Negatív binomiális (optimista) eloszlás súlyfüggvénye;* $n = 2;\; p = 0.4$



26. ábra. *Negatív binomiális (optimista) eloszlás eloszlásfüggvénye;* $n = 2;\; p = 0.4$



27. ábra. *Negatív binomiális (optimista) eloszlások súlyfüggvényei;* $n = 1,\; 2,\; 3,\; 4;\; p = 0.4$



28. ábra. *Negatív binomiális (optimista) eloszlások eloszlásfüggvényei;* $n = 1,\ 2,\ 3,\ 4;\ p = 0.4$

1. Feladat: Nyúlvadászat rafináltabb módon. Vadász barátom hétköznapokon (hétfőtől péntekig) a rosszabbik puskáját használja. Ezzel a puskával minden lövése a többitől függetlenül 0.3 valószínűséggel teríti le a célba vett nyulat. Hétköznapokon addig lövöldöz a nyulakra, míg sikerül négyet leterítenie. Hétvégén (szombaton és vasárnap) az ünnepi puskájával lövöldöz. Az ünnepi puskával minden lövése a többitől függetlenül 0.8 valószínűséggel teríti le a célba vett nyulat. Ünnepnapokon addig lövöldöz a nyulakra, míg sikerül hatot leterítenie.

Első kérdés: Azt mesélik, hogy egy bizonyos napon, melyről nem tudjuk, hogy a hét melyik napja volt, 7 lövést adott le a nyúlvadászaton, hogy teljesítse aznapi feladatát. Ez a bizonyos nap vajon hétköznap volt, vagy hétvége?

Második kérdés: Hogyan lehet következtetni a napi lövések számából arra, hogy a lövések napja vajon hétköznap volt, vagy hétvége?

Megoldás: Tekintsünk egy hétköznapot, amikor négy nyúl leterítése a feladat. Az ehhez szükséges lövések száma optimista negatív binomiális eloszlást követ 4 és 0.3 paraméterekkel. Ezért egy hétköznapon annak a valószínűsége, hogy k lövés dördül,

$$\text{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) = \texttt{NEGBINOM.DIST(k-4 ; 4 ; 0.3 ; FALSE)}$$

Egy hétvégi napon, amikor hat nyúl leterítése a feladat, a lövések száma optimista negatív binomiális eloszlást követ 6 és 0.8 paraméterekkel. Ezért hétvégi napokon annak a valószínűsége, hogy k lövés dördül,

$$\text{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ ) = \texttt{NEGBINOM.DIST(k-6 ; 6 ; 0.8 ; FALSE)}$$

Mivel a napok $5/7$ -e hétköznap, $2/7$ -e hétvégi nap, annak a valószínűsége, hogy egy véletlenszerűen választott napon k lövés dördül,

$$\frac{5}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) + \frac{2}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )$$

Ha egy véletlenszerűen választott napon k lövés dördül, akkor annak a valószínűsége, hogy ez a nap hétköznap volt:

$$\text{P}(\ \text{hétköznap} \mid k \text{ lövés}\ ) = \frac{\frac{5}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétköznap}\ )}{\frac{5}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) + \frac{2}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )}$$

Annak a valószínűsége pedig, hogy hétvégi nap volt:

$$\text{P}(\ \text{hétvégi nap} \mid k \text{ lövés}\ ) = \frac{\frac{2}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )}{\frac{5}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétköznap}\ ) + \frac{2}{7} \cdot \text{P}(\ k \text{ lövés} \mid \text{hétvégi napon}\ )}$$

Tekintve, hogy a képletek számlálóiban, illetve nevezőiben szereplő feltételes valószínűségek numerikus értékét – a fentebb adott képletekkel – Excelben ki lehet számolni a szóbajövő k értékekre, a keresett feltételes valószínűségekre is tudunk táblázatot készíteni:

k	P(hétköznap $\mid$ k lövés)	P(hétvégi nap $\mid$ k lövés)
4	1.00	0.00
5	1.00	0.00
6	0.27	0.73
7	0.31	0.69
8	0.44	0.56
9	0.62	0.38
10	0.79	0.21
11	0.90	0.10
12	0.96	0.04
13	0.99	0.01
14	0.99	0.01
15	1.00	0.00

Táblázat: *Hétköznap és hétvégi nap? - rafináltabb vadászat esetén*

Válasz az első kérdésre: A táblázatból kiolvassuk, hogy $k = 7$ esetén a hétköznap valószínűsége 0.31 , a hétvégi nap valószínűsége 0.69

Válasz a második kérdésre: A táblázat azt is mutatja, hogy $k = 6, 7, 8$ esetén a hétvégi nap a valószínűbb, más lövészszám esetén a hétköznap a valószínűbb.

1.5.6. Negatív binomiális eloszlás (pesszimista)

Súlyfüggvény:

$$p(x) = \binom{x+n-1}{n-1} p^n (1-p)^x \quad \text{ha} \quad x = 0, 1, 2, \ldots \quad \text{(kombinatorikus formula)}$$

$$p(x) = \binom{-n}{x} p^n \left(-(1-p) \right)^x \quad \text{ha} \quad x = 0, 1, 2, \ldots \quad \text{(analitikus formula)}$$

Megjegyzés: Vegyük észre, hogy az optimista negatív binomiális eloszláshoz képest ez az eloszlás annyiban más, hogy most a lehetséges értékek az $x = 0, 1, \ldots$ számok, míg az optimista esetben a lehetséges értékek az $x = n, n+1, n+2, \ldots$ számok. Szemléletesen úgy fogalmazhatunk, hogy a pesszimista negatív binomiális eloszlás grafikonja az optimistából úgy származtatható, hogy a grafikont n egységgel balra toljuk.

Mikor használjuk:

1. **Egy eseménnyel kapcsolatban:** Ha egy p valószínűségű eseményre egy kísérletsorozatot végzünk, akkor az alábbi valószínűségi változó követ ilyen eloszlást:

   $$X = \text{ahány kudarc van az } n \text{-ik siker előtt}$$

   Paraméterek jelentése:

   $$p = \text{az esemény valószínűsége}$$

   $$n = \text{ahányadik bekövetkezésre vadászunk}$$

2. **Több eseménnyel kapcsolatban:** Végtelen sok független, külön-külön p valószínűségű esemény sorozata esetén az alábbi valószínűségi változó követ ilyen eloszlást:

$$X \;=\; \text{ahány kudarcos esemény van az } n \text{-ik sikeres előtt}$$

Paraméterek jelentése:

$$p \;=\; \text{az események közös valószínűség értéke}$$

$$n \;=\; \text{ahányadik bekövetkezésre vadászunk}$$

A súlyfüggvény képletének levezetése: Ha egy p valószínűségű A eseményre végzett független kísérletsorozattal kapcsolatban X jelölöli azt, hogy az A esemény az n -edik bekövetkezése előtt hányszor nem következett be, akkor az $X = x$ esemény azt jelenti, hogy az első $(x + n - 1)$ kísérlet során az A esemény $(n - 1)$ -szer következett be, és az A esemény az $(x + n)$ -ik kísérletnél is bekövetkezik. Annak a valószínűsége, hogy az első $(x + n - 1)$ kísérlet során az A esemény $(n - 1)$ -szer következik be, a binomiális eloszlás képlete szerint

$$\binom{x+n-1}{n-1} p^{n-1}(1-p)^{(x+n-1)-(n-1)} \;=\; \binom{x+n-1}{n-1} p^{n-1}(1-p)^{x}$$

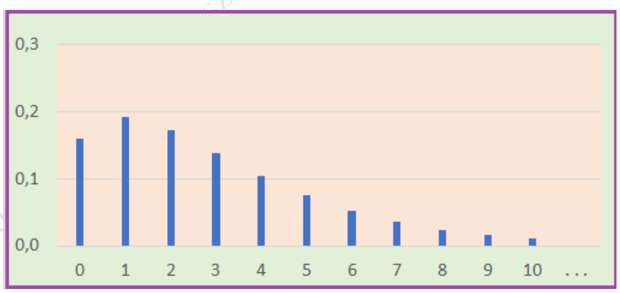
Mivel az $(x + n)$ -ik kísérletnél az A esemény bekövetkezik, a keresett valószínűséget úgy kapjuk meg, hogy ezt a valószínűséget még megszorozzuk p -vel:

$$p(x) \;=\; \mathrm{P}\,(X = x) \;=\; \binom{x+n-1}{n-1} p^{n-1}(1-p)^{x}\; p \;=\; \binom{x+n-1}{n-1} p^{n}(1-p)^{x}$$

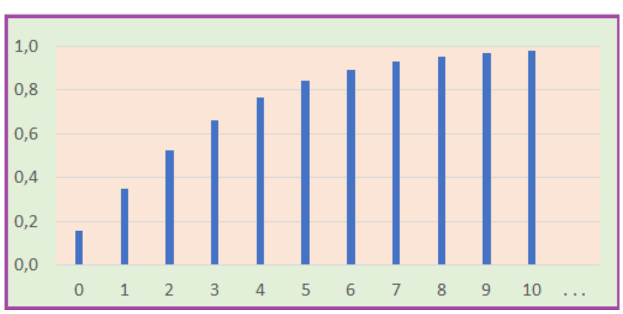
Excel-függvények:

$p(x)$ = `NEGBINOM.DIST(x;n;p;FALSE)` = `NEGBIN.ELOSZLÁS(x;n;p;HAMIS)`

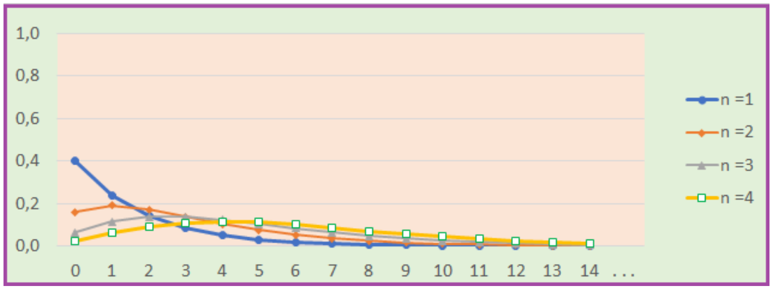
$p(x)$ = `NEGBINOMM.DIST(x;n;p;TRUE)` = `NEGBIN.ELOSZLÁS(x;n;p;IGAZ)`



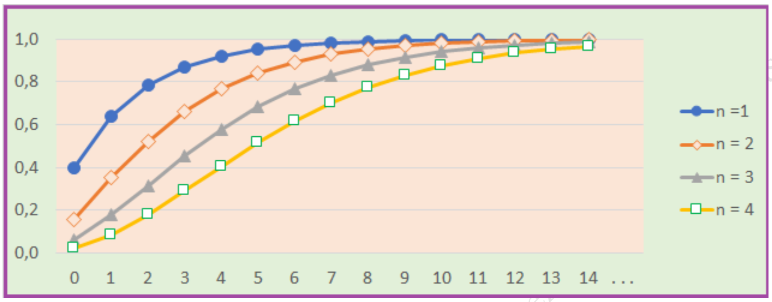
29. ábra. *Negatív binomiális (pesszimista) eloszlás súlyfüggvénye;* $n = 2;\; p = 0.4$



30. ábra. *Negatív binomiális (pesszimista) eloszlás eloszlásfüggvénye;* $n = 2;\; p = 0.4$



31. ábra. *Negatív binomiális (pesszimista) eloszlások súlyfüggvényei;* $n = 1,\ 2,\ 3,\ 4;\ p = 0.4$



32. ábra. *Negatív binomiális (pesszimista) eloszlások eloszlásfüggvényei;* $n = 1,\ 2,\ 3,\ 4;\ p = 0.4$

Megjegyzés: Vegyük észre, hogy az (optimista, pesszimista) negatív binomiális eloszlások az (optimista, pesszimista) geometriai eloszlások általánosításai. $n = 1$ esetén az (optimista, pesszimista) negatív binomiális eloszlás a megfelelő geometriai eloszlásba egyszerűsödik.

1.6. Poisson eloszlás

1.6.1. Poisson eloszlás egydimenzióban

Súlyfüggvény:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad \text{ha} \quad x = 0, 1, 2, \ldots$$

Mikor használjuk: Sok független, külön-külön kis valószínűségű esemény kapcsán az alábbi valószínűségi változó ilyen eloszlást követ:

$$X = \text{ahány esemény bekövetkezik}$$

Paraméter jelentése:

$$\lambda = \text{ahány esemény } \textit{általában, átlagosan} \text{ bekövetkezik}$$

azaz

$$\lambda = \text{az eloszlás várható értéke (ezt a fogalmat a következő fejezetben tanuljuk)}$$

azaz

$$\lambda = \text{az események valószínűségeinek összege}$$

Megjegyzés: Ha az események száma n, és az események egyforma valószínűségűek, és ez a közös érték p, akkor

$$\lambda = np$$

Excel-függvények:

$$p(x) = \texttt{POISSON(x;}\lambda\texttt{;FALSE)} = \texttt{POISSON(x;}\lambda\texttt{;HAMIS)}$$
$$F(x) = \texttt{POISSON(x;n;p;TRUE)} = \texttt{POISSON(x;n;p;IGAZ)}$$

Megjegyzés: Be lehet látni, hogy ha λ egy fix pozitív szám, akkor akármilyen x pozitív egész szám esetén az

$$n \to \infty\,, \qquad p \to 0\,, \qquad n \cdot p \to \lambda$$

feltételek mellett

$$\lim \binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

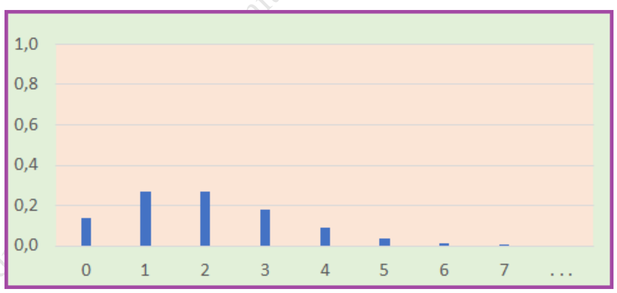
Ez azt jelenti, hogy ha n nagy és p pici úgy, hogy szorzatuk körülbelül λ, akkor a binomiális eloszlás x -ik tagja közelítőleg egyenlő a Poisson eloszlás x -ik tagjával:

$$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda}$$

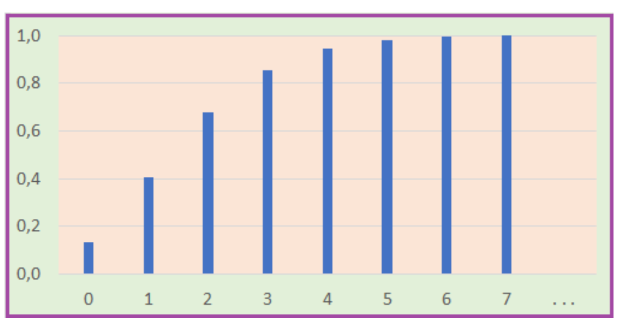
Igazából ez a tény támasztja alá azt, amit fentebb mondtunk: sok független, külön-külön kis valószínűségű esemény kapcsán az

$$X = \text{ahány esemény bekövetkezik}$$

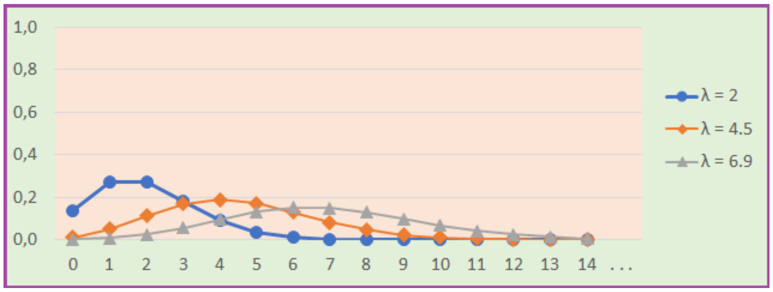
valószínűségi változó eloszlását binomiális eloszlás helyett vehetjük Poisson eloszlásnak.



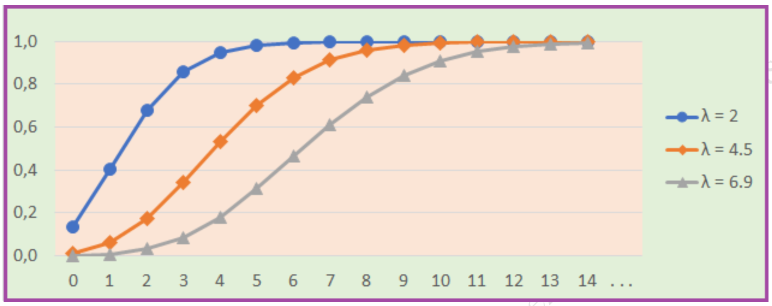
33. ábra. *Poisson eloszlás súlyfüggvénye;* $\lambda = 2$



34. ábra. *Poisson eloszlás eloszlásfüggvénye;* $\lambda = 2$



35. ábra. *Poisson eloszlások súlyfüggvényei;* $\lambda = 2,\ 4.5, 6.9$



36. ábra. *Poisson eloszlások eloszlásfüggvényei;* $\lambda = 2,\ 4.5, 6.9$

1. feladat: Hány hal lesz az *ÖREG* halász hálójában? Történt egyszer, hogy egy nagy tó partján álldogálva nézegettem, ahogy a sok apró halacska egymással és az öreg, alig-alig látó halásszal mit sem törődve össze-vissza úszkált. A halász éppen a hálóját készült kiemelni, amikor kedvesen így szólt hozzám: "Fiatalember! Ha eltalálod, hogy hány hal lesz a hálóban, meghívlak vacsorára". Szeretem a sült halat, és éhes is voltam. Csak annyit kérdeztem az öreg halásztól, hogy milyen gyakran szokott üres lenni a háló. Ő erre azt felelte:

Sok éves tapasztalatom szerint mondhatom neked, hogy az eseteknek a 6 % -ában üres a háló!

Kicsit gondolkodtam, számolgattam, és 2 halra tippeltem. Hogyan gondolkodtam, miért éppen 2 halra tippeltem?

