

Nevezetes diszkrét eloszlások (5. előadás)

$X : \Omega \rightarrow \mathbb{R}$ diszkrét valószínűségi változó eloszlása: a lehetséges értékei és a hozzájuk tartozó valószínűségek.

- binomiális eloszlás: n független kísérlet, mindegyik p valószínűséggel sikerül, X : hány sikerült. $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, ahol $k = 0, 1, 2, \dots, n$.
- Poisson-eloszlás: az X valószínűségi változó Poisson-eloszlású λ paraméterrel, ha

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

- hipergeometriai eloszlás: visszatevés nélküli mintavételnél a húzott fekete golyók száma; $\mathbb{P}(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$.
- geometriai eloszlás: minden kísérlet p valószínűséggel sikerül; Y : hányadik az első sikeres. $\mathbb{P}(Y = k) = (1 - p)^{k-1} p$, ahol $k = 1, 2, \dots$

A binomiális eloszlás közelítése

Tegyük fel, hogy egy biztosító $n = 100000$ ügyfelének mindegyike egy év alatt egymástól függetlenül $p = 0,0001$ valószínűséggel okoz balesetet. A balesetet okozó ügyfelek számának (ezt jelöljük X -szel) **várható értéke**:

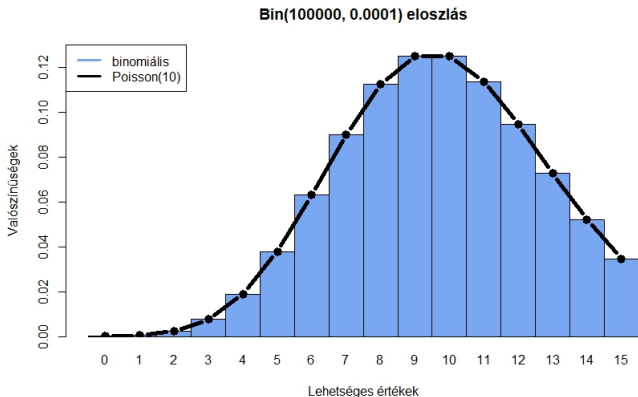
$$\mathbb{E}(X) = np = 100000 \cdot 0,0001 = \mathbf{10}.$$

Annak valószínűsége, hogy **pontosan k ügyfél okoz balesetet**:

$$\begin{aligned}\mathbb{P}(X = k) &= \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} = \\ &= \frac{100000 \cdot 99999 \cdot \dots \cdot (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx \\ &\approx \frac{100000^k \cdot 0,0001^k}{k!} \left(1 - \frac{10}{100000}\right)^{100000} \approx \frac{10^k}{k!} e^{-10},\end{aligned}$$

felhasználva, hogy $\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$ tetszőleges $x > 0$ -ra.

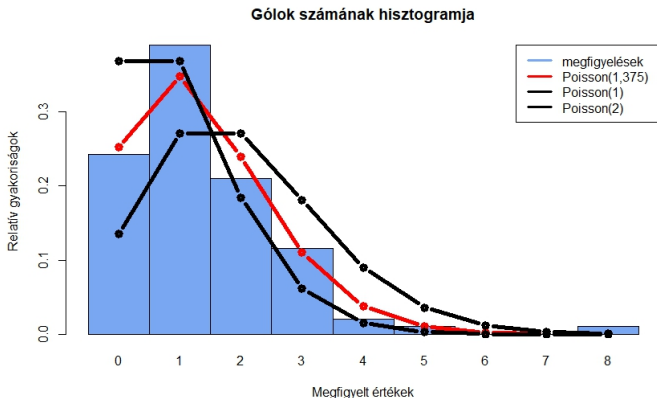
A binomiális eloszlás közelítése



Binomiális eloszlás $n = 100000$ renddel és $p = 0,0001$ paraméterrel (vízszintes tengely: k , oszlopok magassága: $\mathbb{P}(X = k)$).

Feketével a $\frac{10^k}{k!} e^{-10}$ függvény (ez lesz a Poisson(10)-eloszlás).

Poisson-eloszlás



A gólok számának hisztogramja $n = 95$ mérkőzésen, és különböző paraméterű Poisson-eloszlások

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt:**
 - ▶ lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
 - ▶ a balesetek száma egy városban egy hét vagy egy hónap alatt;
 - ▶ a földrengések száma egy év alatt
- egy rendszerbe beérkező igények száma egy adott időszakban:
 - ▶ egy üzletbe beérkező vásárlók száma egy óra alatt
 - ▶ egy weboldal letöltéseinek száma egy óra alatt

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt:**
 - ▶ lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
 - ▶ a balesetek száma egy városban egy hét vagy egy hónap alatt;
 - ▶ a földrengések száma egy év alatt
- egy rendszerbe beérkező igények száma egy adott időszakban:
 - ▶ egy üzletbe beérkező vásárlók száma egy óra alatt
 - ▶ egy weboldal letöltéseinek száma egy óra alatt
- **általában:** véletlen időközönként bekövetkező események száma adott időszak alatt

Poisson-folyamat: a t idő alatt bekövetkező események száma Poisson-eloszlású ct paraméterrel.

