Vetier András

Valószínűségszámítás

2. rész

Vetier András

Valószínűségszámítás

2. rész

Nevezetes diszkrét eloszlások

TYPOTEX

Magyarország Kormánya
Európai Unió
Európai Regionális Fejlesztési Alap
BEFEKTETÉS A JÖVŐBE
SZÉCHENYI 2020

A mű elektronikus kiadása
a VEKOP-2.1.1-15-2016-00152 sz.
projekt keretén belül készült.

ISBN 978 963 493 033 4

Kedves Olvasó!
Köszönjük, hogy kínálatunkból választott olvasnivalót!
Újabb kiadványainkról és akcióinkról a *www.typotex.hu*
és a *facebook.com/typotexkiado* oldalakon értesülhet.

Typotex Kiadó
Alapította Votisky Zsuzsa, 1989
A kiadó az 1795-ben alapított Magyar Könyvkiadók
és Könyvterjesztők Egyesülésének tagja.
Felelős kiadó: Németh Kinga
Főszerkesztő: Horváth Balázs
A kötetet gondozta: Erő Zsuzsa
A borítót készítette: Szalay Éva

Tartalom

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 ír oda,
- 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. április 17.

Üdvözlettel,

Vetier András

1. fejezet

Nevezetes eloszlások

1.1. Egyenletes eloszlások

1.1.1. Egyenletes eloszlás egy dimenzióban

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

Súlyfüggvény:

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

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 két dimenzióban

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

Súlyfüggvény:

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

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

$$X = \text{ a piros kockán felül lévő szám,}$$
$$Y = \text{ a zöld kockán felül lévő szám,}$$
$$p(x, y) = \frac{1}{36} \qquad (x = 1, 2, 3, 4, 5, 6;\; y = 1, 2, 3, 4, 5, 6).$$

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:

$$p(x_1, x_2, \ldots x_r) = \frac{1}{n} \quad \text{minden lehetséges } (x_1, x_2, \ldots x_r) \text{ pontra.}$$

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

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

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

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

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

1.2. Hipergeometrikus eloszlások

1.2.1. Hipergeometrikus eloszlás

Súlyfüggvény:

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

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:

$$n = \text{a húzások száma,}$$

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

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

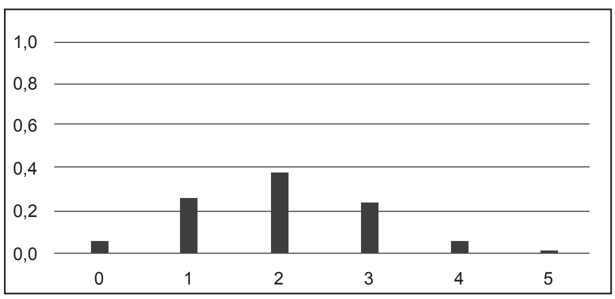
Ha valamiért fontos, akkor az eloszlás neve előtt a paraméterek jelét vagy numerikus értékeiket az itt megadott sorrendben adjuk meg.

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 $\binom{K}{x}$-féleképpen 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:

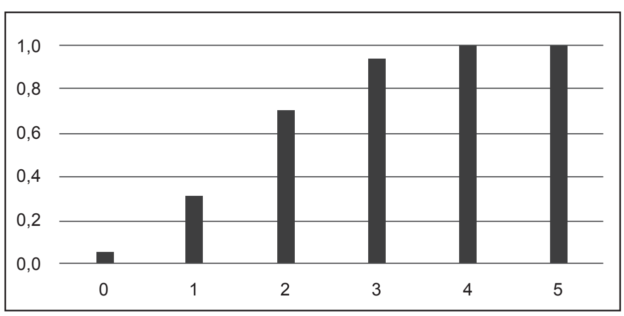
$$\binom{K}{x}\binom{N-K}{n-x}.$$

Ezért a keresett súlyfüggvényérték:

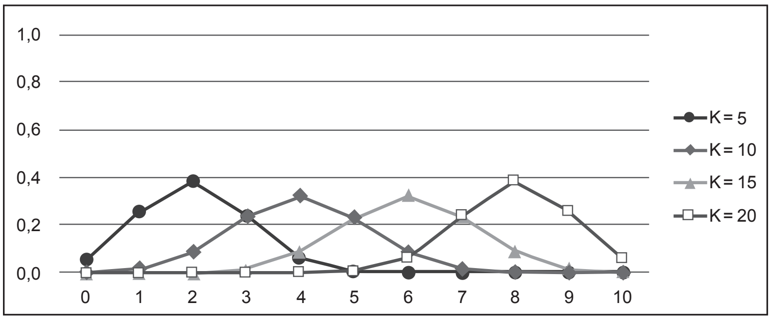
$$p(x) = \mathrm{P}\,(\,X = x\,) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}.$$



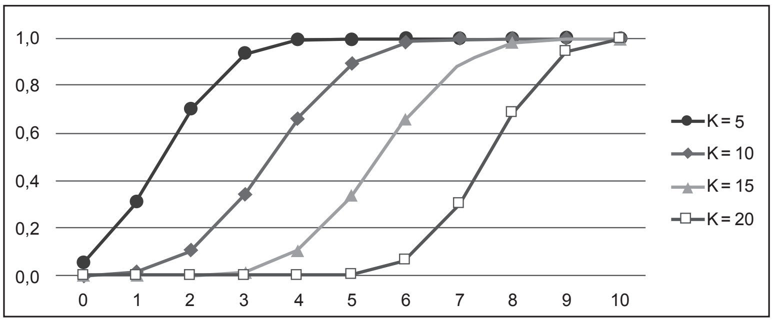
1.1. ábra. *Hipergeometrikus eloszlás súlyfüggvénye;* $n = 10;\; K = 5;\; N = 25$



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



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



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

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

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

azaz

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

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 az 1.1. táblázatba foglaltuk:

N	$\mathrm{P}(15 \text{ piros})$
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

1.1. táblázat. *A* $\mathrm{P}(15 \textit{ piros})$ *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 240 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? Városunkban szeretnék egy felnőtt bridzsklubot 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 felnőttek 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 $\mathrm{P}(\,42 \text{ igen}\,)$ valószínűséget, és az értékeket az 1.2. táblázatba foglaltuk:

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.

K	$\mathrm{P}(42 \text{ igen})$
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

1.2. táblázat. *A* $\mathrm{P}(42$ *igen*$)$ *valószínűség* K *függvényében*

Megjegyzés: Ha a felnőtt emberek közül történő véletlen választás módszere nem garantálja, hogy minden embert legfeljebb egyszer kérdezünk meg, tehát ismétlődés is előforulhat, akkor a $\mathrm{P}(42 \text{ igen})$ 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 az 1.2. 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

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

$$0 \le y \le \min(n, K_2),$$

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

Mikor hasz+náljuk: 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

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

$$Y = \text{ahányszor zöld golyót húzunk},$$

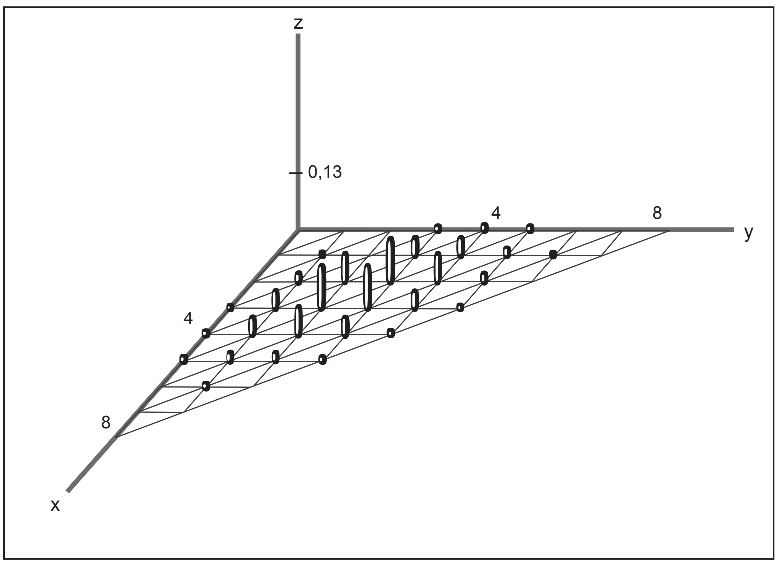
ilyen eloszlást követ.

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 súlyfüggvényérték:

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



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