Megoldás: Nyilván azt kellett kigondolnom, hogy a hálóban hány hal a legvalószínűbb. Így okoskodtam: A hálóban lévő halak száma – mint valószínűségi változó – Poisson eloszlást követ, hiszen

- a tóban *sok* a hal, és
- a sok hal mindegyike egymással mit sem törődve össze-vissza úszkál, ezért *egymástól függetlenül* kerülnek vagy nem kerülnek a hálóba, továbbá
- minden hal esetén az az esemény, hogy ő a hálóba kerül (ez az esemmény a hal számára felettébb szomorú), *kis valószínűségű*, hiszen ez a valószínűség a háló (akármilyen) méretének és a tó méretének a hányadosa, ami nyilván pici.

Mivel – sok éves tapasztalata alapján – a halász elárulta, hogy az eseteknek a 6 % -ában üres a háló, a 0 darab hal valószínűségét 0.06 -nak vettem, vagyis gondolatban felállatottam a

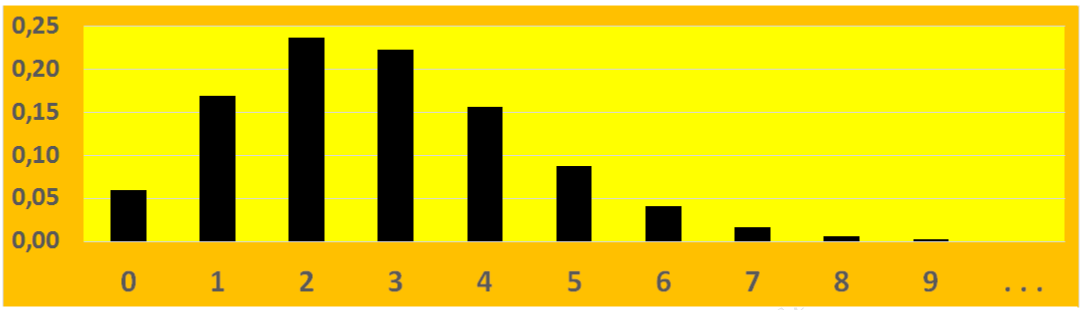
$$\frac{\lambda^0}{0!}\mathrm{e}^{-\lambda} \;=\; 0.06 \qquad \text{azaz} \qquad \mathrm{e}^{-\lambda} \;=\; 0.06$$

egyenletet, aminek megoldása $\lambda \;=\; -\ln(0.06) \;=\; 2.81$.

A 2.81 paraméterű Poisson eloszlás – melyet én akkor még fejből is tudtam – így fest:

x	0	1	2	3	4	5	6	7	8	9	...
$p(x)$	0.06	0.17	0.24	0.22	0.16	0.09	0.04	0.02	0.01	0.00	...

Táblázat: 2.81 *paraméterű Poisson eloszlás*



37. ábra. *Poisson eloszlás* $\lambda = 2.81$ *paraméterrel*

A táblázatból is és az ábrából is jól látható (amit a Poisson eloszlás móduszára vonatkozó ismereteinkből a táblázat és az ábra nélkül is tudunk), hogy az eloszlás módusza 2 -vel egyenlő, vagyis 2 hal a legvalószínűbb. Ezért 2 halra tippeltem – ez volt a legjobb tipp.

Természetesen a 3 hal, vagy az 1 hal, vagy a 4 hal sem rossz tipp, bár ezek a tippek kicsit kisebb valószínűséggel vezetnek sikerre. Viszont – például – a 9 , 10 , 11 vagy több hal valószínűségei – mint számolással ellenőrizhető – még együttesen is 0.001 -nél kevesebbet tesznek ki, ezért az ilyen tippek nem kecsegtetnek a siker reményével.

2. feladat: Hány hal lesz az *IFJÚ* halász hálójában? Érdekes utána járni, hogy hogyan kellett volna gondolkodnom, ha a halász fiatal lett volna, és – sok éves tapasztalat hiányában – így felelt volna:

"A legutóbbi 100 merítés során kifigyeltem, hogy pontosan 6 -szor volt üres a háló!"

Igaz, hogy 6 osztva 100 -zal, megfelel az előző történetben mondott 6 % -nak, de mégis mást jelent. Az öreg halász által mondott 6 % -ot – mivel sok éves tapasztalatra épült – a 0 darab hal valószínűségének tekinthettük. Az ifjú halász által mondott " 100 merítés során 6 -szor volt üres a háló" kijelentés csak egyetlen (100 merítésből álló) megfigyelés!

Abból a tényből, hogy egy alkalommal 100 merítésből 6 -szor üres a háló, nem vehetjük a 0 darab hal valószínűségét 6 % -nak. Az, hogy 100 merítésből 6 -szor üres a háló, megtörténhet még akkor is, ha a 0 darab hal valószínűsége eltér 0.06 -tól.

A 0 darab hal ismeretlen valószínűségét jelöljük p -vel. A p segítségével könnyen felírhatjuk annak valószínűségét, hogy 100 merítésből pontosan 6-szor üres a háló:

$$\text{P(} 100 \text{ merítésből } 6 \text{ -szor üres a háló)} = \binom{100}{6} p^6 (1-p)^{100-6}$$

A jobb oldalon álló képlet p függvényeként egyszerűen vizsgálható:

p	**0.015**	0.030	0.045	**0.060**	0.075	0.090	0.105	0.120	0.135	**0.150**
$\binom{100}{6}p^6(1-p)^{100-6}$	**0.003**	0.050	0.131	**0.166**	0.139	0.089	0.047	0.022	0.009	**0.003**

Táblázat: *A p-től való függés szemléltetése*

A táblázatból is kitűnik, és az analízis eszközeivel is könnyen belátható, hogy a

$$\binom{100}{6}p^6(1-p)^{100-6}$$

kifejezés értéke

- $p < 0.060$ esetén növekszik
- $p = 0.060$ esetén a maximális érték 0.166 -dal egyenlő
- $0.060 < p$ esetén csökken

A táblázatban vastagítással jelezzük a tényt, hogy $p =$**0.015** -nél is és $p =$**0.150** -nél is a kifejezés értéke csak **0.003** . A 0.003 érték a maximális 0.166 értékhez képest elhanyagolhatóan kicsi. Ezért a p -vel jelölt ismeretlen valószínűséget 0.015 és 0.150 közöttinek feltételezzük. Leginkább a $p = 0.060$ érték tűnik jogosnak, de a 0.015 és 0.150 közötti p értékeket sem zárjuk ki. A $p < 0.015$ és a $p > 0.150$ lehetőségeket viszont már elvetjük.

Ez a döntés kétségtelenül egy szubjektív döntés, melyhez hasonlókat a mindennapi életben nap mint nap megteszünk: például szívesebben ülünk egy olyan autóba, amiben van légzsák, mint egy olyanba, amelyikben nincs. A személyi sérülés valószínűsége egy esetleges balesetnél a légzsák esetén lényegesen kisebb, mintha nem lenne légzsák.

Az előző feladat megoldásában elmagyaráztuk, hogy a hálóban lévő halak száma – mint valószínűségi változó – Poisson eloszlást követ. Az eloszlás paraméterét a

$$\frac{\lambda^0}{0!}\mathrm{e}^{-\lambda} = p \qquad \text{azaz} \qquad \mathrm{e}^{-\lambda} = p$$

egyenletből a $\lambda = -\ln(p)$ formula alapján kapjuk meg:

p	0.015	0.030	0.045	0.060	0.075	0.090	0.105	0.120	0.135	0.150
$\lambda = -\ln(p)$	**4.20**	3.51	3.10	2.81	2.59	2.41	2.25	2.12	2.00	**1.90**

Táblázat: *p és λ kapcsolata*

E táblázatban vastagítással kiemeltük a $p = 0.015$ és a $p = 0.150$ szélső esetekhez tartozó λ értékeket:

- $p = 0.015$ esetén $\lambda =$ **4.20**
- $p = 0.150$ esetén $\lambda =$ **1.90**

Ezekhez a paraméter értékekhez tartozó Poisson eloszlások táblázatai így festenek:

x	0	1	2	3	**4**	5	6	7	8	9	...
$p(x)$	0.015	0.063	0.132	0.185	**0.194**	0.163	0.114	0.069	0.036	0.017	...

Táblázat: *Poisson eloszlás* $\lambda = \mathbf{4.20}$ *paraméterrel*

x	0	**1**	2	3	4	5	6	7	8	9	...
$p(x)$	0.150	**0.285**	0.270	0.171	0.081	0.031	0.010	0.003	0.001	0.000	...

Táblázat: *Poisson eloszlás* $\lambda = \mathbf{1.90}$ *paraméterrel*

A táblázatokból látható (amit a Poisson eloszlás móduszára vonatkozó ismereteinkből a táblázatok nélkül is tudunk):

- $\lambda = 4.20$ (vagyis $p = 0.015$) esetén a Poisson eloszlás legvalószínűbb tagja, azaz módusza 4 -gyel egyenlő
- $\lambda = 1.90$ (vagyis $p = 0.150$) esetén a Poisson eloszlás legvalószínűbb tagja, azaz módusza 1 -gyel egyenlő

A $0.015 < p < 0.150$ valószínűségeknek megfelelő $1.90 < \lambda < 4.20$ paraméter értékek melletti Poisson eloszlások móduszai nyilván 1 és 4 közé esnek.

Ezért – összegezve a gondolatokat – ezt kapjuk: *lehetséges, hogy akár* 1 *vagy* 2 *vagy* 3 *vagy* 4 *hal a legvalószínűbb, de a* 2 *hal tűnik a legjobb tippnek*, hiszen – fentebb – láttuk, hogy a

$$\text{P}(\ 100 \text{ merítésből } 6 \text{ -szor üres a háló}) = \binom{100}{6} p^6 (1-p)^{100-6}$$

valószínűség $p = 0.060$ esetén a legnagyobb, ami $\lambda = 2.81$ -nek felel meg, és $\lambda = 2.81$ esetén a Poisson eloszlás legvalószínűbb tagja, vagyis a módusza 2 -vel egyenlő.

Ez az eredmény összhangban van az előző feladat megoldásával, de ugyanakkor érzékelteti, hogy nem mindegy, hogy valaki a sok éves tapasztalata alapján képes egy valószínűséget megmondani, vagy csak 100 kísérlet eredménye alapján a relatív gyakorisággal közelíti azt.

3. feladat: Kullancsok a futóversenyen. A "Kocogj velünk!" mozgalom keretében tavaly futóversenyt rendeztek a Duna-kanyarban. A pályát sajnos kullanccsal fertőzött területen át vezették. Kiderült, hogy a versenyzők közül 300 -an találtak magukban egy, 75 -en pedig két kullancsot. Ennek alapján becsüljük meg, hogy körülbelül hányan indultak a versenyen!

Megoldás első része:

Ha azt vizsgáljuk, hogy egy kiszemelt versenyzőben, mondjuk a "Futó Botond" nevűben, hány kullancs lesz a verseny után, akkor egy valószínűségi változót kapunk. Mivel a sok kullancs mindegyike – a többitől függetlenül – kis valószínűséggel kerül ebbe a versenyzőbe, ez a valószínűségi változó Poisson eloszlást követ valamilyen λ paraméterrel. Ezért – a Poisson eloszlás képlete szerint – annak a valószínűsége, hogy 1 kullancs kerül Futó Botondba, $\lambda e^{-\lambda}$. A 2 kullancs valószínűsége pedig $\frac{\lambda^2}{2} e^{-\lambda}$. Ha a versenyzők számát N -nel jelöljük, és a valószínűségeket relatív gyakoriságokkal helyettesítjük, akkor az alábbi egyenleteket állíthatjuk fel:

$$\frac{300}{N} = \lambda e^{-\lambda} \qquad\qquad \frac{75}{N} = \frac{\lambda^2}{2} e^{-\lambda}$$

Ezt az egyenletrendszert könnyű megoldani. Először elosztjuk a második egyenletet az elsővel. A kapott új egyenletben a baloldal törtben N -nel, a jobboldaliban λ -val és $e^{-\lambda}$ -val egyszerűsítve ezt kapjuk:

$$\frac{75}{300} = \frac{\lambda}{2} \qquad \text{vagyis} \qquad \lambda = \frac{150}{300} = 0.5$$

Ezek után az első egyenletből N -re ez jön ki:

$$N = \frac{300}{\lambda e^{-\lambda}} = \frac{300}{0.5\, e^{-0.5}} = 989.2$$

A versenyzők száma természetesen egész szám. Az, hogy itt N -re nem egész jött ki, annak a következménye, hogy a kiszemelt versenyzőben az 1 , illetve 2 kullancs valószínűségét relatív gyakoriságokkal közelítettük. Ezért a feladatban feltett kérdésre kézenfekvő a közelítő válasz: **Körülbelül 1000 versenyző volt a versenyen**.

Megoldás második része - *(Extra tananyag)*:

Az emberben óhatatlanul felmerül a kérdés: vajon mennyire körülbelül a "körülbelül 1000 "? A probléma elemzése érdekében kiszámoljuk most, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 300 *versenyzőben lesz* 1 *kullancs*

Ha minden egyes versenyzővel kapcsolatban az 1 kullancs valószínűségét $\frac{300}{1000} = 0.3$ -nak vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűségre az 1000 -ed rendű, 0.3 paraméterű binomiális eloszlás szerint Excellel a

```
BINOM.DIST( 300 ; 1000 ; 0.3 ; FALSE )
```

képlet adódik. A valószínűség numerikus értéke 0.027. Azt is kiszámolhatjuk, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 75 *versenyzőben lesz* 2 *kullancs*

Ha minden egyes versenyzővel kapcsolatban a 2 kullancs valószínűségét $\frac{75}{1000} = 0.075$ -nek vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűségre az 1000 -ed rendű, 0.075 paraméterű binomiális eloszlás szerint Excellel a

```
BINOM.DIST( 75 ; 1000 ; 0.075 ; FALSE )
```

képlet adódik. A valószínűség numerikus értéke 0.048. Azt is kiszámoljuk, hogy mi a valószínűsége annak, hogy 1000 versenyző esetén

pontosan 300 *versenyzőben lesz* 1 *kullancs, és*
pontosan 75 *versenyzőben lesz* 2 *kullancs*

azaz

pontosan 300 *versenyzőben lesz* 1 *kullancs, és*
pontosan 75 *versenyzőben lesz* 2 *kullancs, és*
pontosan $1000 - 300 - 75$ *versenyzőben lesz 0 kullancs vagy 2-nél több kullancs*

Ha – úgyanúgy, mint fentebb – minden egyes versenyzővel kapcsolatban az 1 kullancs valószínűségét $\frac{300}{1000} = 0.3$ -nak, a 2 kullancs valószínűségét $\frac{75}{1000} = 0.075$ -nek vesszük, és feltételezzük, hogy a versenyzőkben a kullancsok száma egymástól független, akkor a keresett valószínűség az 1000 -ed rendű, $(0.3; 0.075)$ paraméterű polinomiális eloszlással adódik:

$$\frac{(1000)!}{(300)!\,(75)!\,(1000-300-75)!}\ (0.3)^{300}\ (0.075)^{75}\ (1-0.3-0.075)^{1000-300-75}$$

A polinomiális eloszlás nincs beépítve az Excelbe, és a polinomiális eloszlás képletében szereplő faktoriálisokat sem tudja az Excel kiszámolni. Viszont a polinomiális eloszlást kétdimenziós normális eloszlással közelíthetjük, aminek eredményeképpen a valószínűségre (hat tizedesjegyre kerekítve) 0.001269 jön ki. Azon, hogy egy nagyon kis valószínűség érték jött ki, nem szabad meglepődni: érezhetően kicsi annak esélye, hogy 1000 versenyző közül

pontosan 300 *versenyzőben lesz* 1 *kullancs, és*
pontosan 75 *versenyzőben lesz* 2 *kullancs*

Az embernek eszébe jut, hogy elvileg olyan szélsőséges helyzet is lehetne, hogy – mondjuk – csak 375 versenyző indul a versenyen, és a sors úgy hozza, hogy

300 *versenyzőben lesz* 1 *kullancs, és*
75 *versenyzőben lesz* 2 *kullancs*

Avagy 5000 versenyző esetén sem zárható ki, hogy

300 *versenyzőben lesz* 1 *kullancs, és*
75 *versenyzőben lesz* 2 *kullancs*
4625 *versenyzőben pedig* 0 *kullancs vagy 2-nél több kullancs*

Ezért elbizonytalanodunk, hogy az egyenletrendszerből adódó 898 , avagy 1000 , amit kerekítéssel kaptunk, vajon tényleg jó közelítése a versenyzők számának? A bizonytalanság eloszlatása céljából elemezzük most a problémát, és megnézzük, hogy különböző N és λ értékek esetén mi a valószínűsége annak, hogy

pontosan 300 *versenyzőben lesz* 1 *kullancs, és*
pontosan 75 *versenyzőben lesz* 2 *kullancs*

A valószínűségeket az Excel segítségével számoltunk ki, és táblázatba rendezve adjuk meg. Mivel a valószínűségek értéke nagyon pici, a táblázatban a valószínűségek értékeinek millószorosát (egészekre kerekítve) írtuk be:

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	λ
750	0	0	0	0	0	0	10	1	0	
800	0	0	0	0	0	35	39	0	0	
850	0	0	0	0	2	374	10	0	0	
900	0	0	0	0	96	388	2	0	0	
950	0	0	0	0	833	55	2	0	0	
1000	0	0	0	1	1 294	1	0	0	0	
1050	0	0	0	16	448	0	0	0	0	
1100	0	0	0	125	41	0	0	0	0	
1150	0	0	0	295	1	0	0	0	0	
1200	0	0	0	242	0	0	0	0	0	
1250	0	0	0	77	0	0	0	0	0	
N										

Táblázat: *Valószínűségek millószorosai*

A táblázatból kitűnik, hogy – arányaikat tekintve – "kicsi" és "kicsi" között is nagy a különbség! Annak a valószínűsége, hogy

pontosan 300 *versenyzőben lesz* 1 *kullancs, és*
pontosan 75 *versenyzőben lesz* 2 *kullancs*

a táblázatban látható értékek közül 1000 versenyző esetén a legnagyobb, és már 950 vagy 1050 versenyző esetén is jóval kisebb, de 900 vagy annál kevesebb, illetve 1100 vagy annál több versenyző esetén sokkal-sokkal kisebb.