Ekkor a t idő alatt bekövetkező események számának várható értéke:

A Poisson-eloszlás alkalmazásai

- **ritkán bekövetkező események száma adott időszak alatt:**
 - ▶ lórúgás halálos áldozatainak száma a porosz hadseregben (ez volt az első statisztikai példa)
 - ▶ a balesetek száma egy városban egy hét vagy egy hónap alatt;
 - ▶ a földrengések száma egy év alatt
- egy rendszerbe beérkező igények száma egy adott időszakban:
 - ▶ egy üzletbe beérkező vásárlók száma egy óra alatt
 - ▶ egy weboldal letöltéseinek száma egy óra alatt
- **általában:** véletlen időközönként bekövetkező események száma adott időszak alatt

Poisson-folyamat: a t idő alatt bekövetkező események száma Poisson-eloszlású ct paraméterrel.

Ekkor a t idő alatt bekövetkező események számának várható értéke: ct , azaz az intervallum hosszával arányos.

Programozós házi feladat október 9., 10:15-ig (3 pont)

- Tegyük fel, hogy egy város n lakosa közül mindenki egymástól függetlenül p valószínűséggel támogat egy adott pártot. Válasszunk tetszőleges $n \geq 1000$ és $0,1 < p < 0,9$ számokat. Milyen eloszlású a pártot támogatók száma? Mennyi a várható értéke, mennyi a szórása?
- Sorsoljunk 500 elemű mintát a támogatók számának eloszlásából. Készítsünk a kapott adatsorból hisztogramot.
- Határozzuk meg a kapott adatsor átlagát (`mean(...)`) és (korrigált tapasztalati) szórását (`sd(...)`). Hasonlítsuk össze a kapott értékeket az első kérdésben kiszámolt várható értékkel és szórással.

Házi feladat október 9., 10:15-ig, megoldás

Tegyük fel, hogy egy város n lakosa közül mindenki egymástól függetlenül p valószínűséggel támogat egy adott pártot. Válasszunk tetszőleges $n \geq 1000$ és $0,1 < p < 0,9$ számokat. Milyen eloszlású a pártot támogatók száma? Mennyi a várható értéke, mennyi a szórása?

A pártot támogatók száma **binomiális eloszlású** n renddel és p paraméterrel.

Legyen $n = 8000$ és $p = 0,36$. Ekkor az eloszlás várható értéke:

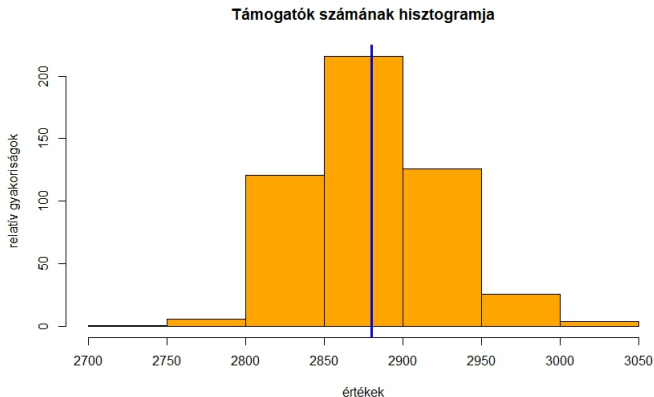
$$\mathbb{E}(X) = np = 8000 \cdot 0,36 = 2880.$$

A szórása:

$$D(X) = \sqrt{np(1-p)} = \sqrt{8000 \cdot 0,36 \cdot 0,64} = 42,93.$$

```
> minta<-rbinom(500, size=8000, prob=0.36)           500 elemű minta  
> hist(minta, col="orange", main="Támogatók számának hisztogramja",  
xlab="értékek", ylab="relatív gyakoriságok")         hisztogram  
> lines(abline(v=2880, lwd="3", col="blue"))         várható érték
```

Házi feladat október 9., 10:15-ig: megoldás



500 elemű minta histogramja $n = 8000$ rendű és $p = 0,36$ paraméterű binomiális eloszlásból, és a várható érték: $np = 2880$.

Házi feladat október 9., 10:15-ig, megoldás

Tegyük fel, hogy egy város n lakosa közül mindenki egymástól függetlenül p valószínűséggel támogat egy adott pártot. Válasszunk tetszőleges $n \geq 1000$ és $0,1 < p < 0,9$ számokat. Milyen eloszlású a pártot támogatók száma? Mennyi a várható értéke, mennyi a szórása?

A pártot támogatók száma **binomiális eloszlású** n renddel és p paraméterrel.

Legyen $n = 8000$ és $p = 0,36$. Ekkor az eloszlás várható értéke:

$$\mathbb{E}(X) = np = 8000 \cdot 0,36 = 2880.$$

A szórása:

$$D(X) = \sqrt{np(1-p)} = \sqrt{8000 \cdot 0,36 \cdot 0,64} = 42,93.$$

```
> minta<-rbinom(500, size=8000, prob=0.36)           500 elemű minta
```

```
> mean(minta)           átlag
```

```
[1] 2881.032
```

```
> sd(minta)           korrigált tapasztalati szórás
```

```
[1] 42.95935
```

Számolás házi feladat október 9., 10:15-ig (3 pont)

Tegyük fel, hogy ha az A és B focicsapat egymás ellen játszik, akkor az A csapat által rúgott gólok száma Poisson-eloszlású $1,2$ várható értékkel, a B csapat által rúgott gólok száma Poisson-eloszlású $1,8$ várható értékkel, továbbá ez a két valószínűségi változó egymástól független.

- Az A – B labdarúgómérkőzésen mennyi az összes gól számának várható értéke?
- Feltéve, hogy egy mérkőzésen összesen 5 gól születik, mennyi a valószínűsége, hogy ebből k darabot szerzett az A csapat? Itt $k = 0, 1, 2, 3, 4, 5$ lehet.
- Ábrázoljuk a kapott feltételes valószínűségeket k függvényében (kézzel készített ábra is megfelelő).

Házi feladat október 9., 10:15-ig

Tegyük fel, hogy ha az A és B focicsapat egymás ellen játszik, akkor az A csapat által rúgott gólok száma Poisson-eloszlású $1,2$ várható értékkel, a B csapat által rúgott gólok száma Poisson-eloszlású $1,8$ várható értékkel, továbbá ez a két valószínűségi változó egymástól független.

Legyen X az A csapat, Y a B csapat által rúgott gólok száma. Ekkor

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 1,2 + 1,8 = 3,$$

hiszen összeg várható értéke a várható értékek összege, a Poisson-eloszlás várható értéke pedig a paramétere.

A feltételes valószínűség definíciója alapján:

$$\mathbb{P}(X = k | X + Y = 5) = \frac{\mathbb{P}(X = k, X + Y = 5)}{\mathbb{P}(X + Y = 5)} = \frac{\mathbb{P}(X = k)\mathbb{P}(Y = 5 - k)}{\mathbb{P}(X + Y = 5)}.$$

Házi feladat október 9., 10:15-ig

$$\mathbb{P}(X = k)\mathbb{P}(Y = 5 - k) = \frac{1,2^k}{k!} \cdot e^{-1,2} \cdot \frac{1,8^{5-k}}{(5-k)!} \cdot e^{-1,8} = \frac{1}{5!} \cdot e^{-3} \cdot \binom{5}{k} 1,2^k \cdot 1,8^{5-k}.$$

Másrészt $\mathbb{P}(X + Y = 5)$ ezeknek az összege (hiszen minden lehetőséget felsoroltunk). A binomiális tétel alapján

$$\mathbb{P}(X + Y = 5) = \sum_{k=0}^5 \mathbb{P}(X = k)\mathbb{P}(Y = 5 - k) = \frac{1}{5!} \cdot e^{-3} \cdot \sum_{k=0}^5 \binom{5}{k} 1,2^k \cdot 1,8^{5-k} = \frac{3^5}{5!} e^{-3}.$$

Valójában azt láttuk be, hogy **az összesen rúgott gólok száma is Poisson-eloszlású**.

$$\begin{aligned} \mathbb{P}(X = k | X + Y = 5) &= \frac{\mathbb{P}(X = k)\mathbb{P}(Y = 5 - k)}{\mathbb{P}(X + Y = 5)} = \frac{\frac{1}{5!} \cdot e^{-3} \cdot \binom{5}{k} 1,2^k \cdot 1,8^{5-k}}{\frac{3^5}{5!} e^{-3}} = \\ &= \binom{5}{k} \cdot \left(\frac{1,2}{3}\right)^k \cdot \left(\frac{1,8}{3}\right)^{5-k} = \binom{5}{k} 0,4^k \cdot 0,6^{5-k}. \end{aligned}$$