Ezért józan, elfogadható következtetésnek tűnik: körülbelül 1000 versenyző indult a versenyen, ahol a "körülbelül" azt jelenti, hogy 950 -nél kevesebb vagy 1050 -nél több versenyző gyakorlatilag kizárt.

1.6.2. Poisson eloszlás kétdimenzióban

Súlyfüggvény:

$$p(x,y) \;=\; \frac{\lambda_1^x}{x!}\,e^{-\lambda_1}\;\cdot\;\frac{\lambda_2^y}{y!}\,e^{-\lambda_2} \qquad \text{ha} \quad x,y=0,1,2,\ldots$$

Mikor használjuk: Amikor egy olyan kétdimenziós valószínűségi változóval van dolgunk, melynek koordinátái függetlenek, és külön-külön Poisson eloszlást követnek valamilyen paraméterekkel.

Paraméterek jelentése:

$\lambda_1 \;=\;$ az első koordináta "átlagos" értéke (pontosabban: várható értéke)

$\lambda_2 \;=\;$ a második koordináta "átlagos" értéke (pontosabban: várható értéke)

A várható érték fogalmát a következő fejezetben tanuljuk.

1.7. A csaló vándor és a Bölcs Király

Feladat: Még ifjú voltam és bohó, amikor meseország királya megengedte nekem, hogy a Varázs-tóból aranyhalakat fogjak. Engedélyt adott arra, hogy a hálómat 7 -szer belemerítsem a tóba. Nagy izgalommal mentem a tóhoz. Azóta is előttem van a látvány, ahogy a sok apró aranyhalacska egymással és hálómmal mit sem törődve össze-vissza úszkált a hatalmas tó minden zugában. Senki sem volt a közelben. Ezért csalárd módon 7 -nél jóval többször merítettem be a hálómat. Hazafelé menet találkoztam a királlyal. Megkérdezte tőlem:

– Mondd, hányszor volt hálódban pontosan 3 halacska?

– Bizony hatszor, felség – feleltem az igazságnak megfelelően, mert attól féltem, hogy ha a csalás után még hazudok is, akkor a fejemre szakad az ég,

– Nem tévedsz? – kérdezte kicsit gyanakodva.

– Nem - feleltem határozottan. Kis gondolkodás után a király így szólt:

– Fiam, nem bízom abban, hogy egy 0.001 -nél kisebb valószínűségű esemény bekövetkezzen. Sokkal inkább arra gyanakszom, hogy te becsaptál, és kb. 25 -ször vagy még ennél is többször merítettél a tóból. Ezért az aranyhalakat dobd vissza a tóba, és takarodj országomból!

Hogyan gondolkodott a Bölcs Király?

Megoldás: A Király egyetemista korában rendesen átgondolta a valószínűségszámítás anyagot. Jól ismerte a véletlen törvényeit. Így gondolkodott:

Egy-egy merítésnél a kifogott halak száma Poisson eloszlást követ, hiszen a sok hal mindegyike a többitől függetlenül kis valószínűséggel kerül a hálóba. (A függetlenség abból következik, hogy a halak egymással mit sem törődve úszkáltak, a valószínűség pedig kb. a háló területének és a tó területének a hányadosa.) Ezért annak az eseménynek, hogy pontosan 3 hal kerül a hálóba, p valószínűsége: a

$$p \;=\; \frac{\lambda^3}{3!}\mathrm{e}^{-\lambda}$$

ahol λ valamilyen általunk nem ismert pozitív paraméter. λ szerinti deriválással könnyű meghatározni, hogy ennek a kifejezésnek a maximuma $\lambda=3$ esetén adódik, és a maximum értéke

$$\frac{3^3}{3!}\mathrm{e}^{-3} \;=\; 0.224$$

Ezért

$$p \;\leq\; 0.224$$

Ha én nem csalok, és 7 -szer merítek a hálómmal, akkor azoknak a merítéseknek a száma, amikor pontosan 3 hal van a hálóban, binomiális eloszlást követő valószínűségi változó $n = 7$ és p paraméterekkel. (A binomiális eloszlást az indokolja, hogy az egyes merítések eredménye nem befolyásolja a többi merítés eredményét, és mind a 7 esetben p valószínűséggel következik be, hogy pontosan 3 hal van a hálóban.) Tehát ha nem csaltam volna, akkor annak a valószínűsége, hogy 6 -szor fogok 3 halat, ennyi:

$$7 \cdot p^6 \cdot (1-p)^1$$

p szerinti deriválással könnyű meggyőződni, hogy ez a kifejezés 0 és $\frac{6}{7}$ közötti p értékekre monoton növekedő. Ezért a $p \leq 0.224$ egyenlőtlenségből következik, hogy

$$7 \cdot p^6 \cdot (1-p)^1 \ \leq 7 \cdot 0.224^6 \cdot (1-0.224)^1 \ = \ 0.000686 \ < \ 0.001$$

Ezt a kis valószínűséget szembeállítva a 7 -ed rendű binomiális eloszlás többi tagjának lényegesen nagyobb értékével, a király elvetette azt a hipotézist, hogy én 7 -szer merítettem a hálómmal.

Ezek után még így gondolkodott:

Ha a merítéseim számát N -nel jelöljük, és az N merítés során 6 -szor volt 3 hal a hálóban, akkor a $\frac{6}{N}$ relatív gyakoriságot körülbelül egyenlőnek vehetjük a 3 hal kifogásának a p valószínűségével:

$$\frac{6}{N} \ \approx \ p$$

vagyis

$$\frac{6}{p} \ \approx \ N$$

Mivel $p \leq 0.224$, az adódik, hogy

$$N \ \approx \ \frac{6}{p} \geq \frac{6}{0.224} \ = \ 26.78$$

vagyis N értéke körülbelül 25 vagy 26 vagy 27 vagy még nagyobb.

1.8. Példa nem normált eloszlásra *(Extra tananyag)*

A gyorsan felejtő diákról szóló példában, melyet a 6. fejezet **Feladatok vizsgálatokról, vizsgákról** című fejezet 2. pontjában tárgyaltunk feltettük, hogy az idő múlásával a diák fárad, tudása hanyatlik, ezért, ha arra sor kerül, az n -ik vizsgán a bukás valószínűsége – az n függvényében – az alábbi (mesterségesen kitalált) matematikai képlet szerint nő :

$$\text{P(az } n\text{-ik vizsga sikertelen | az } n\text{-ik vizsgára sor kerül)} \ = \ \frac{0.6 + \frac{0.4}{n+1}}{0.6 + \frac{0.4}{n}} \quad (\ n = 1, 2, \ldots)$$

A feladat utáni megjegyzésben kiszámoltuk, hogy annak a valószínűsége, hogy a diák éppen az n-ik vizsgán megy át, minden $n = 1, 2, \ldots$ értékre ennyi:

$$\text{P(a diák éppen az } n\text{-ik vizsgán megy át)} \ = \ \frac{0.4}{n(n+1)}$$

Ezért, ha valaki azt a valószínűségi változót vizsgálja, hogy hányadik vizsgán megy át a diák, akkor ennek a valószínűségi változónak az eloszlását az alábbi súlyfüggvény adja meg:

$$p(n) \ = \ \frac{0.4}{n(n+1)} \qquad n = 1, 2, \ldots$$

Ez az eloszlás nem normált, mert

$$\sum_{n=1}^{\infty} p(n) \;=\; \sum_{n=1}^{\infty} \frac{0.4}{n(n+1)} \;=\; \sum_{n=1}^{\infty} \left(\frac{0.4}{n} - \frac{0.4}{n+1} \right) \;=\; 0.4 < 1$$

hiszen a

$$\sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) \;=$$

$$\left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \dots$$

összeg – teleszkopikus tulajdonsága miatt – nyilván 1-gyel egyenlő.

Megjegyzés: 6. fejezet 2. pontjában kiszámoltuk annak a valószínűségét, hogy a diák vizsgája valaha is sikerül. Erre ott 0.4 -et kaptunk. Ugyanezt fejezi ki az a tény, hogy a

$$\sum_{n=1}^{\infty} p(n) \;=$$

összeg értéke 0.4 -gyel egyenlő.

1.9. A nevezetes eloszlások mindegyike normált – bizonyítások *(Extra tananyag)*

Egyenletes eloszlás

$$\sum_{x} p(x) \;=\; n \cdot \frac{1}{n} \;=\; 1$$

Hipergeometrikus eloszlás

$$\sum_{x} p(x) \;=\; \sum_{x=\max(0,n-B)}^{\min(n,A)} \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} \;=\; 1$$

A bizonyításban felhasználtuk az alábbi azonosságot:

$$\sum_{x=\max(0,n-B)}^{\min(n,A)} \binom{A}{x} \binom{B}{n-x} \;=\; \binom{A+B}{n}$$

Ezt az azonosságot így láthatjuk be: Ha egy dobozban A piros és B zöld golyó van, és mi n golyót húzunk visszatevés nélkül, akkor azoknak a kombinációknak a száma, melyekben pontosan x piros golyó van, a $\binom{A}{x}\binom{B}{n-x}$ képlettel adható meg, ezért az összes kombinációk száma

$$\sum_{x=\max(0,n-B)}^{\min(n,A)} \binom{A}{x} \binom{B}{n-x}$$

Ugyanakkor az összes kombinációk száma nyilván $\binom{A+B}{n}$.

Binomiális eloszlás

$$\sum_{x} p(x) \;=\; \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \;=\; (p + (1-p))^n \;=\; 1^n \;=\; 1$$

A bizonyításban felhasználtuk az alábbi – "binomiális tétel" néven ismert – azonosságot:

$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} \;=\; (a+b)^n$$

Esetünkben $a = p$, $b = 1-p$ volt.

Polinomiális eloszlás

$$\sum_{(x,y)} p(x,y) \;=$$

$$= \sum_{(x,y)} \frac{n!}{x!\; y!\; (n-x-y)!} \;(p_1)^x\;(p_2)^y\;(1-p_1-p_2)^{n-x-y} \;=\; (\,p_1+p_2+(1-p_1-p_2)\,)^n \;=\; 1^n \;=\; 1$$

A bizonyításban felhasználtuk az alábbi – "polinomiális tétel" néven ismert – azonosságot:

$$= \sum_{(x,y)} \frac{n!}{x!\; y!\; (n-x-y)!} \;a^x\; b^y\; c^{n-x-y} \;=(\,a+b+c\,)^n$$

Esetünkben $a = p_1$, $b = p_2$, $c = 1-p_1-p_2$ volt.

Geometriai eloszlás – pesszimista

$$\sum_x p(x) \;=\; \sum_{x=0}^{\infty}(1-p)^x p \;=\; \frac{p}{1-(1-p)} \;=\; \frac{p}{p} \;=\; 1$$

Felhasználtuk a végtelen mértani sorokra vonatkozó "első tag osztva egy mínusz kvóciens" összegzési szabályt:

$$\sum_{x=0}^{n} q^x a \;=\; \frac{a}{1-q} \;=\; \frac{p}{p} \;=\; 1$$

Geometriai eloszlás – optimista

$$\sum_x p(x) \;=\; \sum_{x=1}^{\infty}(1-p)^{x-1} p \;=\; \frac{p}{1-(1-p)} \;=\; \frac{p}{p} \;=\; 1$$

Itt is felhasználtuk a végtelen mértani sorokra vonatkozó "első tag osztva egy mínusz kvóciens" összegzési szabályt.

$$\sum_{x=0}^{n} q^x a \;=\; \frac{a}{1-q} = \frac{p}{p} \;=\; 1$$

Poisson eloszlás

$$\sum_x p(x) \;=\; \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} \;=\; e^{-\lambda}\; \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \;=\; e^{-\lambda}\; e^{\lambda} \;=\; 1$$

Felhasználtuk az exponenciális eloszlás jól ismert Taylor-sorát:

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \;=\; e^{\lambda}$$

Poisson eloszlás kétdimenzióban

$$\sum_{(x,y)} p(x,y) \;=\; \sum_{(x,y)} \left(\frac{\lambda_1^x}{x!}\, e^{-\lambda_1} \cdot \frac{\lambda_2^y}{y!}\, e^{-\lambda_2} \right) \;=\; \left(\sum_x \; \frac{\lambda_1^x}{x!}\, e^{-\lambda_1} \right) \cdot \left(\sum_y \; \frac{\lambda_2^y}{y!}\, e^{-\lambda_2} \right) \;=\; 1 \cdot 1 \;=\; 1$$

2. Módusz megkeresése

2.1. Előkészületek *(Extra tananyag)*

Könnyű megérteni az alábbi három triviális tényt, ha az Olvasó papírt és ceruzát ragad, és – például – pálcikákkal szemléleteti a szóban forgó eloszlásokat.

1. tény: Ha egy eloszlás $p(k)$ súlyfüggvényére igaz, hogy

- $p(k-1) < p(k)$ ha $k \le k_0$ (a súlyfüggvény k_0 -ig növekszik)
- $p(k-1) > p(k)$ ha $k > k_0$ (utána csökken)

akkor ez azt jelenti, hogy

- a $p(0), p(1), \ldots, p(k_0)$ sorozat szigorúan monoton növekedő
- a $p(k_0), p(k_0+1), \ldots$ sorozat szigorúan monoton csökkenő

ezért a súlyfüggvény a maximális értékét a k_0 helyen veszi fel,
vagyis az eloszlás módusza a k_0 szám.

2. tény: Ha egy eloszlás $p(k)$ súlyfüggvényére igaz, hogy

- $p(k-1) < p(k)$ ha $k < k_0$ (a súlyfüggvény (k_0-1) -ig növekszik)
- $p(k-1) = p(k)$ ha $k = k_0$ ((k_0-1)-nél és k_0 -nál egyforma az értéke)
- $p(k-1) > p(k)$ ha $k > k_0$ (utána csökken)

akkor ez azt jelenti, hogy

- a $p(0), p(1), \ldots, p(k_0-1)$ sorozat szigorúan monoton növekedő
- $p(k_0-1) = p(k_0)$
- a $p(k_0), p(k_0+1), \ldots$ sorozat szigorúan monoton csökkenő

ezért a súlyfüggvény a maximális értékét a (k_0-1) és a k_0 helyeken veszi fel,
vagyis az eloszlásnak két módusza van: (k_0-1) és k_0 .

3. tény: Ha $\lfloor x_0 \rfloor$ -lal jelöljük az x_0 valós szám egész részét, vagyis azt az egész számot, amit x_0 -ból kapunk, ha lefelé kerekíjük, és k egész szám, akkor

- a $k \le x_0$ egyenlőtlenség ekvivalens a $k \le \lfloor x_0 \rfloor$ egyenlőtlenséggel
- a $k > x_0$ egyenlőtlenség ekvivalens a $k > \lfloor x_0 \rfloor$ egyenlőtlenséggel

E három tény együtt adja a következő módszert.

2.2. Módszer a módusz képletének meghatározására

Egyetlen módusz esete: Ha egy eloszlás $p(k)$ súlyfüggvényére és egy x_0 valós számra igaz, hogy

- $p(k-1) < p(k)$ minden $k \le x_0$ -ra, és
- $p(k-1) > p(k)$ minden $k > x_0$ -ra,

akkor az eloszlás módusza az $\lfloor x_0 \rfloor$ szám.

Két módusz esete: Ha egy eloszlás $p(k)$ súlyfüggvényére és egy x_0 egész számra igaz, hogy

- $p(k-1) < p(k)$ ha $k < x_0$ -ra
- $p(k-1) = p(k)$ ha $k = x_0$ -ra
- $p(k-1) > p(k)$ ha $k > x_0$ -ra

akkor az eloszlásnak két módusza van: $x_0 - 1$ és x_0 .

1. Példa: Binomiális eloszlás módusza. A binomiális eloszlás súlyfüggvénye

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

Ezért a $p(k-1) < p(k)$ egyenlőtlenség most ezt jelenti:

$$\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} < \binom{n}{k} p^k (1-p)^{n-k}$$

Triviális egyszerűsítések és alakítások következnek:

$$\frac{n!}{(k-1)!(n-k+1)!}\, p^{k-1} (1-p)^{n-k+1} \;<\; \frac{n!}{k!(n-k)!}\, p^k (1-p)^{n-k}$$

$$\frac{1-p}{n-k+1} \;<\; \frac{p}{k}$$

$$k\,(1-p) \;<\; (n-k+1)\,p$$

$$k \;-\; k\,p \;<\; n\,p \;-\; k\,p \;+\; p$$

Tehát a $p(k-1) < p(k)$ egyenlőtlenség ekvivalens a

$$k < (n+1)\,p$$

egyenlőtlenséggel. A $p(k-1) > p(k)$ egyenlőtlenség pedig nyilván ekvivalens a

$$k > (n+1)\,p$$

egyenlőséggel. Ez azt jelenti, hogy $(n+1)\,p$ tölti be azt a szerepet, amit a módusz meghatározására mondott módszerben az x_0 töltött be. Ezért ha $(n+1)\,p$ nem egész szám, akkor az eloszlásnak egyetlen módusza a $\lfloor (n+1)\,p \rfloor$ szám. Ha $(n+1)\,p$ egész szám, $p(k-1) = p(k)$ is fenn tud állni $k = (n+1)\,p$ esetén. Ezért ha $(n+1)\,p$ egész szám, akkor az eloszlásnak két módusza van: $(n+1)\,p - 1$ és $(n+1)\,p$.

2. Példa: Poisson eloszlás módusza. A Poisson eloszlás súlyfüggvénye

$$p(k) = \frac{\lambda^k}{k!}\, e^{-\lambda} \qquad (k = 0, 1, 2, \ldots)$$

Ezért a $p(k-1) < p(k)$ egyenlőtlenség most ezt jelenti:

$$\frac{\lambda^{k-1}}{(k-1)!}\, e^{-\lambda} < \frac{\lambda^k}{k!}\, e^{-\lambda}$$

Triviális egyszerűsítések adják, hogy az egyenlőtlenség ekvivalens ezzel:

$$k < \lambda$$

A $p(k-1) > p(k)$ egyenlőtlenség pedig ekvivalens a

$$k > \lambda$$

egyenlőséggel. Ez azt jelenti, hogy λ tölti be azt a szerepet, amit a módusz meghatározására mondott módszerben az x_0 töltött be. Ezért ha λ nem egész szám, akkor az eloszlásnak egyetlen módusza a $\lfloor \lambda \rfloor$ szám. Ha λ egész szám, akkor $p(k-1) = p(k)$ is fenn tud állni $k = \lambda$ esetén. Ezért ha λ egész szám, akkor az eloszlásnak két módusza van: $\lambda - 1$ és λ .