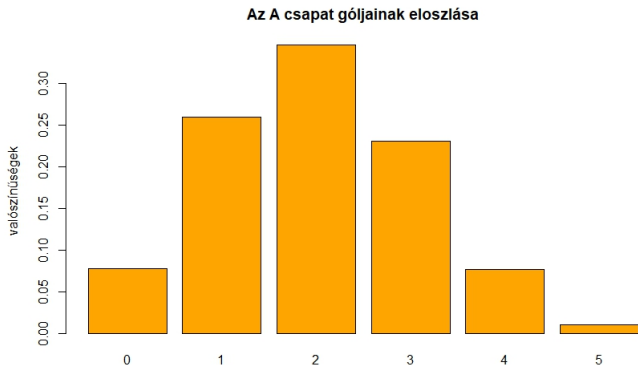
Házi feladat október 9., 10:15-ig

```
> x=0:5
```

```
> y<-dbinom(x, size=5, prob=0.4)
```

```
> names(y)<-c("0", "1", "2", "3", "4", "5")
```

```
> barplot(y, col="orange", main="Az A csapat góljainak eloszlása",  
xlab="értékek", ylab="valószínűségek")
```



Hipergeometriai eloszlás: példa

Egy sportcsapat $N = 20$ **tagja** közül $M = 9$ **balkezes**.

A pályán egyszerre $n = 7$ **különböző** játszik.

Tegyük fel, hogy minden hétfős összeállítás egyformán valószínű.

Milyen eloszlású a pályán a **balkezes játékosok száma**, X ?

Hipergeometriai eloszlás: példa

Egy sportcsapat $N = 20$ **tagja** közül $M = 9$ **balkezes**.

A pályán egyszerre $n = 7$ **különböző** játszik.

Tegyük fel, hogy minden hétfős összeállítás egyformán valószínű.

Milyen eloszlású a pályán a **balkezes játékosok száma, X ?**

A hétfős összeállítások száma:

k balkezes játékos kiválasztása:

$7 - k$ jobbkezes játékos kiválasztása:

A jó lehetőségek száma összesen:

Hipergeometriai eloszlás: példa

Egy sportcsapat $N = 20$ **tagja** közül $M = 9$ **balkezes**.

A pályán egyszerre $n = 7$ **különböző** játszik.

Tegyük fel, hogy minden hétfős összeállítás egyformán valószínű.

Milyen eloszlású a pályán a **balkezes játékosok száma, X** ?

A hétfős összeállítások száma: $\binom{20}{7}$

k balkezes játékos kiválasztása: $\binom{9}{k}$

$7 - k$ jobbkezes játékos kiválasztása: $\binom{11}{7-k}$

A jó lehetőségek száma összesen: $\binom{9}{k} \cdot \binom{11}{7-k}$ ←

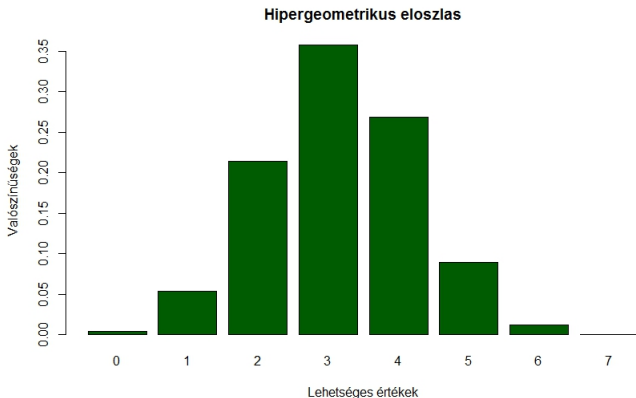
szorzás: bármely balkezes választás bármely jobbkezesessel jó

osztás: minden lehetőség egyformán valószínű

$$\mathbb{P}(X = k) = \frac{\binom{9}{k} \cdot \binom{11}{7-k}}{\binom{20}{7}}$$

visszatevés nélküli mintavétel

Hipergeometriai eloszlás: példa



A kiválasztott balkezes játékosok számának eloszlása hipergeometriai eloszlás, $N = 20$, $M = 9$, $n = 7$

vízszintes: k , oszlopok magassága: $\mathbb{P}(X = k) = \frac{\binom{9}{k} \cdot \binom{11}{7-k}}{\binom{20}{7}}$.

Hipergeometriai eloszlás

Legyenek N, M, n pozitív egészek úgy, hogy $1 \leq n \leq M \leq N$. Az X valószínűségi változó **hipergeometriai eloszlású**, ha

$$\mathbb{P}(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n).$$

- **visszatevés nélküli mintavételnél** a húzott fekete golyók száma: N golyó, ebből M fekete, n -szer húzunk visszatevés nélkül
- lottósorsolásnál a találatok száma, X , hipergeometrikus eloszlású $N = 90$, $M = 5$, $n = 5$ paraméterekkel:

$$\mathbb{P}(X = k) = \mathbb{P}(k \text{ találat}) = \frac{\binom{5}{k} \binom{85}{5-k}}{\binom{90}{5}} \quad k = 0, 1, 2, 3, 4, 5.$$