2.3. Nevezetes eloszlások móduszai – formulák

Hipergeometrikus eloszlás

Ha $(n+1)\,\frac{M+1}{N+2}$ egész szám, akkor két módusz van: $(n+1)\,\frac{M+1}{N+2} - 1$ és $(n+1)\,\frac{M+1}{N+2}$

Ha $(n+1)\,\frac{M+1}{N+2}$ nem egész szám, akkor egy módusz van: $\left\lfloor (n+1)\,\frac{M+1}{N+2} \right\rfloor$

Binomiális eloszlás

Ha $(n+1)\,p$ egész, akkor két módusz van: $(n+1)\,p - 1$ és $(n+1)\,p$

Ha $(n+1)\,p$ nem egész, akkor egy módusz van: $\lfloor (n+1)\,p \rfloor$

Indikátor eloszlás

Ha $p = \frac{1}{2}$, akkor két módusz van: 0 és 1

Ha $p < \frac{1}{2}$, akkor egy módusz van: 0

Ha $p > \frac{1}{2}$, akkor egy módusz van: 1

Optimista geometriai eloszlás

módusz: 1

Pesszimista geometriai eloszlás

módusz: 0

Optimista negatív binomiális eloszlás

Ha $\frac{n-1}{p}$ egész szám, akkor két módusz van: $\frac{n-1}{p}$ és $\frac{n-1}{p}+1$

Ha $\frac{n-1}{p}$ nem egész szám, akkor egy módusz van: $\left\lfloor \frac{n-1}{p} \right\rfloor + 1$

Pesszimista negatív binomiális eloszlás

Ha $\frac{n-1}{p}$ egész, akkor két módusz van: $\frac{n-1}{p}-n$ és $\frac{n-1}{p}-n+1$

Ha $\frac{n-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor \frac{n-1}{p} \right\rfloor - n + 1$

Poisson eloszlás

Ha λ egész, akkor két módusz van: $\lambda - 1$ és λ

Ha λ nem egész, akkor egy módusz van: $\lfloor \lambda \rfloor$

3. Szimuláció

3.1. A $[0;1]$ intervallum felosztásának módszere

Vegyünk fel 1 -nél kisebb pozitív számoknak egy véges sok tagból álló, növekedő sorozatát! Íme – például – egy 3 tagból sorozat:

$$0.2\,,\quad 0.6\,,\quad 0.7$$

Most egy táblázatot alkotunk, melynek első sora 0 -val kezdődik, aztán következnek a sorozat tagjai, a második sorba pedig írjunk négy akármilyen számot, például a 10 , 20 , 30 , 40 számokat:

0	0.2	0.6	0.7
10	20	30	40

Táblázat: *Ahogy Excelben készítjük*

Ha egy ilyen táblázatot készítünk Excelben, akkor

angolul: `HLOOKUP( RAND() ; * ; 2 )` magyarul: `VKERES( VÉL() ; * ; 2 )`

utasítás, ahol * helyén a táblázatra való hivatkozás áll, számunkra előnyösen működik. A generált véletlen számot összehasonlítja a táblázat első sorában lévő számokkal, és megkeresi az első sorban azt a számot, amelyik

- vagy egyenlő a generált számmal,
- vagy éppen alatta van a generált értéknek,

és az utasítás értékéül a második sorból veszi azt az elemet, ami az első sorban talált elem alatt van. Tehát

- ha a véletlen érték 0 és 0.2 közé esik, vagy 0 -val egyenlő, akkor a függvény értéke 10 ,
- ha a véletlen érték 0.2 és 0.6 közé esik, vagy 0.2 -del egyenlő, akkor a függvény értéke 20 ,
- ha a véletlen érték 0.6 és 0.7 közé esik, vagy 0.6 -del egyenlő, akkor a függvény értéke 30 ,
- ha a véletlen érték 0.7 és 1 közé esik, vagy 0.7 -del egyenlő, akkor a függvény értéke 40 .

Mivel

- a $[\,0\,;\,0.2\,]$ intervallum hossza 0.2 ,
- a $[\,0.2\,;\,0.6\,]$ intervallum hossza 0.4 ,
- a $[\,0.6\,;\,0.7\,]$ intervallum hossza 0.1 ,
- a $[\,0.7\,;\,1\,]$ intervallum hossza 0.3 ,

a függvény értéke

- 0.2 valószínűséggel 10 ,
- 0.4 valószínűséggel 20 ,
- 0.1 valószínűséggel 30 ,
- 0.3 valószínűséggel 40 .

Tehát a

```
HLOOKUP( RAND() ; * ; 2 )        VKERES( VÉL() ; * ; 2 )
```

Excel függvénnyel generált valószínűségi változó eloszlása

x	10	20	30	40
$p(x)$	0.2	0.4	0.1	0.3

Táblázat: *A generált valószínűségi változó eloszlása*

Könnyű észrevenni, hogy milyen számokat kell az Excel táblázatba írni ahhoz, hogy adott véges diszkrét eloszlást követő valószínűségi változót kapjunk ezzel a módszerrel.

Például a megadott eloszlás legyen a következő:

x	1	2	3	4	5	6
$p(x)$	0.05	0.10	0.15	0.3	0.25	0.15

Táblázat: *Adott eloszlás*

Ekkor az Excel táblázatot így kell elkészíteni:

0.00	0.05	0.15	0.30	0.60	0.85
1	2	3	4	5	6

Táblázat: *Ahogy Excelben csináljuk*

Az Excel táblázat első sorában a 0 után álló számokat összegzéssel kaptuk meg a megadott eloszlás valószínűségeiből:

$$0.05 = 0 + 0.05$$

$$0.15 = 0 + 0.05 + 0.10$$

$$0.30 = 0 + 0.05 + 0.10 + 0.15$$

$$0.60 = 0 + 0.05 + 0.10 + 0.15 + 0.30$$

$$0.85 = 0 + 0.05 + 0.10 + 0.15 + 0.30 + 0.25$$

4. Tömegpont rendszerek súlypontja és tehetetlenségi nyomatéka

Előrebocsátunk két, a fizikából ismert képletet. Tekintsünk egy tömegpont-rendszert a számegyenesen, mely egy véges vagy megszámlálhatóan végtelen sok pontból álló S halmazra koncentrálódik. Ha x eleme S -nek, akkor az x pontban lévő tömeg mennyiségét jelöljük $p(x)$ -szel. A pontrendszer **súlypont**ja – mint ismeretes

$$\frac{\sum_x x\, p(x)}{\sum_x p(x)}$$

A kifejezésben itt – és a későbbiekben is – a szummázás az összes lehetséges x -re értendő. A

$$\sum_x (x-c)^2\, p(x)$$

érték pedig a pontrendszernek **a** c **pontra vonatkozó tehetetlenségi nyomatéka**.

Megjegyzés: Vegyük észre, hogy ha az össztömeg 1-gyel egyenlő, azaz $\sum_x p(x) = 1$, akkor a súlypontot megadó képlet így egyszerűsödik:

$$\sum_x x\, p(x)$$

5. Egydimenziós adatrendszerek

5.1. Átlag

Adatoknak (számoknak) egy sorozatát *adatrendszer*nek hívjuk. Egy adatrendszer valamilyen értelemben vett közepét, az *átlag* fogalmát, mindenki ismeri.

1. Példa: Adatrendszer 5 elemből. A könnyebb követhetőség kedvéért az alábbi adatrendszer csak 5 elemből áll:

adatok	54	55	44	44	60

Táblázat: *Adatok*

A sor végén feltüntetjük az átlagot, ami 51.4 :

						átlag
adatok	54	55	44	44	60	51.4

Táblázat: *Adatok és az átlaguk*

Az adatok átlagát *első momentum*nak is szokás hívni. Egy külön sorban feltüntetjük az adatok négyzeteit is és azok átlagát is:

						átlag
adatok	54	55	44	44	60	51.4
adatok négyzetei	2 916	3 025	1 936	1 936	3 600	2 682.6

Táblázat: *Az adatok négyzetei*

5.2. Második momentum

Az adatok négyzeteinek az átlagát az adatrendszer *második momentum*ának (esetleg 0 *-ra vonatkozó második momentum*ának) hívjuk. A példánkban szereplő adatrendszer második momentuma $2\,682.6$. Az összegét – például Excellel – mátrixok szorzataként is ki lehet számolni. Az alábbi mátrix szorzás, majd osztás az adatok számával (ami itt 5) a második momentumot adja:

$$\boxed{\begin{array}{ccccc} 54 & 55 & 44 & 44 & 60 \end{array}} \cdot \boxed{\begin{array}{c} 54 \\ 55 \\ 44 \\ 44 \\ 60 \end{array}} \cdot \frac{1}{5} = 2\,682.6$$

Ha a négyzetre emelés előtt egy c számot kivonunk az adatrendszer minden tagjából, akkor a c *-re vonatkozó második momentum*ot kapjuk meg. Az alábbi táblázatban $c = 50$, és az 50 -re vonatkozó második momentum 42.6:

						átlag
adatok	54	55	44	44	60	
adatok és c különbségei	4	5	-6	-6	10	
különbségek négyzete	16	25	36	36	100	42.6

Táblázat: *Második momentum az* 50 *-re vonatkozólag*

A $c = 50$ -re vonatkozó (itt $c = 50$) második momentumot – a (0 -ra vonatkozó) "sima" második momentumhoz hasonlóan – a megfelelő sor- és oszlopvektorok szorzásával, és utána a darabszámmal való osztással is megkaphatjuk:

$$\boxed{\begin{array}{ccccc} 4 & 5 & -6 & -6 & 10 \end{array}} \cdot \boxed{\begin{array}{c} 4 \\ 5 \\ -6 \\ -6 \\ 10 \end{array}} \cdot \frac{1}{5} = 42.6$$

A $c = 50$ -re vonatkozó második momentum arról ad felvilágosítást, hogy

- az adatrendszer vajon a c érték köré tömörül-e – ilyenkor a $c = 50$ -re vonatkozó második momentum értéke kisebb , vagy
- a $c = 50$ értéktől jobban szóródik – ilyenkor a c -re vonatkozó második momentum értéke nagyobb

5.3. Variancia, szórás

Speciálisan kiszámolhatjuk az adatoknak az átlagtól, itt most 51.4 -től vett különbségét is:

						átlag
adatok	54	55	44	44	60	51.4
adatok és átlag különbségei	2.6	3.6	-7.4	-7.4	8.6	0

Táblázat: *Különbségek*

A kapott adatrendszert az eredeti adatrendszer *centralizált*jának hívjuk. A centralizált adatok azt mutatják, hogy az eredeti adatok hogyan helyezkednek el az átlagukhoz viszonyítva. Ha a centralizált érték pozitív, akkor az eredeti adat az átlag felett van, ha ha a centralizált érték negatív, akkor az eredeti adat az átlag alatt van. A centralizált érték nagysága pedig az átlagtól való távolsággal egyenlő.

Egy centralizált adatrendszer átlaga nyilván mindig 0 . A centralizált adatrendszer második momentumát az eredeti adatrendszer *varianciá*jának vagy *szórásnégyzeté*nek hívjuk. A mi adatrendszerünk varianciája 40.64 . A variancia négyzetgyökét, aminek numerikus értéke itt 6.37 , *szórás*nak nevezzük.

						átlag	
centralizált adatok	2.6	3.6	-7.4	-7.4	8.6	0	
centralizált adatok négyzetei	6.76	12.96	54.76	54.76	73.96	40.64	6.37
						variancia	szórás

Táblázat: *Variancia és szórás*

Hangsúlyozzuk, hogy a varianciát – például Excellel – mátrixok szorzásával is ki lehet számolni. A centralizált adatokból alkotott sorvektort jobbról szorozva a transzponáltjával, majd osztva az adatok számával (ami itt 5), a varianciát kapjuk:

$$\boxed{\begin{array}{ccccc} 2.6 & 3.6 & -7.4 & -7.4 & 8.6 \end{array}} \cdot \boxed{\begin{array}{c} 2.6 \\ 3.6 \\ -7.4 \\ -7.4 \\ 8.6 \end{array}} \cdot \frac{1}{5} = 40.64$$

A variancia és a szórás arról adnak felvilágosítást, hogy

- az adatrendszer az átlag köré összetömörül-e – ilyenkor a variancia értéke is és a szórás értéke is kisebb, vagy
- jobban szóródik – ilyenkor a variancia értéke is és a szórás értéke is nagyobb

A következő két adatrendszer mindgyikének az átlaga 100, de a variancia és a szórás a második adatrendszer esetében nagyobb, mint az első esetében Ezzel tudjuk numerikusan kifejezni azt a szemléletes tényt, hogy a második adatrendszer jobban szóródik az átlag körül, mint az első.

adatok	99	100	97	99	105	9	3
						variancia	szórás

Táblázat: *A variancia és a szórás – itt kisebb*

adatok	95	100	85	95	125	225	15
						variancia	szórás

Táblázat: *A variancia és a szórás – itt nagyobb*

Megjegyzés: Elég természetes dolognak tűnne, hogy egy adatrendszer szóródását az átlagtól való (mindenféle négyzetre emeléstől mentes) távolságoknak az átlagával, az ún. *átlagos abszolút eltérés*sel jellemezzük:

						átlag	
adatok	54	55	44	44	60	51.4	
adatok és átlag különbségei	2.6	3.6	-7.4	-7.4	8.6	0	(mindig 0)
távolságok az átlagtól	2.6	3.6	7.4	7.4	8.6	5.92	átlagos abszolút eltérés

Táblázat: *Átlagos abszolút eltérés*

Ezzel a természetes fogalommal szépen lehet numerikusan számolni, dolgozni, de – mint az később kiderül – a variancia és a szórás olyan technikai előnyökkel bírnak, amelyek mind az elméletben, mind a gyakorlatban háttérbe szorítják az átlagos abszolút eltérést. Megjegyezzük, hogy egy adatrendszer átlagos abszolút eltérése legfeljebb akkora lehet, mint a szórása.

Vegyük észre, hogy a kiindulásul felvett adatrendszer varianciáját úgy is megkaphatjuk, hogy a második momentumból kivonjuk az első momentum négyzetét:

$$40.64 = 2\,682.6 - (51.4)^2$$

Ez az összefüggés mindig igaz:

$$\text{variancia} = \text{második momentum} - (\text{ első momentum })^2$$

A bizonyítást az Olvasóra bízzuk.

Megjegyzés: Minden olyan kalkulátor, ami tud statisztikai üzemmódban dolgozni, egy-egy gombnyomásra kiadja a betátplált adatrendszer átlagát, varianciáját, szórását. Az Excelben az átlagra, varianciára, szórásra beépített függvények vannak.

Egyszerű algebrai összefüggés a Steiner egyenlőség, és annak következménye a Steiner egyenlőtlenség:

Steiner egyenlőség: Akármilyen c -re igaz, hogy

$$c \text{ -re vonatkozó második momentum} = \text{variancia} + (c \text{ - átlag })^2$$

Steiner egyenlőtlenség: Akármilyen c-re igaz, hogy

$$c \text{ -re vonatkozó második momentum} \geq \text{variancia}$$

$c = \text{átlag}$ esetén az egyenlőség, más c értékekre az egyenlőtlenség teljesül.

5.4. Medián

Itt teszünk említést az a*datrendszer mediánjá*nak fogalmáról.

- Ha az adatokat növekvő sorrendbe rendezzük, akkor – páratlan sok tag esetén – a sorrendben egyértelmű középső tagot nevezzük mediánnak.
- - Ha páros sok tag van, akkor két "szinte középső" tag van. Vannak, akik ezeknek az átlagát nevezik mediánnak
  - Mások mindkettőt mediánnak nevezik.
  - És vannak akik nemcsak ezt a két értéket, hanem minden közöttük lévő számot is mediánnak neveznek.

  Páros sok tag esetén három kissé különböző deiníció jól elfér egymás mellett és a fejünkben is. Mindhárom hozzáállásnak van előnye is, hátránya is.

  - Az első definíció szerint a medián egyértelmű.
  - A második definíció esetén mindkét medián az adathalmaz eleme. Ha valamiért kell, akkor az átlagukat bármikor ki lehet számítani.
  - A harmadik definíció a legbátrabb és sok szempontból a legegészségesebb: a mediánok egy zárt intervallumot alkotnak. Az intervallum belsejében minden elemre vitathatatlanul igaz, hogy felezi az adathalmazt: tőle balra és jobbra az adathalmaznak ugyanannyi eleme van.

Páros sok tag esetén az Excelben a `MEDIAN` függvény az első definíciónak megfelelően működik.

1. Példa: A $\{6, 7, 15, 3, 4\}$ adatrendszer mediánja 6, hiszen az elemeket nagyság szerint sorba rendezve a $3, 4, 6, 7, 15$ sorozatot kapjuk. Ennek a sorozatnak közepén a 6 áll, ő a medián.

2. Példa: A $\{7, 2, 15, 3, 20, 4\}$ adatrendszer mediánja nem egyértelmű, hiszen az elemeket nagyság szerint sorba rendezve a $2, 3, 4, 7, 15, 20$ sorozatot kapjuk. Ennek a sorozatnak közepén a 4 és a 7 áll, ezért – ízléstől függően – mediánnak a

- a 4 és a 7 számtani közepét, az 5.5 -öt, vagy
- a 4 -et is és a 7 -et is, vagy
- bármely 4 és 7 közötti számot (őket is beleértve) szokás venni.

Az Excel az első lehetőséggel dolgozik. Matematikusi szemmel nézve a harmadik lehetőség a legkorrektebb.

6. Valószínűségi változók és eloszlások várható értéke, varianciája, szórása

6.1. Várható érték

Tekintsünk egy diszkrét X valószínűségi változót. X lehetséges értékeinek halmazát jelöljük S -sel, az S valamely elemét x -szel, az x -hez tartozó valószínűséget pedig $p(x)$ -szel. Ekkor a

$$\sum_x x\, p(x)$$

számot az X valószínűségi változó **várható érték**ének nevezzük (Várható érték angolul: expected value). A várható értéket E(X) -szel jelöljük:

$$\mathrm{E}(\ X\ ) = \sum_x x\, p(x)$$

A kifejezésben – mint korábban is és a későbbiekben is – a szummázás az összes lehetséges x -re értendő.