A hipergeometriai eloszlás várható értéke és szórása

Ha az X valószínűségi változó hipergeometriai eloszlású M, N, n paraméterekkel, azaz

$$\mathbb{P}(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n),$$

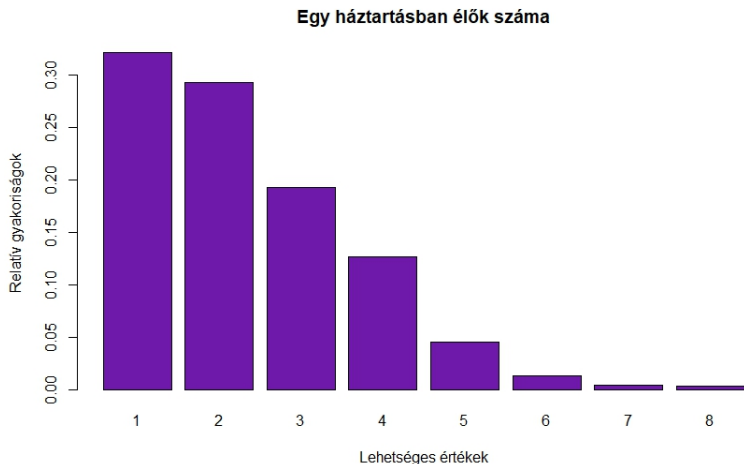
akkor

$$\mathbb{E}(X) = \frac{M}{N} n; \quad D(X) = \sqrt{n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}}.$$

Például, ha $N = 20$ játékos közül $M = 9$ balkezes, és $n = 7$ -et választunk visszatevés nélkül, akkor a balkezes játékosok számának, X -nek **várható értéke** és **szórása**:

$$\mathbb{E}(X) = \frac{9}{20} \cdot 7 = \mathbf{3,15}; \quad D(X) = \sqrt{7 \cdot \frac{9}{20} \cdot \left(1 - \frac{9}{20}\right) \cdot \frac{13}{19}} = \mathbf{1,09}.$$

Megfigyelések



Egy háztartásban élők számának hisztogramja (forrás: KSH, 2011)
($n = 4105698$ a háztartások száma)

Geometriai eloszlás

Egy közvéleménykutatásban mindenki a többiektől függetlenül $0,2$ valószínűséggel válaszol egy adott kérdésre. Jelölje Y , hogy hány embert kell megkérdezni, míg találunk egy válaszadót.

$$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$$

$$\mathbb{P}(Y = 2) = \mathbb{P}(\text{az első nem válaszol, a második igen}) =$$

Geometriai eloszlás

Egy közvéleménykutatásban mindenki a többiektől függetlenül $0,2$ valószínűséggel válaszol egy adott kérdésre. Jelölje Y , hogy hány embert kell megkérdezni, míg találunk egy válaszadót.

$$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$$

$$\mathbb{P}(Y = 2) = \mathbb{P}(\text{az első nem válaszol, a második igen}) = 0,8 \cdot 0,2;$$

$$\mathbb{P}(Y = 3) = \mathbb{P}(\text{az első kettő nem válaszol, a harmadik igen}) =$$

Geometriai eloszlás

Egy közvéleménykutatásban mindenki a többiektől függetlenül $0,2$ valószínűséggel válaszol egy adott kérdésre. Jelölje Y , hogy hány embert kell megkérdezni, míg találunk egy válaszadót.

$$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$$

$$\mathbb{P}(Y = 2) = \mathbb{P}(\text{az első nem válaszol, a második igen}) = 0,8 \cdot 0,2;$$

$$\mathbb{P}(Y = 3) = \mathbb{P}(\text{az első kettő nem válaszol, a harmadik igen}) = 0,8^2 \cdot 0,2;$$

$$\mathbb{P}(Y = k) = \mathbb{P}(\text{az első } k - 1 \text{ nem válaszol, a } k. \text{ igen}) =$$

Geometriai eloszlás

Egy közvéleménykutatásban mindenki a többiektől függetlenül $0,2$ valószínűséggel válaszol egy adott kérdésre. Jelölje Y , hogy hány embert kell megkérdezni, míg találunk egy válaszadót.

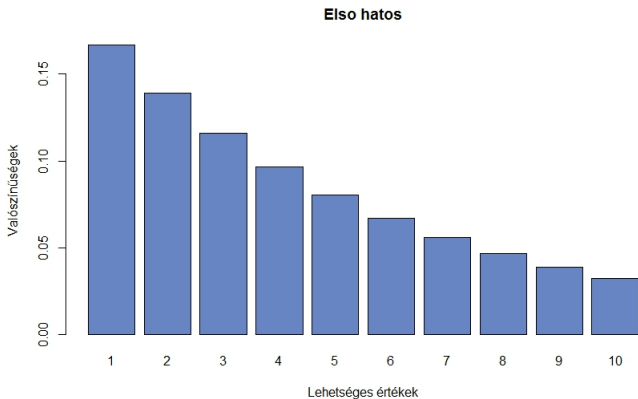
$$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$$

$$\mathbb{P}(Y = 2) = \mathbb{P}(\text{az első nem válaszol, a második igen}) = 0,8 \cdot 0,2;$$

$$\mathbb{P}(Y = 3) = \mathbb{P}(\text{az első kettő nem válaszol, a harmadik igen}) = 0,8^2 \cdot 0,2;$$

$$\mathbb{P}(Y = k) = \mathbb{P}(\text{az első } k - 1 \text{ nem válaszol, a } k. \text{ igen}) = 0,8^{k-1} \cdot 0,2.$$