Ha egy eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, akkor a

$$\sum_x x\, p(x)$$

számot az **eloszlás várható érték**ének nevezzük.

1. Megjegyzés: Ezeket a képleteket a súlypont képletével összevetve azonnal szembetűnik, hogy a súlypont

$$\frac{\sum_x x \cdot p(x)}{\sum_x p(x)}$$

általános, fizikából tanult képlete a $\sum_x p(x) = 1$ feltétel teljesülése miatt a várható érték képletébe megy át. Tehát egy valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a várható érték egybeesik a tömegeloszlás súlypontjával.

Kicsit általánosabban, ha a szóban forgó valószínűségi változó mellett még egy $y = t(x)$ transzformációt (függvényt) is tekintünk, és képezzük a $t(X)$ valószínűségi változót, akkor a

$$\mathrm{E}\,(\ t(X)\ ) = \sum_x t(x)\, p(x)$$

számot **a transzformációval nyert** $t(X)$ **valószínűségi változó várható érték**ének nevezzük.

Ha egy eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk a számegyenesen, és tekintünk egy $y = t(x)$ transzformációt (függvényt) is, akkor a

$$\sum_x t(x)\, p(x)$$

számot a szóban forgó eloszlásból **a** t **transzformációval kapott eloszlás várható értéké**nek nevezzük.

Még általánosabban, egy (X, Y) kétdimenziós valószínűségi változó esetén ha még egy $z = t(x, y)$ transzformációt (függvényt) is tekintünk, és képezzük a $t(X, Y)$ valószínűségi változót, akkor a

$$\mathrm{E}\,(\ t(X,Y)\ ) = \sum_{(x,y)} t(x,y)\, p(x,y)$$

számot **a transzformációval nyert** $t(X,Y)$ **valószínűségi változó várható érték**ének nevezzük.

Ha egy síkbeli eloszlást csak úgy önmagában – mindenféle valószínűségi változó nélkül – vizsgálunk, és tekintünk egy $y = t(x,y)$ transzformációt (függvényt) is, akkor a

$$\sum_{(x,y)} t(x,y)\; p(x,y)$$

számot a szóban forgó eloszlásból **a** t **transzformációval kapott eloszlás várható értéké**nek nevezzük.

Még ennél is általánosabban, ha egy absztrakt S halmazon vizsgálunk egy eloszlást, és tekintünk egy $y = t(x)$ transzformációt (függvényt) is, mely az S -ről a számegyenesre képez akkor a

$$\sum_{x \in S} t(x)\; p(x)$$

számot a szóban forgó eloszlásból **a** t **transzformációval kapott eloszlás várható értéké**nek nevezzük. Ha valamely véletlen jelenségben az S halmaz elemei véletlenszerűen bukkannak fel, és X -szel jelöljük a felbukkanó véletlen elemet, akkor ez a szám egyúttal **a** $t(X)$ **valószínűségi változó várható értéke** is.

2. Megjegyzés: Ha az $y = t(x)$ függvény a négyzetre emelés függvénye, azaz $t(x) = x^2$, akkor a

$$\sum_x x^2\; p(x)$$

kifejezéshez jutunk. Ezt az értékét a valószínűségi változó avagy a szóban forgó eloszlás **második momentum**ának nevezzük. A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a második momentum egybeesik a tömegeloszlásnak a 0 pontra vonatkozó tehetetlenségi nyomatékával.

3. Megjegyzés: Ha c egy fix szám, és az $y = t(x)$ függvény az $x - c$ különbség négyzete, azaz $t(x) = (x-c)^2$, akkor a

$$\sum_x (x-c)^2\; p(x)$$

kifejezéshez jutunk. Ezt az értékét a valószínűségi változó avagy a szóban forgó eloszlás c **-re vonatkozó második momentum**ának nevezzük. A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a második momentum egybeesik a tömegeloszlásnak a c pontra vonatkozó tehetetlenségi nyomatékával.

4. Megjegyzés: Ha a t függvény az identitás függvény, azaz $t(x) = x$, akkor a

$$\sum_x x\; p(x)$$

kifejezéshet jutunk, ami a várható érték képlete. Ezt a képletet a második momentummal összevetve látjuk, hogy a várható értéket jogos **első momentum**nak is nevezni.

5. Megjegyzés: Ha a lehetséges értékeket indexezve soroljuk fel: $x_1, x_2, \ldots$, és a hozzájuk tartozó valószínűségeket pedig $p_1, p_2, \ldots$ -vel jelöljük, akkor a képletek így festenek:

$$\mathrm{E}(\ X\ ) = \sum_i x_i\; p_i$$

$$\mathrm{E}\,(\ t(X)\ ) = \sum_i t(x_i)\; p_i$$

6. Megjegyzés: Jelöljük X -szel a kockadobás eredményét. X^2 várható értékét szemléletesen úgy is kiszámolhatnánk, hogy a dobókocka oldalaira új címkéket ragasztanánk $1, 4, 9, 16, 25, 36$ számokkal. Így nem meglepő, hogy

E($t(X)$) képletében a $p(x)$ értékek változatlanok maradnak, míg x-eket mindenhol formálisan kicseréljük $t(x)$ -re.

7. Megjegyzés: Ha X lehetséges értékeinek S halmaza végtelen sok elemet tartalmaz, akkor a fenti sorok végtelen sorok, ezért ilyenkor előfordulhat, hogy a sor nem konvergens, hanem értéke plusz- vagy mínusz végtelen, vagy akár az is, hogy a sor értéke nem definiált, mert a sor pozitív tagjaiból álló sor és a negatív tagjaiból álló sor is divergál, és ezért a $-\infty + \infty$ alakú, nem értelmezhető eset áll elő. Ha X lehetséges értékeinek S halmaza csak véges sok elemet tartalmaz, akkor ilyen nem fordulhat elő, és a várható érték egy jól definiált szám.

6.2. Feltételes várható érték egy eseményen belül

Példa: Fiatal házaspár egy vagy több gyerekkel. Korábbi példánk szerint, ha egy fiatal házaspárt véletlenszerűen választunk, és tekintjük az

$$X = \text{ gyerekek száma}$$

valószínűségi változót, akkor a valószínűségi változó eloszlása:

x	0	1	2	3
$p(x)$	0.2	0.4	0.3	0.1

Táblázat: *Gyerekek számának eloszlása*

Egyszerű számolással kapjuk, hogy az X valószínűségi változó várható értéke

$$\text{E}(\ X\ ) = 0 \cdot 0.2 + 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.3$$

Tehát ha véletlenszerűen választunk sok családot, akkor körülbelül 1.3 .

Kérdés: Ha a kiválasztott családok közül csak a gyerekes családok gyerekei számának az átlagát vesszük, akkor ez az átlag körülbelül mennyivel egyenlő?

A válasz egyszerű: korábban meghatároztuk az X valószínűségi változó $X \geq 1$ feltétel melletti (feltételen belüli) feltételes eloszlását:

x	1	2	3
$p(x)$	$\frac{4}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Táblázat: *Feltételes eloszlás az* $X \geq 1$ *feltétel mellett*

Ennek várható értéke:

$$1 \cdot \frac{4}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{13}{8} = 1.625$$

Általánosan is megfogalmazzuk a példában leírtakat: Ha S a számegyenesnek egy részhalmaza, és tekintünk egy

$$\{p(x),\ x \in S\}$$

eloszlást S-en, továbbá A részhalmaza S-nek, akkor a

$$\left\{ \frac{p(x)}{P(A)},\ x \in A \right\}$$

feltételes eloszlás

$$\sum_{x \in A} x\ \frac{p(x)}{P(A)}$$

várható értékét az eredeti eloszlásnak **az** A **esemény bekövetkezése melletti feltételes várható értéké**nek vagy **az** A **-n belüli feltételes várható értéké**nek nevezzük, és $\mathrm{E}(\ X \mid A\ )$ -val jelöljük.

A fenti példával kapcsolatban ezt írhatjuk:

$$\mathrm{E}(X \mid X \geq 1\ ) = 1.625$$

6.3. Geometriai eloszlás jellemzése feltételes várható értékekkel *(Extra tananyag)*

Az alábbi két állítás mindegyike fontos jellemzése a geometriai eloszlásoknak. Az elsőnek a bizonyítása triviális, nem is adjuk meg. A másodiknak a bizonyítása kissé nehéznek tűnhet, de igazából egyszerű és érdekes.

1. Állítás: Ha egy pozitív egész értékű, diszkrét X valószínűségi változóra teljesül, hogy az

$$\mathrm{P}(\ X = s + 1 \mid X\ >\ s\ )$$

feltételes valószínűség nem függ s -től, akkor X optimista geometriai eloszlást követ.

Bizonyítás: Triviális.

2. Állítás: Ha egy pozitív értékű, diszkrét X valószínűségi változóra teljesül, hogy az

$$\mathrm{E}(\ X\ -\ s \mid X\ >\ s\ )$$

feltételes várható érték nem függ s-től, akkor X optimista geometriai eloszlást követ.

Bizonyítás: Az $\mathrm{E}(\ X - s \mid X > s\ )$ feltételes várható értéket felírjuk összegekkel:

$$\mathrm{E}(\ X - s \mid X\ >\ s\ )\ = \frac{1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots}{p_{s+1} + p_{s+2} + p_{s+3} + \ldots}$$

Ha ez a feltételes várható érték egy c konstanssal egyenlő, akkor minden s -re, igaz, hogy

$$\frac{1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots}{p_{s+1} + p_{s+2} + p_{s+3} + \ldots} = c$$

A nevezővel átszorozva ez adódik:

$$1\,p_{s+1} + 2\,p_{s+2} + 3\,p_{s+3} + \ldots \;\; = c\,(p_{s+1} + p_{s+2} + p_{s+3} + \ldots)$$

Írjuk fel ezt a legutolsó egyenletet s helyett $s-1$-re is:

$$1\,p_s + 2\,p_{s+1} + 3\,p_{s+2} + \ldots \;\; = \; c\,(p_s + p_{s+1} + p_{s+2} + \ldots)$$

Az alsó egyenletből kivonva a felsőt, a következő, minden s-re fennálló egyenletet kapjuk:

$$p_s + p_{s+1} + p_{s+2} + \ldots \; = \; c\,p_s$$

Írjuk fel ugyanezt az egyenletet s helyett $s+1$-re:

$$p_{s+1} + p_{s+2} + p_{s+3} + \ldots \; = \; c\,p_{s+1}$$

Most vonjuk ki a felsőből az alsót, ezt kapjuk:

$$p_s = \; c\;(\,p_s - p_{s+1}\,)$$

ahonnan

$$c\,p_{s+1} \;\; = \;\; (c-1)\;p(s)$$

vagyis

$$p_{s+1} = \; \frac{c-1}{c}\,p(s)$$

ami világosan mutatja, hogy a $p_1, p_2, p_3, \ldots$ sorozat minden eleme az előző elemből egy konstanssal való szorzással adódik, ezért a sorozat egy geometriai sorozatot alkot. Tehát tényleg egy geometriai eloszlásról van szó.

Geometriai eloszlás egy számsorozat elemeinek halmazán: Ha az a_1 , a_2 , $\ldots$ sorozat egy számtani sorozatot alkot, akkor

$$p(a_k) = (1-p)^{k-1}p \qquad (k = 1, 2, \ldots)$$

képlettel definiált eloszlás neve: **geometriai eloszlás a** a_1 , a_2 , $\ldots$ **számsorozat elemeinek halmazán**.

6.4. Variancia és szórás

Ha c egy szám, akkor a valószínűségi változó, illetve az eloszlás c -re vonatkozó második momuentumának nevezzük a

$$\sum_x (x-c)^2\;p(x)$$

értéket. A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a c -re vonatkozó második momuentum egybeesik a tömegeloszlásnak a c pontra vonatkozó tehetetlenségi nyomatékával. Ha c az eloszlás várható értéke, akkor a második momuentum neve **variancia** vagy **szórásnégyzet**, amit VAR(X) -szel jelölünk:

$$\mathrm{VAR}(\;X\;) \;\; = \;\; \sum_x (x - \mathrm{E}(\;X\;))^2\;p(x)$$

A valószínűségi eloszlást tömegeloszlásként is elképzelve látjuk, hogy a variancia egybeesik a tömegeloszlásnak a súlypontjára vonatkozó tehetetlenségi nyomatékával.

A variancia négyzetgyökét, a

$$\sqrt{\sum_x (x-c)^2\;p(x)}$$

számot **szórás**nak hívjuk, és SD(X)-vel jelöljük:

$$\mathrm{SD}(\ X\ )\ =\ \sqrt{\mathrm{VAR}(\ X\ )}\ =\ \sqrt{\sum_x (x-\mathrm{E}(\ X\ ))^2\ p(x)}$$

1. Megjegyzés: Emlékeztetünk rá, hogy egy valószínűségi változó, illetve egy eloszlás 0 -ra vonatkozó második momentumát, a

$$\sum_x x^2\ p(x)$$

számot egyszerűen csak **második momuentum**nak nevezzük. Könnyű belátni, hogy igaz a

$$\text{variancia}\ =\ \text{második momentum} - (\ \text{első momentum}\ )^2$$

összefüggés.

2. Megjegyzés: Nyilvánvalóan igaz a Steiner egyenlőség, és annak következménye a Steiner egyenlőtlenség:

Steiner egyenlőség: Akármilyen c-re igaz, hogy

$$\mathrm{E}\left(\ (X-c)^2\ \right)\ =\ \mathrm{VAR}(\ X\ ) + (\ c\ - \mathrm{E}(\ X\ ))^2$$

Steiner egyenlőtlenség: Akármilyen c -re igaz, hogy

$$\mathrm{E}\left(\ (X-c)^2\ \right)\ \geq\ \mathrm{VAR}(\ X\ )$$

$c\ = \mathrm{E}(\ X\ )$ esetén az egyenlőség, más c értékekre az egyenlőtlenség teljesül.

7. Nagy számok törvényei

Mind elméleti, mind gyakorlati szempontból fontos tény, hogy nagy számú kísérletek eredményeinek átlaga, szórása és sok egyéb jellemzője – bár függnek a véletlentől – mégis egy fajta stabilitást mutatnak: értékük közel van egy olyan értékhez, amit elméleti úton – a kísérletek elvégzése nélkül – is meg lehet határozni. Erre a tényre, mint a nagy számok törvényeire szoktunk hivatkozni.

7.1. NSZT a kísérleti eredmények átlagára

Heurisztikus megfogalmazás: Képzeljük el, hogy egy X valószínűségi változóra sok kísérletet végzünk. A kísérletek számát jelölje N, a kísérleti eredményeket, melyek véletlen számok, jelölje $X_1, X_2, \ldots, X_N$. Ha kísérleti eredményeket átlagoljuk, akkor ez az átlag – nagy kísérletszám esetén – körülbelül az X várható értékével egyenlő:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \sum_x x\, p(x) = \mathrm{E}(\ X\ )$$

Vázlatos bizonyítás: Az X valószínűségi változó lehetséges értékei legyenek

$$x_1,\ x_2, \ldots$$

Figyelem! Az $X_1, X_2, \ldots, X_N$ nagybetűk kísérleti eredményeket, tehát véletlen értékeket jelölnek. A $x_1,\ x_2, \ldots$ kisbetűk segítségével pedig a lehetséges értékeket soroljuk fel. Ha például az X valószínűségi változó egy dobókockával dobott véletlen számot jelöl, és – mondjuk – százszor dobunk, akkor a nagybetűvel jelölt X_1 jelenti az első dobás eredményét, X_2 jelenti a második dobás eredményét, és így tovább, X_{100} jelenti a századik dobás eredményét. A kisbetűvel jelölt $x_1,\ x_2,\ x_3,\ x_4,\ x_4,\ x_6$ értékek pedig az $1, 2, 3, 4, 5, 6$ számokat jelentik.
Jelöljék

$$N_1,\ N_2, \ldots$$

a lehetséges értékek gyakoriságait az N kísérlet során. Tehát minden i esetén N_i azt mutatja, hogy x_i hányszor következik be az N kísérlet során. Tudjuk, hogy minden i -re az N_i/N relatív gyakoriság – nagy N esetén – közel van a p_i valószínűséghez.
Az

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

kifejezés számlálója N darab tagból áll. Közöttük az x_1 érték N_1 -szer, az x_2 érték N_2 -ször, $\ldots$ fordul elő. Ezért a számláló értéke

$$X_1 + X_2 + \ldots + X_N = N_1 \cdot x_1 + N_2 \cdot x_2 + \ldots$$

a tört értéke pedig

$$\frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{N_1 \cdot x_1 + N_2 \cdot x_2 + \ldots}{N} =$$

$$= \frac{N_1}{N} \cdot x_1 + \frac{N_2}{N} \cdot x_2 + \ldots \approx p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots = \mathrm{E}(\ X\ )$$

Precíz megfogalmazás *(Extra tananyag)*:

1. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ X\ ) = \sum_x x\ p(x)$ sor értéke egyértelműen definiált véges szám, ami teljesül, ha

- a sor véges sok tagból áll, azaz az X valószínűségi változónak csak véges sok lehetséges értéke van, vagy
- végtelen sok tag esetén – ha a sor abszolút konvergens, azaz $\sum_x |x| \; p(x) < \infty$.

Ekkor $N \to \infty$ esetén az

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \;\to\; \mathrm{E}(\, X \,)$$

esemény valószínűsége 1 -gyel egyenlő.

2. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, X \,) = \sum_x x \; p(x)$ sor értéke egyértelműen plusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{x>0} \; x \; p(x)$ divergens, és

   (b) a negatív tagokat tartalmazó $\sum_{x<0} \; x \; p(x)$ konvergens.

   Ekkor $N \to \infty$ esetén az

   $$\frac{X_1 + X_2 + \ldots + X_N}{N} \;\to\; +\infty$$

   esemény valószínűsége 1 -gyel egyenlő.

3. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\, X \,) = \sum_x x \; p(x)$ sor értéke egyértelműen mínusz végtelen, ami akkor teljesül, ha

   (a) pozitív tagokat tartalmazó $\sum_{x>0} \; x \; p(x)$ konvergens, és

   (b) a negatív tagokat tartalmazó $\sum_{x<0} \; x \; p(x)$ divergens.

   Ekkor $N \to \infty$ esetén az

   $$\frac{X_1 + X_2 + \ldots + X_N}{N} \;\to\; -\infty$$

   esemény valószínűsége 1 -gyel egyenlő.

7.2. NSZT a kísérleti eredmények függvényének az átlagára

Heurisztikus megfogalmazás: Képzeljük el, hogy egy X valószínűségi változóra sok kísérletet végzünk. A kísérletek számát jelölje N , a kísérleti eredményeket, melyek véletlen számok, jelölje $X_1, X_2, \ldots, X_N$. Ha kísérleti eredményeket egy $t(x)$ függvénybe helyettesíthetjük, majd az így kapott

$$t(X_1) \; , \; t(X_2) \; , \; \ldots , t(X_N)$$

értékeket átlagoljuk, akkor ez az átlag – nagy kísérletszám esetén – körülbelül a $t(X)$ várható értékével egyenlő:

$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \;\approx\; \sum_x t(x) \; p(x) \;=\; \mathrm{E}\,(\; t(X) \;)$$

Vázlatos bizonyítás: Az X valószínűségi változó lehetséges értékei legyenek

$$x_1, \; x_2, \ldots$$

Jelölje

$$N_1, \; N_2, \ldots$$

a lehetséges értékek gyakoriságait az N kísérlet során. Tehát minden i esetén N_i azt mutatja, hogy x_i hányszor következik be az N kísérlet során. Tudjuk, hogy minden i -re az N_i/N relatív gyakoriság – nagy N esetén – közel van a p_i valószínűséghez. A

$$\frac{t(X_1)+t(X_2)+\ldots+t(X_N)}{N}$$

kifejezés számlálója N darab tagból áll. Közöttük a t függvény argumentumaként az x_1 érték N_1 -szer, az x_2 érték N_2 -ször, $\ldots$ fordul elő. Ezért a számláló értéke

$$t(X_1)+t(X_2)+\ldots+t(X_N)=N_1\cdot t(x_1)+N_2\cdot t(x_2)+\ldots$$

a tört értéke pedig

$$\frac{t(X_1)+t(X_2)+\ldots+t(X_N)}{N}=\frac{N_1\cdot t(x_1)+N_2\cdot t(x_2)+\ldots}{N}=$$

$$=\frac{N_1}{N}\cdot t(x_1)+\frac{N_2}{N}\cdot t(x_2)+\ldots\approx p_1\cdot t(x_1)+p_2\cdot t(x_2)+\ldots=\mathrm{E}(\ t(X)\ ,)$$

Precíz megfogalmazás *(Extra tananyag)*:

1. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ t(X)\ )=\sum_x t(x)\ p(x)$ sor értéke egyértelműen definiált véges szám, ami teljesül, ha
   (a) a sor véges sok tagból áll, azaz az X valószínűségi változónak csak véges sok lehetséges értéke van, vagy
   (b) végtelen sok tag esetén – ha a sor abszolút konvergens, azaz $\sum_x |t(x)|\ p(x)<\infty$.

   Ekkor $N\to\infty$ esetén a
   $$\frac{t(X_1)+t(X_2)+\ldots+t(X_N)}{N}\ \to\ \mathrm{E}(\ t(X)\ )$$
   esemény valószínűsége 1-gyel egyenlő.

2. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ t(X)\ )=\sum_x t(x)\ p(x)$ sor értéke egyértelműen plusz végtelen, ami akkor teljesül, ha
   (a) pozitív tagokat tartalmazó $\sum_{t(x)>0}\ x\ p(x)$ divergens, és
   (b) a negatív tagokat tartalmazó $\sum_{t(x)<0}\ x\ p(x)$ konvergens.

   Ekkor $N\to\infty$ esetén a
   $$\frac{t(X_1)+t(X_2)+\ldots+t(X_N)}{N}\ \to\ +\infty$$
   esemény valószínűsége 1 -gyel egyenlő.

3. Tegyük fel, hogy a várható értéket definiáló $\mathrm{E}(\ t(X)\ )\ =\ \sum_i t(x_i)\ p_i\ =\ \sum_i t(x)\ p(x)$ sor értéke egyértelműen mínusz végtelen, ami akkor teljesül, ha
   (a) pozitív tagokat tartalmazó $\sum_{t(x)>0}\ x\ p(x)$ kovergens, és
   (b) a negatív tagokat tartalmazó $\sum_{t(x)<0}\ x\ p(x)$ divergens.

   Ekkor $N\to\infty$ esetén a
   $$\frac{t(X_1)+t(X_2)+\ldots+t(X_N)}{N}\ \to\ -\infty$$
   esemény valószínűsége 1 -gyel egyenlő.

7.3. NSZT a második momentumra

Az $X_1, X_2, \ldots, X_N$ kísérleti eredmények második momentumára igaz, hogy nagy kísérletszám esetén körülbelül az X második momentumával egyenlő:

$$\frac{X_1^2 + X_2^2 + \ldots + X_N^2}{N} \approx \sum_x x^2 \, p(x) = \mathrm{E}(\ X^2\ )$$

Kicsit általánosabban: az $X_1, X_2, \ldots, X_N$ kísérleti eredmények c -re vonatkozó második momentumára igaz, hogy nagy kísérletszám esetén körülbelül az X -nek a c -re vonatkozó második momentumával egyenlő:

$$\frac{(\ X_1 - c\ )^2 + (\ X_2 - c\ )^2 + \ldots + (\ X_N - c\ )^2}{N} \approx \sum_x (\ x - c\ )^2 \, p(x) = \mathrm{E}\big(\ (X - c)^2\ \big)$$

Bizonyítás: Ha az előző részben a speciális $t(x) = x^2$, illetve a $t(x) = (x - c)^2$ függvényt vesszük akkor azonnal megkapjuk az állításokat.

7.4. NSZT a varianciára

Az $X_1, X_2, \ldots, X_N$ kísérleti eredmények

$$\frac{(\ X_1 - \overline{X}\ )^2 + (\ X_2 - \overline{X}\ )^2 + \ldots + (\ X_N - \overline{X}\ )^2}{N}$$

varianciájára igaz, hogy nagy kísérletszám esetén körülbelül az X varianciájával egyenlő:

$$\frac{(\ X_1 - \overline{X}\ )^2 + (\ X_2 - \overline{X}\ )^2 + \ldots + (\ X_N - \overline{X}\ )^2}{N} \approx \sum_x (x - \mathrm{E}(\ X\ ))^2 \, p(x) = \mathrm{VAR}(\ X\ )$$

Vázlatos bizonyítás: Ugyanis, ha az $\overline{X}_N$ átlag helyére az $\mathrm{E}(\ X\ )$ várható értéket írjuk, akkor azt kapjuk, hogy

$$\frac{(\ X_1 - \overline{X}\ )^2 + (\ X_2 - \overline{X}\ )^2 + \ldots + (\ X_N - \overline{X}\ )^2}{N} \approx$$

$$\approx \frac{(\ X_1 - \mathrm{E}(\ X\ )\ )^2 + (\ X_2 - \mathrm{E}(\ X\ )\ )^2 + \ldots + (\ X_N - \mathrm{E}(\ X\ )\ )^2}{N}$$

A jobb oldalon álló átlag az $(\ X - \mathrm{E}(\ X\ )\ )^2$ valószínűségi változó várható értékéhez van közel, ami éppen az X varianciájával egyenlő:

$$\frac{(\ X_1 - \mathrm{E}(\ X\ )\ )^2 + (\ X_2 - \mathrm{E}(\ X\ )\ )^2 + \ldots + (\ X_N - \mathrm{E}(\ X\ )\ )^2}{N} \approx \mathrm{E}\big(\ (\ X - \mathrm{E}(\ X\ )\ )^2\ \big) = \mathrm{VAR}(\ X\ )$$

7.5. NSZT a szórásra

Így aztán nagy kísérletszám esetén az $X_1, X_2, \ldots, X_N$ kísérleti eredmények szórása – igen általános feltételek mellett – körülbelül az X szórásával egyenlő:

$$\sqrt{\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_N - \overline{X})^2}{N}} \approx \sqrt{\sum_x (x - \mathrm{E}(X))^2\, p(x)} = \mathrm{SD}(X)$$

7.6. NSZT a mediánra

Nagy kísérletszám esetén az $X_1, X_2, \ldots, X_N$ kísérleti eredményekből számított medián körülbelül az X eloszlásából számított (elméleti) mediánnal egyenlő.

8. A várható érték, variancia és szórás általános tulajdonságai

Ebben a fejezetben felsoroljuk a várható érték, a variancia és a szórás legfontosabb általános tulajdonságait. Ezek a tulajdonságok diszkrét és folytonos valószínűségi változókra, eloszlásokra egyaránt igazak. Arra bíztatom a Kedves Olvasót, hogy a tulajdonságokat

- gondolja meg,
- értelmezze őket a saját szemlélete szerint (jól!), és
- ellenőrizze konkrét példákra készített szimulációkkal!

A szimulációk készítésének élménye a tulajdonságok megtapasztalását teszi lehetővé, ami nagy mértékben megkönnyíti a szabályok megértését, megjegyzését.

A lentebbi képletekben X, Y, X_1, X_2, …, X_n valószínűségi változókat a, b, c, n, a_1, a_2, …, a_n, konstansokat jelentenek.

8.1. Várható érték tulajdonságai

1. **Összeadás – általános eset:**

   Összeg várható értéke egyenlő a tagok várható értékeinek az összegével:

   (a) Két tagra:
   $$\mathrm{E}(\ X+Y\ ) \ = \ \mathrm{E}(\ X\ ) + \mathrm{E}(\ Y\ )$$

   (b) Több tagra:
   $$\mathrm{E}(\ X_1+X_2+\ldots+X_n\ ) \ = \ \mathrm{E}(\ X_1\ ) + \mathrm{E}(\ X_2\ ) + \ldots + \mathrm{E}(\ X_n\ )$$

2. **Konstanssal való szorzás:**

   Ha egy valószínűségi változót egy konstanssal szorzunk, akkor a várható érték is a konstanssal szorzódik:
   $$\mathrm{E}(\ c \cdot X\ ) \ = \ c \ \cdot \ \mathrm{E}(\ X\ )$$

3. **Linearitás:**

   Egy lineáris kombináció várható értéke egyenlő a tagok várható értékeinek ugyanolyan lineáris kombináviójátval:

   (a) Két tagra:
   $$\mathrm{E}(\ a\,X \ + \ b\,Y\ ) \ = \ a\,\mathrm{E}(\ X\ ) \ + \ b\,\mathrm{E}(\ Y\ )$$

   (b) Több tagra:
   $$\mathrm{E}(\ a_1\,X_1 + a_2\,X_2 + \ldots + a_n\,X_n\ ) \ = \ a_1\,\mathrm{E}(\ X_1\ ) + a_2\,\mathrm{E}(\ X_2\ ) + \ldots + a_n\,\mathrm{E}(\ X_n\ )$$

4. **Összeadás – azonos várható értékek esetén:**

   Ha a tagok várható értékei egyenlők, akkor az összeg várható értéke egyenlő a várható értékek (μ -vel jelölt) közös értéke szorozva a tagok számával:
   $$\mathrm{E}(\ X_1+X_2+\ldots+X_n\ ) \ = \ n\,\mu$$

5. **Átlagolás – azonos várható értékek esetén:**

   Ha a tagok várható értékei egyenlők, akkor az átlag várható értéke egyenlő a várható értékek (μ -vel jelölt) közös értékével:

$$\mathrm{E}\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \right) \;=\; \mu$$

6. **Független valószínűségi változók szorzata:**

   Független valószínűségi változók szorzatának várható értéke egyenlő a várható értékek szorzatával:

   (a) Két tényezőre:

$$\mathrm{E}(\; X \cdot Y \;) \;=\; \mathrm{E}(\; X \;) \;\cdot\; \mathrm{E}(\; Y \;)$$

   (b) Több tényezőre:

$$\mathrm{E}(\; X_1 \cdot X_2 \cdot \;\ldots\; \cdot X_n \;) \;=\; \mathrm{E}(\; X_1 \;) \;\cdot\; \mathrm{E}(\; X_2 \;) \;\cdot\; \ldots \;\cdot\; \mathrm{E}(\; X_n \;)$$

Bizonyítás: Az 1-5. állítások kézenfekvőek. A 6. állítás bizonyítása: A $t(x,y) = xy$ függvényre alkalmazva a várható érték

$$\mathrm{E}\,(\; t(X,Y)\;) = \sum_{(x,y)} t(x,y)\;\; p(x,y)$$

általános képletét, ezt kapjuk:

$$\mathrm{E}\,(\; X \cdot Y \;) = \sum_{(x,y)} x \cdot y \;\cdot\; p(x,y) = \sum_{(x,y)} x \cdot y \;\cdot\; p_1(x) \cdot p_2(y) =$$

$$= \sum_{(x,y)} x \cdot p_1(x) \;\cdot\; y \cdot p_2(y) = \sum_{x} x \cdot p_1(x) \;\cdot\; \sum_{y} y \cdot p_2(y) = \mathrm{E}\,(\; X \;) \;\cdot\; \mathrm{E}\,(\; Y \;)$$

8.2. Variancia tulajdonságai

1. **Összeadás (két tagra) – általános eset:**

   Összeg varianciája egyenlő a varianciák összegével plusz a kovariancia kétszerese:

$$\mathrm{VAR}(\; X + Y \;) = \mathrm{VAR}(\; X \;) + \mathrm{VAR}(\; Y \;) + 2\,\mathrm{COV}(\; X, Y \;)$$

   ahol COV $(\; X, Y \;)$ az X és Y közötti kovarianciát jelentik. A kovariancia fogalmát később tanuljuk.

2. **Összeadás (két tagra) – független tagok esetén:**

   Független valószínűségi változók összegének a varianciája egyenlő a tagok varianciáinak az összegével:

$$\mathrm{VAR}(\; X + Y \;) = \mathrm{VAR}(\; X \;) + \mathrm{VAR}(\; Y \;)$$

3. **Konstanssal való szorzás:**

   Ha egy valószínűségi változót egy konstanssal szorzunk, akkor a variancia a konstans négyzetével szorzódik:

$$\mathrm{VAR}(\; c \cdot X \;) \;=\; c^2 \;\cdot\; \mathrm{VAR}(\; X \;)$$

4. **Összeadás – független tagok és azonos varianciák esetén:**

   Ha a tagok függetlenek, és varianciáik egyenlők, akkor az összeg varianciája egyenlő varianciák (σ^2 -val jelölt) közös értéke szorozva a tagok számával:

$$\mathrm{VAR}(\; X_1 + X_2 + \ldots + X_n \;) \;=\; n\,\sigma^2$$

5. **Átlagolás – független tagok és azonos varianciák esetén:**
Ha a tagok függetlenek, és varianciáik egyenlők, akkor az átlag varianciája egyenlő varianciák (σ^2 -val jelölt) közös értéke osztva a tagok számával:

$$\mathrm{VAR}\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \right) = \frac{\sigma^2}{n}$$

Megjegyzés: Nem függetlenek valószínűségi változókra ezek a szabályok általában nem igazak.

Bizonyítás: Csak a 2. állítás szorul magyarázatra, a többi kézenfekvő. A 2. állítás bizonyítása, amikor $+$ jel áll a bal oldalon:

A variancia definícióját az $X + Y$ valószínűségi változóra alkalmazzuk:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big([\, X + Y - \mathrm{E}(\, X + Y \,)\,]^2 \Big)$$

Felhasználjuk, hogy $\mathrm{E}(\, X + Y \,) = \mathrm{E}(\, X \,) + \mathrm{E}(\, Y \,)$, és ezt kapjuk:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big([\, X + Y - \mathrm{E}(\, X \,) - \mathrm{E}(\, Y \,)\,]^2 \Big)$$

A szögletes zárójelen belül átrendezzük a tagokat:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big([\, (\, X - \mathrm{E}(\, X \,)\,) + (\, Y - \mathrm{E}(\, Y \,)\,)\,]^2 \Big)$$

Elvégezzük a négyzetre emelést:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big((\, X - \mathrm{E}(\, X \,)\,)^2 + (\, Y - \mathrm{E}(\, Y \,)\,)^2 + 2 \cdot (\, X - \mathrm{E}(\, X \,)\,) \cdot (\, Y - \mathrm{E}(\, Y \,)\,) \Big)$$

A három tag összegének a várható értékére három várható érték összegére esik szét:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big((\, X - \mathrm{E}(X)\,)^2 \Big) + \mathrm{E}\Big((\, Y - \mathrm{E}(\, Y \,)\,)^2 \Big) + \mathrm{E}\,(\, 2 \cdot (\, X - \mathrm{E}(\, X \,)) \cdot (\, Y - \mathrm{E}(\, Y \,)\,)\,)$$

A harmadik tagban a függetlenség miatt a szorzat várható értéke egyenlő a várható értékek szorzatával,

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{E}\Big((\, X - \mathrm{E}(\, X \,)\,)^2 \Big) + \mathrm{E}\Big((\, Y - \mathrm{E}(\, Y \,)\,)^2 \Big) + 2 \cdot \mathrm{E}\,(\, X - \mathrm{E}(\, X \,)\,) \cdot \mathrm{E}\,(\, Y - \mathrm{E}(Y)\,)$$

Felhasználva, hogy

$$\mathrm{E}\,(\, X - \mathrm{E}(\, X \,)\,) = 0 \;\text{ és }\; \mathrm{E}\,(\, Y - \mathrm{E}(\, Y \,)\,) = 0$$

kiadódik, hogy a harmadik tag eltűnik:

$$\mathrm{VAR}\,(X + Y\,) = \mathrm{E}\Big((\, X - \mathrm{E}(X)\,)^2 \Big) + \mathrm{E}\Big((\, Y - \mathrm{E}(Y)\,)^2 \Big)$$

Itt pedig ráismerünk a varianciák összegére:

$$\mathrm{VAR}(\, X + Y \,) = \mathrm{VAR}(\, X \,) + \mathrm{VAR}(\, Y \,)$$

Megjegyzés: Az Olvasó győződjön meg arról, hogy a fenti formulákban a bal oldali összegekben a $+$ jel mindenhol kicserélhető $\pm$ jelre is:

$$\mathrm{VAR}(\, X \pm Y \,) = \mathrm{VAR}(\, X \,) + \mathrm{VAR}(\, Y \,)$$

$$\mathrm{VAR}(\, X_1 \pm X_2 \pm \ldots \pm X_n \,) = \mathrm{VAR}(\, X_1 \,) + \mathrm{VAR}(\, X_2 \,) + \ldots + \mathrm{VAR}(\, X_n \,)$$

$$\mathrm{VAR}(\, X_1 \pm X_2 \pm \ldots \pm X_n \,) = n\,\sigma^2$$

$$\mathrm{VAR}\left(\frac{X_1 \pm X_2 \pm \ldots \pm X_n}{n} \right) = \frac{\sigma^2}{n}$$

8.3. Szórás tulajdonságai

1. **Konstanssal való szorzás:**
   Ha egy valószínűségi változót egy konstanssal szorzunk, akkor a szórás a konstans abszolút értékével szorzódik:
   $$\mathrm{SD}(\ c\,X\ )\ =\ |c|\ \mathrm{SD}(\ X\ )$$
2. **Négyzetgyök szabály az összegre:**
   Ha a tagok függetlenek, és varianciáik egyenlők, akkor az összeg szórása egyenlő szórások (σ -val jelölt) közös értéke szorozva a tagok számának a négyzetgyökével.
   $$\mathrm{SD}(\ X_1 + X_2 + \ldots + X_n\ )\ =\ \sqrt{n}\,\sigma$$
3. **Négyzetgyök szabály az átlagra:**
   Ha a tagok függetlenek, és varianciáik egyenlők, akkor az átlag szórása egyenlő szórások (σ -val jelölt) közös értéke osztva a tagok számának a négyzetgyökével.
   $$\mathrm{SD}\left(\ \frac{X_1 + X_2 + \ldots + X_n}{n}\ \right)\ =\ \frac{\sigma}{\sqrt{n}}$$

Egyik állítás sem szorul különösebb magyarázatra.