Példa: geometriai eloszlás



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

Geometriai eloszlás

- független kísérleteket végzünk;
- mindegyik p valószínűséggel sikerül;
- Y : hányadik kísérlet az első sikeres.

Geometriai eloszlás

- független kísérleteket végzünk;
- mindegyik p valószínűséggel sikerül;
- Y : hányadik kísérlet az első sikeres.

Definíció

Az Y valószínűségi változó **geometriai eloszlású** p paraméterrel, ha lehetséges értékei:

$$1, 2, 3 \dots$$

és minden $1 \leq k$ egészre

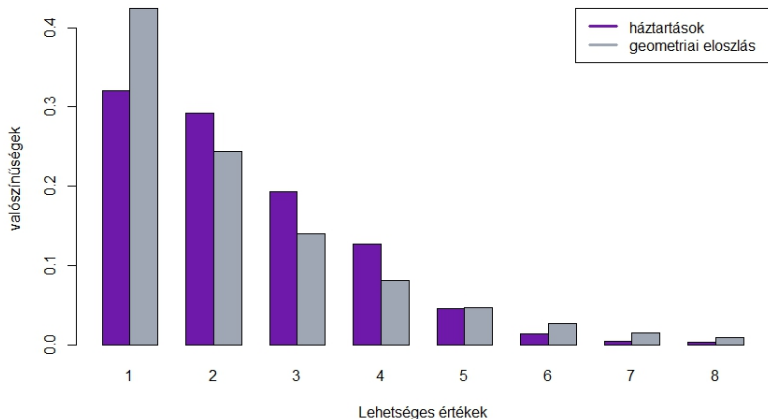
$$\mathbb{P}(Y = k) = (1 - p)^{k-1} p.$$

($0 < p < 1$.) Jelölés: $\text{Geo}(p)$. Másik elnevezés: Pascal-eloszlás.

Mivel $\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1$, ez valóban valószínűségeloszlás (annak valószínűsége, hogy sosem sikerül a kísérlet, 1).

Megfigyelések

Háztartások és geometriai eloszlás



Egy háztartásban élők száma és geometriai eloszlás $p = \frac{1}{X} = 0,423$ -vel
($n = 4105698$ a háztartások száma)

Geometriai eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó geometriai eloszlású p paraméterrel, azaz

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p \quad (k = 1, 2, \dots),$$

akkor

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

Geometriai eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó geometriai eloszlású p paraméterrel, azaz

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p \quad (k = 1, 2, \dots),$$

akkor

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

Példa. Tegyük fel, hogy egy adott pártot mindenki a többiektől függetlenül $p = 0,06$ valószínűséggel támogat. Jelölje X , hogy hány embert kell megkérdezni, míg az első olyan embert megtaláljuk, aki ezt a pártot támogatná. Ekkor

$$\mathbb{E}(X) = \frac{1}{0,06} = 16,67; \quad D(X) = \sqrt{\frac{0,94}{0,06^2}} = 16,16.$$

Nevezetes diszkrét eloszlások várható értéke és szórása

- Ha X binomiális eloszlású n renddel és p paraméterrel:

$$\mathbb{E}(X) = np; \quad D(X) = \sqrt{np(1-p)}.$$

- Ha X hipergeometriai eloszlású N, M, n paraméterekkel:

$$\mathbb{E}(X) = \frac{M}{N}n; \quad D(X) = \sqrt{n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}}.$$

- Ha X Poisson-eloszlású λ paraméterrel:

$$\mathbb{E}(X) = \lambda; \quad D(X) = \sqrt{\lambda};$$

- Ha X geometriai eloszlású p paraméterrel:

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

Valószínűségi vektorváltozó

Gyakran több mennyiséget egyszerre vizsgálunk, vagy ugyanazt többször megmérjük. Például: véletlenszerűen választott ember havi jövedelme, kiadása rezsire, illetve élelmiszerre; vagy egy helyen öt különböző műszerrel megmérjük a hőmérsékletet.