Megjegyzés: Az Olvasó győződjön meg arról, hogy a fenti formulákban a bal oldali összegekben a $+$ jel mindenhol kicserélhető $\pm$ jelre is:

$$\mathrm{SD}^2(\ X \pm Y\ )\ =\ \mathrm{SD}^2(\ X\ ) + \mathrm{SD}^2(\ Y\ )$$
$$\mathrm{SD}(\ X \pm Y\ )\ =\ \sqrt{\mathrm{SD}^2(\ X\ ) + \mathrm{SD}^2(\ Y\ )}$$
$$\mathrm{SD}(\ X_1 \pm X_2 \pm \ldots \pm X_n\ )\ =\ \sqrt{n}\,\sigma$$
$$\mathrm{SD}\left(\ \frac{X_1 \pm X_2 \pm \ldots \pm X_n}{n}\ \right)\ =\ \frac{\sigma}{\sqrt{n}}$$

9. Nevezetes eloszlások várható értéke, varianciája, szórása – formulák

9.1. Hipergeometrikus eloszlás

paraméterek: M , N , n

várható érték: E(X) $= n \, \frac{M}{N}$

variancia: VAR(X) $= n \, \frac{N-n}{N-1} \, \frac{M}{N} \, (1 - \frac{M}{N})$

szórás: SD(X) $= \sqrt{n \, \frac{N-n}{N-1} \, \frac{M}{N} \, (1 - \frac{M}{N})}$

módusz:

Ha $(n+1) \, \frac{M+1}{N+2}$ egész szám, akkor két módusz van: $(n+1) \, \frac{M+1}{N+2}$ és $(n+1) \, \frac{M+1}{N+2} - 1$

Ha $(n+1) \, \frac{M+1}{N+2}$ nem egész szám, akkor egy módusz van: $\left\lfloor (n+1) \, \frac{M+1}{N+2} \right\rfloor$

9.2. Binomiális eloszlás

paraméterek: n és p

várható érték: E(X) $= n \, p$

variancia: VAR(X) $= n \, p \, (1-p)$

szórás: SD(X) $= \sqrt{n \, p \, (1-p)}$

módusz:

Ha $(n+1) \, p$ egész, akkor két módusz van: $(n+1) \, p$ és $(n+1) \, p - 1$

Ha $(n+1) \, p$ nem egész, akkor egy módusz van: $\lfloor (n+1) \, p \rfloor$

9.3. Indikátor eloszlás

paraméter: p

várható érték: E(X) $= p$

variancia: VAR(X) $= p \, (1-p)$

szórás: SD(X) $= \sqrt{p \, (1-p)}$

módusz:

Ha $p = \frac{1}{2}$, akkor két módusz van: 0 és 1

Ha $p < \frac{1}{2}$, akkor egy módusz van: 0

Ha $p > \frac{1}{2}$, akkor egy módusz van, az 1

9.4. Optimista geometriai eloszlás

paraméter: p

várható érték: $\mathrm{E}(\ X\ )\ =\ \frac{1}{p}$

variancia: $\mathrm{VAR}(\ X\ )\ =\ \frac{(1-p)}{p^2}$

szórás: $\mathrm{SD}(\ X\ )\ =\ \frac{\sqrt{(1-p)}}{p}$

módusz: 1

9.5. Pesszimista geometriai eloszlás

paraméter: p

várható érték: $\mathrm{E}(\ X\ )\ =\ \frac{1}{p}-1\ =\ \frac{1-p}{p}$

variancia: $\mathrm{VAR}(\ X\ )\ =\ \frac{(1-p)}{p^2}$

szórás: $\mathrm{SD}(\ X\ )\ =\ \frac{\sqrt{(1-p)}}{p}$

módusz: 0

9.6. Optimista negatív binomiális eloszlás

paraméterek: $n\ ,\ p$

várható érték: $\mathrm{E}(\ X\ )\ =\ \frac{n}{p}$

variancia: $\mathrm{VAR}(\ X\ )\ =\ \frac{n\ (1-p)}{p^2}$

szórása: $\mathrm{SD}(\ X\ )\ =\ \frac{\sqrt{n\ (1-p)}}{p}$

módusza:

Ha $\frac{n-1}{p}$ egész, akkor két módusz van: $\frac{n-1}{p}+1$ és $\frac{n-1}{p}$

Ha $\frac{n-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor \frac{n-1}{p} \right\rfloor + 1$

9.7. Pesszimista negatív binomiális eloszlás

paraméterek: n, p

várható érték: $\mathrm{E}(\ X\ ) \ = \ \frac{n}{p} - n \ = \ n\,\frac{1-p}{p}$

variancia: $\mathrm{VAR}(\ X\ ) \ = \ \frac{n\,(1-p)}{p^2}$

szórás: $\mathrm{SD}(X) \ = \ \frac{\sqrt{n\,(1-p)}}{p}$

módusz:

Ha $\frac{n-1}{p}$ egész, akkor két módusz van: $\frac{n-1}{p} - n + 1$ és $\frac{n-1}{p} - n$

Ha $\frac{n-1}{p}$ nem egész, akkor egy módusz van: $\left\lfloor \frac{n-1}{p} \right\rfloor - n + 1$

9.8. Poisson eloszlás

paraméter: λ

várható érték: λ

variancia: λ

szórás: $\sqrt{\lambda}$

módusz:

Ha λ egész, akkor két módusz van: λ és $\lambda - 1$

Ha λ nem egész, akkor egy módusz van: $\lfloor \lambda \rfloor$

Feladat: Telefonhívások. Tanszékünkre a délelőtti órákban óránként átlagosan 3.3 telefonhívás érkezik. Mi a valószínűsége annak, hogy

1. délelőtt 9 és 10 óra között pontosan 4 hívás érkezik?
2. délelőtt 9 és 11 óra között pontosan 4 hívás érkezik?
3. délelőtt 9 és 11 óra között 4 -nél több hívás érkezik?

Megoldás: Egy adott időintervallumban a hívások száma Poisson eloszlást követ. Egy óra alatt a várható érték 3.3 , két óra alatt 6.6 . Ezért

1.
$$\text{P(délelőtt 9 és 10 között pontosan 4 hívás)} \ =$$
$$= \frac{(3.3)^4}{4!}\,\mathrm{e}^{-3.3} \ = \ \texttt{POISSON(4 ; 3.3 ; FALSE)} \ = \ 0.18$$

2.

$$\text{P(délelőtt } 9 \text{ és } 11 \text{ között pontosan 4 hívás)} =$$

$$= \frac{(6.6)^4}{4!}\, \mathrm{e}^{-6.6} = \texttt{POISSON(4 ; 6.6 ; FALSE)} = 0.11$$

3.

$$\text{P(délelőtt } 9 \text{ és } 11 \text{ között } 4 \text{ -nél több hívás)} =$$

$$= 1 - \text{P(délelőtt 9 és 11 között 4 vagy kevesebb hívás)} =$$

$$= 1 - \sum_{k=0}^{4} \frac{(6.6)^k}{k!}\, \mathrm{e}^{-6.6} = \texttt{1 - POISSON(4 ; 6.6 ; TRUE)} = 0.79$$

Feladat: Csillaghullás. Fogadjuk el, hogy augusztus közepén átlagosan kb. 10 percenként lehet egy-egy csillaghullást látni. Mi a valószínűsége annak, hogy 15 perc alatt (pontosan) 2 csillaghullást lát valaki?

Megoldás: A 15 perc alatti csillaghullások száma Poisson eloszlást követ, hiszen sok a meteorit a világűrben, melyek egymással mit sem törődve száguldoznak, és minden meteoritra igaz, hogy a tekintett 15 perc alatt kis eséllyel lép be a fejünk felett a légkörbe. Mivel átlagosan 10 percenként jönnek a csillaghullások, 15 perc alatt a csillaghullások számának a várható értéke 1.5 . A Poisson eloszlás paramétere 1.5. Ezért a kérdezett valószínűség:

$$\text{P(2 csillaghullás)} = \frac{(1.5)^2}{2!}\, \mathrm{e}^{-1.5} =$$

$$= \texttt{POISSON(2 ; 1.5 ; FALSE)} = 0.25$$

9.9. Példa: Ha eltalálod, mindet neked adom

Az alábbi feladat is egy kedves, érdeklődő diák ötlete alapján született:

Feladat: A Poisson eloszlással való ismerkedésünk elején szerepelt a **Hány hal lesz az *ÖREG* halász hálójában?** című feladat (lásd: 45. oldal). Ebben a feladatban a halász ezt mondta nekem:

Ha eltalálod, hogy hány hal lesz a hálóban, meghívlak vacsorára.

Ott kiderült, hogy a vacsorán való részvételem esélye akkor a legnagyobb, ha 2 halra tippelek.

Ha a célom nem a vacsorán való részvétel, hanem valami más, akkor nyilván máshogyan kell taktikáznom. Kérdés: hogyan gondolkodtam és taktikáztam volna, ha a halász ezt mondta volna:

Ma még sokszor kiemelem a hálót.
Mindig tippelhetsz, hogy hány hal lesz a hálóban.
Amikor eltalálod, neked adom a hálóban lévő összes halat.

Megoldás: Kézenfekvő, hogy ebben az esetben arra törekednék, hogy a nyert halak összes száma avagy – ami ugyanazt jelenti – a nyert halak számának az átlaga a lehető legnagyobb legyen. Mivel a hálót sokszor emeli ki a halász, a nyert halak számának az átlaga közelíthető az átlag elméleti értékével, a várható értékkel. Ezért azt kell megkeresni, hogy a nyert halak számának a várható értéke milyen tipp esetén lesz a lehető legnagyobb.

A halak száma a hálóban – mint tudjuk régebbről – Poisson eloszlást követ 2.81 paraméterrel. Ha én minden merítésnél k halra tippelek, akkor minden alkalommal

$$\frac{(2.81)^k}{k!}e^{-2.81} \qquad \text{valószínűséggel} \quad k \quad \text{darab}$$

$$1 - \frac{(2.81)^k}{k!}e^{-2.81} \qquad \text{valószínűséggel} \quad 0 \quad \text{darab}$$

halat nyerek. Ezért a nyert halak számának a várható értéke

$$k \cdot \frac{(2.81)^k}{k!}e^{-2.81} + 0 \cdot \left(1 - \frac{(2.81)^k}{k!}e^{-2.81} \right) = k \cdot \frac{(2.81)^k}{k!}e^{-2.81}$$

Ezért azt kell megkeresni, hogy a

$$k \cdot \frac{(2.81)^k}{k!}e^{-2.81}$$

kifejezésnek milyen k értéknél van a maximuma. Ezt a k értéket akár a numerikus értékek táblázatából, akár a maximum keresésére tanult módszerrel könnyű megtalálni. Legyen az Olvasó feladata ellenőrizni, hogy a maximum a $k = 3$ értékre adódik.

Megjegyzés: Természetesen nem csak úgy lehet taktikázni, hogy mindig ugyanazzal a k számmal tippelünk. Legyen az is az Olvasó feladata, hogy belássa: rosszabbul járnék, ha nem mindig 3 -ra tippelnék.

10. Nevezetes eloszlások várható értékei – bizonyítások

10.1. Egyenletes eloszlás

$$\mathrm{E}(\ X\ )\ =\ \sum_x x\ p(x)\ =\ \sum_{x=A}^{B} x\ \frac{1}{B-A+1} =$$

$$\frac{1}{B-A+1}\ \sum_{x=A}^{B} x\ =\ \frac{1}{B-A+1}\ (B-A+1)\ \frac{A+B}{2}\ =\ \frac{A+B}{2}$$

Mivel az eloszlás szimmetrikus az $\frac{a+b}{2}$ pontra, nyilvánvaló, hogy a várható érték $\frac{a+b}{2}$.

10.2. Hipergeometrikus eloszlás *(Extra tananyag)*

$$\mathrm{E}(\ X\ ) = \sum_x\ x\ p(x) =$$

$$\sum_{x=\max(0,n-B)}^{\min(n,A)}\ x\ \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}\ =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)}\ x\ \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}\ =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)}\ \frac{x\ \binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}\ =$$

$$\sum_{x=\max(1,n-B)}^{\min(n,A)}\ \frac{A\ \binom{A-1}{x-1}\binom{B}{n-x}}{\frac{A+B}{n}\binom{A-1+B}{n-1}}\ =$$

$$n\ \frac{A}{A+B}\ \sum_{x=\max(1,n-B)}^{\min(n,A)}\ \frac{\binom{A-1}{x-1}\binom{B}{n-x}}{\binom{A-1+B}{n-1}}\ =$$

$$n\ \frac{A}{A+B}\ \sum_{x=\max(0,n-1-B)}^{\min(n-1,A-1)}\ \frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}}\ =\ n\ \frac{A}{A+B}$$

A levezetés során $x-1$ et y -nal helyettesítettük, és ennek megfelelően $1+y$ helyére x -et irtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{x=\max(0,n-1-B)}^{\min(n-1,A-1)}\ \frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}}\ =\ 1$$

ami következik abból a tényből, hogy az $A-1, B, n-1$ paraméterű hipergeometrikus eloszlás súlyfüggvénye:

$$\frac{\binom{A-1}{y}\binom{B}{n-1-y}}{\binom{A-1+B}{n-1}}$$

$$(\ \max(\ 0\ ,\ n-1-B\ )\ \le\ x\ \le\ \min(\ n-1\ ,\ A-1\ )\ )$$

10.3. Indikátor eloszlás

10.3.1. Heurisztikus levezetés

Jelentsen A egy p valószínűségű eseményt. Ha egyetlen kísérletet végzünk A -ra, akkor az X valószínűségi változó mutassa azt, hogy A hányszor következik be eme egyetlen kísérlet során:

- X értéke 1, ha A bekövetkezik
- X értéke 0, ha A nem bekövetkezik be

Ha erre az X valószínűségi változóra N kísérletet végzünk, akkor N darab kísérleti eredményt kapunk:

$$X_1, X_2, \ldots, X_N$$

Itt minden X_i vagy 1 vagy 0 , attól függően, hogy i -ik kísérlet során A bekövetkezik-e vagy sem. Egyrészt, mint a nagy számok törvénye alapján tudjuk, az $X_1, X_2, \ldots X_N$ kísérleti eredmények átlaga – nagy N esetén – közelíti X várható értékét:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathrm{E}(\, , X\,)$$

Másrészt a tört számlálójában álló $X_1 + X_2 + \ldots + X_N$ összeg i -ik tagja 1 vagy 0 attól függően, hogy az i -ik kísérlet során A bekövetkezik-e vagy sem, maga az összeg azt mutatja, hogy N kísérlet során hányszor következik be az A esemény. Ezért az összeget N-nel osztva a kapott tört – nagy N esetén – körülbelül egyenlő az A esemény valószínűségével:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx p$$

Mindezekből természetesnek tűnik, hogy

$$\mathrm{E}(\, X\,) = p$$

10.3.2. Bizonyítás

$$\mathrm{E}(\, X\,) = \sum_x x\, p(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

10.4. Binomiális eloszlás

10.4.1. Heurisztikus levezetés

Jelentsen A egy p valószínűségű eseményt. Képzeljük el, hogy n kísérletet végzünk A -ra. Az X valószínűségi változó mutassa azt, hogy A hányszor következik be az n kísérlet során. Ha erre az X valószínűségi változóra N kísérletet végzünk, akkor N kísérleti eredményt kapunk: $X_1, X_2, \ldots, X_N$. Egyrészt, mint a nagy számok törvénye alapján tudjuk, ezek átlaga – nagy N esetén – közelíti X várható értékét:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathrm{E}(\, X\,)$$

Másrészt a bal oldalon álló tört számlálójában álló $X_1 + X_2 + \ldots + X_N$ összeg minden tagja n kísérlet kapcsán mutatja, hogy az A esemény hányszor következik be, maga az összeg azt mutatja, hogy $N \cdot n$ kísérlet során hányszor következik be az A esemény. Ezért az összeget $N \cdot n$ -nel osztva a kapott tört – nagy N esetén – körülbelül egyenlő az A esemény valószínűségével:

$$\frac{X_1 + X_2 + \ldots + X_N}{N \cdot n} \approx p$$

A közelítő egyenlőség mindkét oldalát n-nel szorozva kapjuk, hogy

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx n \cdot p$$

Mindezekből természetesnek tűnik, hogy

$$\mathrm{E}(X) = n \cdot p$$

10.4.2. Bizonyítás

$$\mathrm{E}(X) = \sum_x x\, p(x) =$$

$$\sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$\sum_{x=1}^{n} n \binom{n-1}{x-1}\, p\, p^{x-1}\, (1-p)^{n-x} =$$

$$np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} =$$

$$np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = np$$

A levezetés során $x-1$ et y -nal helyettesítettük, és ennek megfelelően $1+y$ helyére x -et írtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = 1$$

ami következik abból a tényből, hogy az $n-1$ és p paraméterű binomiális eloszlás súlyfüggvénye:

$$\binom{n-1}{y} p^y (1-p)^{n-1-y} \quad \text{if} y = 0, 1, 2, \ldots, n$$

10.5. Geometriai eloszlás (optimista)

10.5.1. Heurisztikus levezetés

Jelentsen A egy p valószínűségű eseményt. Képzeljük el, hogy addig végzünk kísérleteket A -ra, amíg A esemény végre bekövetkezik. Az X valószínűségi változó mutassa azt, hogy hány kísérlet kell ehhez. Ha erre az X valószínűségi változóra N kísérletet végzünk, akkor N kísérleti eredményt kapunk: $X_1, X_2, \ldots, X_N$. Egyrészt, mint a nagy számok törvénye alapján tudjuk, ezek átlaga – nagy N esetén – közelíti X várható értékét:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathrm{E}(\, X\,)$$

Másrészt a számlálóban álló $X_1+X_2+\ldots+X_N$ összeg azt mutatja, hogy hány kísérlet kell az A esemény N -edik bekövetkezéséhez, N osztva az összeggel – nagy N esetén – körülbelül egyenlő az A esemény valószínűségével:

$$\frac{N}{X_1 + X_2 + \ldots + X_N} \approx p$$

vagyis

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \frac{1}{p}$$

Mindezekből természetesnek tűnik, hogy

$$\mathrm{E}(\, X\,) = \frac{1}{p}$$

10.5.2. Bizonyítás

Két bizonyítást is adunk. Az első bizonyítás végtelen mértani sorok összegzésein alapul:

$$\mathrm{E}(\, X\,) = \sum_x x\, p(x) =$$

$$
\begin{array}{cccccccccc}
p & + & 2pq & + & 3pq^2 & + & 4pq^3 & + & \dots & = \\
\\
p & + & pq & + & pq^2 & + & pq^3 & + & \dots & \\
 & + & pq & + & pq^2 & + & pq^3 & + & \dots & \\
 & & & + & pq^2 & + & pq^3 & + & \dots & \\
 & & & & & + & pq^3 & + & \dots & \\
 & & & & & & & \ddots & & \\
 & & & & & & & & & = \\
\frac{p}{1-q} & & & & & & & & & \\
 & + & \frac{pq}{1-q} & & & & & & & \\
 & & & + & \frac{pq^2}{1-q} & & & & & \\
 & & & & & + & \frac{pq^3}{1-q} & & & \\
 & & & & & & & \ddots & & \\
 & & & & & & & & & = \\
1 & & & & & & & & & \\
 & + & q & & & & & & & \\
 & & & + & q^2 & & & & & \\
 & & & & & + & q^3 & & & \\
 & & & & & & & \ddots & & \\
 & & & & & & & & & =
\end{array}
$$

$$\frac{1}{1-q} = \frac{1}{p}$$

A második bizonyítás a hatványsorok elméletén alapszik. A $q = 1 - p$ jelöléssel élve ezt kapjuk:

$$\mathrm{E}(\ X\ ) = \sum_x \ x\ p(x) \ =$$

$$1\ p + 2\ p(1-p) + 3\ p(1-p)^2 + 4\ p(1-p)^3 + \dots \ =$$

$$1\ p + 2\ pq + 3\ pq^2 + 4\ pq^3 + \dots \ =$$

$$p\ \left(1 + 2q + 3q^2 + 4q^3 + \dots\right) \ = \ p\ \frac{1}{(1-q)^2} \ = \ p\ \frac{1}{p^2} \ = \ \frac{1}{p}$$

Felhasználtuk a

$$1 + 2q + 3q^2 + 4q^3 + \dots \ = \ \frac{1}{(1-q)^2}$$

azonosságot, amit az alábbiakban le is vezetünk úgy, hogy először felismerjük, hogy az adott sor egy mértani sor deriváltjaként fogható fel, aztán vesszük a végtelen mértani sor összegképletét, és végül azt deriváljuk:

$$1 + 2q + 3q^2 + 4q^3 + \dots \ =$$

$$\frac{d}{dq}\left(1 + q + q^2 + q^3 + q^4 \dots\right) \ =$$

$$\frac{d}{dq}\left(\frac{1}{1-q}\right) \ = \ \frac{d}{dq}\left((1-q)^{-1}\right) \ = \ (1-q)^{-2} \ = \ \frac{1}{(1-q)^2}$$

10.6. Geometriai eloszlás (pesszimista)

Mivel a pesszimista geometriai eloszlás az optimistából úgy adódik, hogy azt 1 egységgel balra toljuk, a pesszimista geometriai eloszlás várható értéke nem más, mint az optimista geometriai eloszlás várható értéke mínusz 1 .

$$\mathrm{E}(\ X\ ) = \frac{1}{p} - 1$$

10.7. Negatív binomiális eloszlás (optimista)

10.7.1. Heurisztikus levezetés

Jelentsen A egy p valószínűségű eseményt. Képzeljük el, hogy addig végzünk kísérleteket A -ra, amíg A esemény végre n -edszer bekövetkezik. Az X valószínűségi változó mutassa azt, hogy hány kísérlet kell ehhez. Ha erre az X valószínűségi változóra N kísérletet végzünk, akkor N kísérleti eredményt kapunk: $X_1, X_2, \ldots, X_N$. Egyrészt, mint a nagy számok törvénye alapján tudjuk, ezek átlaga – nagy N esetén – közelíti X várható értékét:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathrm{E}(\ X\ )$$

Másrészt a számlálóban álló $X_1 + X_2 + \ldots + X_N$ összeg azt mutatja, hogy hány kísérlet kell az A esemény $N \cdot n$ -edik bekövetkezéséhez, $N \cdot n$ osztva az összeggel – nagy N esetén – körülbelül egyenlő az A esemény valószínűségével:

$$\frac{N \cdot n}{X_1 + X_2 + \ldots + X_N} \approx p$$

vagyis

$$\frac{X_1 + X_2 + \ldots + X_N}{N \cdot n} \approx \frac{1}{p}$$

A közelítő egyenlőség mindkét oldalát n -nel szorozva azt kapjuk, hogy

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \frac{n}{p}$$

Mindezekből természetesnek tűnik, hogy

$$\mathrm{E}(\ X\ ) = \frac{n}{p}$$

10.7.2. Bizonyítás *(Extra tananyag)*

$$\mathrm{E}(\ X\ ) = \sum_x x\, p(x) =$$

$$= \sum_{x=n}^{\infty} x \binom{x-1}{n-1} p^n (1-p)^{x-n} =$$

$$= \sum_{x=n}^{\infty} n \binom{x}{n} \frac{p^{n+1}}{p} (1-p)^{x-n} =$$

$$= \frac{n}{p} \sum_{x=n}^{\infty} \binom{x}{n} p^{n+1} (1-p)^{x-n} =$$

(itt az $y = x + 1, \; x = y - 1$ helyettesítést alkalmazzuk)

$$= \frac{n}{p} \sum_{y=n+1}^{\infty} \binom{y-1}{n} p^{n+1} (1-p)^{y-1-n} \; =$$

$$= \frac{n}{p}$$

(Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{y=n+1}^{\infty} \binom{y-1}{n} p^{1+n} (1-p)^{y-1-n} \; = \; 1$$

ami azért igaz, mert a szumma az $n + 1$, p paraméterű optimista negatív binomiális eloszlás tagjait összegzi.)

10.8. Negatív binomiális eloszlás (pesszimista) *(Extra tananyag)*

Mivel a pesszimista negatív binomiális eloszlás az optimistából úgy adódik, hogy azt n egységgel balra toljuk, a pesszimista negatív binomiális eloszlás várható értéke nem más, mint az optimista geometriai eloszlás várható értéke mínusz n

$$\mathrm{E}(\; X \;) \; = \; \frac{r}{p} - n$$

10.9. Poisson eloszlás

$$\mathrm{E}(\; X \;) = \sum_{x} \; x \, p(x) \; =$$

$$\sum_{x=0}^{\infty} x \; \frac{\lambda^x}{x!} e^{-\lambda} \; =$$

$$\sum_{x=1}^{\infty} x \; \frac{\lambda^x}{x!} e^{-\lambda} \; =$$

$$\sum_{x=1}^{\infty} \lambda \; \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} \; =$$

$$\lambda \; \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} \; =$$

$$\lambda \; \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!} e^{-\lambda} \; = \; \lambda$$

A levezetés során $x - 1$ et y -nal helyettesítettük, és ennek megfelelően $1 + y$ helyére x-et írtunk. Az utolsó lépésnél felhasználtuk, hogy

$$\sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!} e^{-\lambda} \; = \; 1$$

ami következik abból a tényből, hogy a λ paraméterű Poisson eloszlás súlyfüggénye

$$\frac{\lambda^{y}}{y!} e^{-\lambda} \quad (\; y = 0, 1, 2, \dots \;)$$

11. Binomiáls eloszlással kapcsolatos levezetések

11.1. Második momentum

Állítás:

$$\mathrm{E}(\ X^2\ )\ =\ np\ +\ n^2p^2\ -\ np^2$$

Bizonyítás *(Extra tananyag)*:

$$\mathrm{E}(\ X^2\ ) = \sum_x\ x^2\ p(x)\ =$$

$$\sum_{x=0}^{n}\ x^2\ \binom{n}{x} p^x(1-p)^{n-x}\ =$$

$$\sum_{x=1}^{n}\ x^2\ \binom{n}{x} p^x(1-p)^{n-x}\ =$$

$$\sum_{x=1}^{n}\ n\ x\ \binom{n-1}{x-1}\ p^x\ (1-p)^{n-x}\ =$$

$$np\ \sum_{x=1}^{n}\ x\ \binom{n-1}{x-1} p^{x-1}(1-p)^{n-x}\ =$$

Most $(x-1\ )$-et y -nal helyettesítjük, és így $(1+y)$-t írunk x helyére:

$$np\ \sum_{y=0}^{n-1}\ (1+y)\ \binom{n-1}{y} p^y(1-p)^{n-1-y}\ =$$

Az összeg két összegre bontható:

$$np\ \left[\left(\sum_{y=0}^{n-1}\ 1\ \ \binom{n-1}{y} p^y(1-p)^{n-1-y} \right) + \left(\sum_{y=0}^{n-1}\ y\ \ \binom{n-1}{y} p^y(1-p)^{n-1-y} \right) \right] =$$

Itt a szögletes zárójal belsejében álló első szumma az $n-1$ és p paraméterű binomiális eloszlás tagjainak az összege, ami 1. A második szumma az $n-1$ és p paraméterű binomiláis eloszlás várható értéke, ami $(n-1)p$. Ezt kapjuk:

$$np\ \ [\ 1\ +\ (n-1)\ p\ ]\ =\ np\ +\ n^2p^2\ -\ np^2$$

11.2. Variancia és szórás

Állítás:

$$\mathrm{VAR}(\ X\ )\ =\ np(1-p)$$

$$\mathrm{SD}(\ X\ )\ =\ \sqrt{np(1-p)}$$

Bizonyítás. *(Extra tananyag.)* Mint már tudjuk, a binomiális eloszlás második momentuma

$$\mathrm{E}(\ X^2\ )\ =\ n^2p^2 - np^2 + np$$

várható értéke

$$\mathrm{E}(\ X\ )\ =\ np$$

Ezért a varianciája

$$\mathrm{VAR}(\ X\ )\ =\ \left(n^2p^2 - np^2 + np\right) - (np)^2\ =\ np - np^2 = np(1-p)$$

szórása pedig

$$\mathrm{SD}(\ X\ )\ =\ \sqrt{np(1-p)}$$

12. Feltételes várható érték, variancia, szórás

Először a feltételes várható érték fogalmával ismerkedünk meg, azután jön a többi feltételes "izé".

12.1. Feltételes várható érték

Ha X és Y diszkrét valószínűségi változók kapcsán korábban bevezettük a $p_{2|1}(y|x)$, $p_{1|2}(x|y)$ feltételes súlyfüggvényeket (feltételes eloszlásokat). Most a feltételes súlyfüggvényekre támaszkodva bevezetjük először a feltételes várható érték fogalmát.

A feltételes súlyfüggvényből számított

$$\sum_y y \, p_{2|1}(y|x)$$

várható értéket **feltételes várható érték**nek nevezzük. Az $X = x$ feltétel mellett az Y feltételes várható értékét $\mathrm{E}(Y|X = x)$ -szel jelöljük. Tehát

$$\mathrm{E}(\ Y \mid X = x\ ) = \sum_y y \, p_{2|1}(y|x)$$

Az $X = x$ feltétel mellett az Y feltételes várható értéke azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények átlaga.

Hasonlóképpen az $Y = y$ feltétel mellett az X feltételes várható értékét $\mathrm{E}(\ X \mid Y = y\ )$ -nal jelöljük. Nyilván

$$\mathrm{E}(\ X \mid Y = y\ ) \ = \ \sum_x x \, p_{1|2}(x|y)$$

Az $Y = y$ feltétel mellett az X feltételes várható értéke azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X értékek átlaga.

12.2. Feltételes variancia

A feltételes súlyfüggvényből számított varianciát **feltételes varianciá**nak nevezzük.

Az $X = x$ feltétel mellett az Y feltételes varianciája azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y -ra kapott kísérleti eredmények varianciája.

Az $Y = y$ feltétel mellett az X feltételes varianciája azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X -re kapott kísérleti eredmények varianciája.

12.3. Feltételes szórás

A feltételes súlyfüggvényből számított szórást **feltételes szórás**nak nevezzük.

Az $X = x$ feltétel mellett az Y feltételes szórása azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Yv-ra kapott kísérleti eredmények szórása.

Az $Y = y$ feltétel mellett az X feltételes varianciája azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $Y = y$, akkor az érvényes kísérletekben körülbelül mennyi az X -re kapott kísérleti eredmények szórása.

Megjegyzés: A feltételes súlyfüggvényből számított "izé"-t **feltételes "izé"** -nek nevezzük. Az $X = x$ feltétel mellett az Y feltételes "izé" -je azt fejezi ki, hogy sok kísérlet kapcsán, ha csak azokat a kísérleteket fogadjuk el érvényesnek, amikor $X = x$, akkor az érvényes kísérletekben körülbelül mennyi az Y-ra kapott kísérleti eredmények "izé" -je.

12.4. Példák: Ha tudjuk, mennyi az egyik, akkor mennyi a másiknak az "izé" -je?

A példák azzal a feladattal kapcsolatosak, melyet az 1. fejezet "Kombinatorikus alapképletek" című 8. pontjában vettünk: Egy dobozban 45 darab golyó van, melyek közül 10 piros, 15 kék, 20 fehér. Kiveszünk 8 golyót visszatevés nélkül, és megfigyeljük, hogy a kihúzottak között hány piros és hány kék lesz. Az így kapott X és Y valószínűségi változókból adódó (X, Y) kétdimenziós valószínűségi változó $p(x, y)$ súlyfüggvényének táblázatát korábban meghatároztuk. Most csak a feltételes eloszlások táblázataival foglalkozunk.

1. Példa: Ha tudjuk, hány piros, akkor mi a kékek számának várható értéke, varianciája, szórása? A feltételes eloszlások táblázatában minden oszlop aljára odaírtuk a kékek számának feltételes várható értékét, feltételes varianciáját és feltételes szórását:

y x	0	1	2	3	4	5	6	7	8
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.040	0.015	0.003	0.000	0.000	0.000	0.000	0.000	0.000
5	0.145	0.085	0.037	0.009	0.000	0.000	0.000	0.000	0.000
4	0.281	0.231	0.160	0.084	0.026	0.000	0.000	0.000	0.000
3	0.300	0.328	0.320	0.266	0.174	0.070	0.000	0.000	0.000
2	0.173	0.242	0.313	0.369	0.381	0.321	0.176	0.000	0.000
1	0.049	0.086	0.143	0.224	0.327	0.435	0.504	0.429	0.000
0	0.005	0.012	0.024	0.048	0.093	0.174	0.319	0.571	1.000
F.v.é.	3.43	3.00	2.57	2.14	1.71	1.29	0.86	0.43	0.00
F.var.	1.56	1.41	1.25	1.08	0.89	0.69	0.48	0.24	0.00
F.sz.	1.25	1.19	1.12	1.04	0.95	0.83	0.69	0.49	0.00

Táblázat: X *adott értéke mellett*
Y **feltételes** *eloszlása, várható értéke, varianciája, szórása*

2. Példa: Ha tudjuk, hány kék, akkor mi a pirosak számának várható értéke, varianciája, szórása? A feltételes eloszlások táblázatában minden sor jobb szélére odaírtuk a pirosak számának feltételes várható értékét, feltételes varianciáját és feltételes szórását:

y x	0	1	2	3	4	5	6	7	8	F.v.é.	F.var.	F.sz.
8	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0,00
7	0.667	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.33	0.22	0.47
6	0.437	0.460	0.103	0.000	0.000	0.000	0.000	0.000	0.000	0.67	0.43	0.66
5	0.281	0.468	0.222	0.030	0.000	0.000	0.000	0.000	0.000	0.00	0.62	0.79
4	0.177	0.416	0.312	0.088	0.008	0.000	0.000	0.000	0.000	1.33	0.80	0.89
3	0.109	0.340	0.360	0.160	0.029	0.002	0.000	0.000	0.000	1.67	0.96	0.98
2	0.065	0.261	0.367	0.230	0.067	0.008	0.000	0.000	0.000	2.33	1.10	1.05
1	0.038	0.190	0.343	0.286	0.118	0.024	0.002	0.000	0.000	2.33	1.23	1.11
0	0.022	0.132	0.298	0.318	0.174	0.049	0.007	0.000	0.000	2.67	1.35	1.16

Táblázat: Y *adott értéke mellett*
X ***feltételes*** *eloszlása, várható értéke, varianciája, szórása*