Definíció (Valószínűségi vektorváltozó)

Egy $\underline{X} : \Omega \rightarrow \mathbb{R}^n$ függvény valószínűségi vektorváltozó, ha tetszőleges t_1, t_2, \dots, t_n számokra

$$\{\omega \in \Omega : X_1(\omega) \leq t_1, X_2(\omega) \leq t_2, \dots, X_n(\omega) \leq t_n\} \in \mathcal{A},$$

azaz a $\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n)$ valószínűség értelmes.

Itt \underline{X} értékei n dimenziós vektorok, vagyis n hosszú számsorozatok, és X_k ennek a k . koordinátáját (elemét) jelöli.

Peremeloszlás

Valószínűségi vektorváltozó együttes eloszlása: \underline{x}_k lehetséges értékek és $p_k = \mathbb{P}(X = \underline{x}_k)$ valószínűségek ($k = 1, 2, \dots$).

Az $\underline{X} = (X_1, X_2, \dots, X_n)$ függvény pontosan akkor valószínűségi vektorváltozó, ha minden koordinátája valószínűségi változó. Ez alapján:

Definíció

Legyen $\underline{X} = (X_1, X_2, \dots, X_n)$ valószínűségi vektorváltozó. Ekkor az X_k valószínűségi változó eloszlását az \underline{X} valószínűségi vektorváltozó k . peremeloszlásának nevezzük.

Az együttes eloszlás meghatározza a peremeloszlásokat. Fordítva nem, kivéve, ha a koordináták függetlenek.

Együttes eloszlás: példa

Kétszer dobunk szabályos kockával. Legyen X az első dobás, Y pedig a dobott számok közül a nagyobb. Ekkor az (X, Y) valószínűségi vektorváltozó együttes eloszlása:

X/Y	1	2	3	4	5	6	összesen
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	0	$1/18$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	0	0	$1/12$	$1/36$	$1/36$	$1/36$	$1/6$
4	0	0	0	$1/9$	$1/36$	$1/36$	$1/6$
5	0	0	0	0	$5/36$	$1/36$	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
összesen	$1/36$	$1/18$	$1/12$	$1/9$	$5/36$	$1/6$	1

A peremeloszlások:

X : $(1, 1/6), (2, 1/6), (3, 1/6), \dots, (6, 1/6)$

Y : $(1, 1/36), (2, 1/18), \dots, (6, 1/6)$

Függetlenség: példa

Mely valószínűségi változók tekinthetők egymástól függetlennek, és melyek között van kapcsolat? Zsófia egy felmérés véletlenszerűen választott résztvevője.

Zsófia autóinak száma

csapadékmennyiség holnap Budapesten

Zsófia havi jövedelme

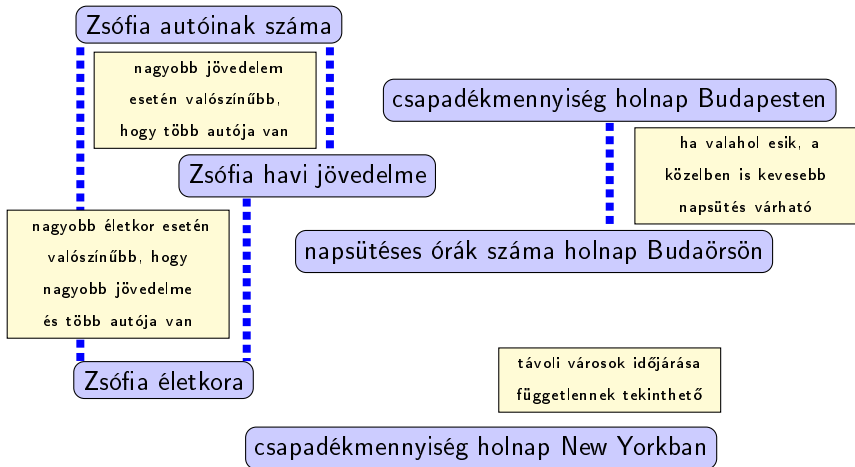
napsütéses órák száma holnap Budaörsön

Zsófia életkora

csapadékmennyiség holnap New Yorkban

Függetlenség: példa

Mely valószínűségi változók tekinthetők egymástól függetlennek, és melyek között van kapcsolat? Zsófia egy felmérés véletlenszerűen választott résztvevője.



Függetlenség: példa

Emlékeztető: az A és B események **függetlenek**, ha

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

Ha például X a csapadékmennyiség holnap Budapesten (mm-ben), és Y New Yorkban, akkor például

$$A : X \leq 5; \quad B : Y \leq 5$$

esetén ez a feltétel így írható:

$$\mathbb{P}(X \leq 5, Y \leq 5) = \mathbb{P}(X \leq 5) \cdot \mathbb{P}(Y \leq 5).$$

Azaz, feltételezve, hogy a két város időjárása egymástól független: annak valószínűsége, hogy **mindkét helyen legfeljebb 5 mm csapadék lesz**, a két esemény **valószínűségének szorzata**.

Valószínűségi változók függetlensége

- **két valószínűségi változóra**: az $X, Y : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\mathbb{P}(X \leq t_1, Y \leq t_2) = \mathbb{P}(X \leq t_1) \cdot \mathbb{P}(Y \leq t_2)$$

teljesül tetszőleges $t_1, t_2 \in \mathbb{R}$ valós számokra.

- **véges sok valószínűségi változóra**: $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) &= \\ &= \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \dots \mathbb{P}(X_n \leq t_n) \end{aligned}$$

teljesül tetszőleges t_1, t_2, \dots, t_n valós számokra.

- **megszámlálható sok valószínűségi változóra**: az $X_1, X_2, X_3 \dots$ valószínűségi változók **függetlenek**, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.

Függetlenség diszkrét esetben

Ha a valószínűségi változók **diszkrét**, azaz lehetséges értékeik halmaza véges vagy megszámlálhatóan végtelen, akkor a függetlenséget az alábbi módon is ellenőrizhetjük.

Az X és Y **diszkrét** valószínűségi változók pontosan akkor **függetlenek**, ha

az X **minden lehetséges x_k értékére** és

az Y **minden lehetséges y_l értékére** teljesül, hogy

$$\mathbb{P}(X = x_k, Y = y_l) = \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_l) \quad (k, l = 1, 2, \dots).$$

Vagyis: annak valószínűsége, hogy X **értéke x_k** és Y **értéke y_l** , ennek a két eseménynek **a valószínűségének a szorzata**.

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy az **első dobott szám** és a **második dobott szám** értéke egymástól független?

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy az **első dobott szám** és a **második dobott szám** értéke egymástól független?

Tipp. a második dobásnál az első dobás értéke „elfelejtődik”, nincs kapcsolat a két dobás között \Rightarrow a két dobott szám **független**.

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy az **első dobott szám** és a **második dobott szám** értéke egymástól független?

Tipp. a második dobásnál az első dobás értéke „elfelejtődik”, nincs kapcsolat a két dobás között \Rightarrow a két dobott szám **független**.

Indoklás. Legyen X az első dobás, Y a második. Legyen például $x_k = 3, y_l = 5$. Ekkor a feltétel teljesül:

$$\frac{1}{36} = \mathbb{P}(X = 3, Y = 5) = \mathbb{P}(X = 3) \cdot \mathbb{P}(Y = 5) = \frac{1}{6} \cdot \frac{1}{6}.$$

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy az **első dobott szám** és a **második dobott szám** értéke egymástól független?

Tipp. a második dobásnál az első dobás értéke „elfelejtődik”, nincs kapcsolat a két dobás között \Rightarrow a két dobott szám **független**.

Indoklás. Legyen X az első dobás, Y a második. Legyen például $x_k = 3, y_l = 5$. Ekkor a feltétel teljesül:

$$\frac{1}{36} = \mathbb{P}(X = 3, Y = 5) = \mathbb{P}(X = 3) \cdot \mathbb{P}(Y = 5) = \frac{1}{6} \cdot \frac{1}{6}.$$

Hasonlóképpen tetszőleges (x_k, y_l) lehetséges értékekre (azaz 1 és 6 közötti egészekre) igaz, hogy

$$\frac{1}{36} = \mathbb{P}(X = x_k, Y = y_l) = \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_l) = \frac{1}{6} \cdot \frac{1}{6}.$$

Ezért valóban **a két dobás egymástól független**.

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy a dobott számok **összege** és **szorzata** független egymástól?

Valószínűségi változók függetlensége: példa

Kétszer dobunk egy szabályos dobókockával. Igaz-e, hogy a dobott számok **összege** és **szorzata** független egymástól?

Tipp: minél nagyobb az összeg, annál valószínűbb, hogy a szorzat értéke is inkább nagy lesz \Rightarrow **nem függetlenek**.

Indoklás: legyen X az összeg, Y a szorzat. Ha például $X = 2$: ez csak úgy lehet, hogy mindkét dobás 1-es, vagyis ekkor Y értéke biztosan 1. Ezért ha például $x_1 = 2$ és $y_2 = 2$ -t választunk, $X = 2$ és $Y = 2$ egyszerre nem következhetnek be, és így:

$$\begin{aligned} 0 &= \mathbb{P}(X = 2, Y = 2) \neq \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 2) = \\ &= \mathbb{P}(11) \cdot \mathbb{P}(12 \text{ vagy } 21) = \frac{1}{36} \cdot \frac{1}{18} > 0. \end{aligned}$$

Vagyis az $x_1 = 2$ és $y_2 = 2$ párra nem teljesül az előírt feltétel, az **összeg és szorzat nem függetlenek**.

A várható érték tulajdonságai

- (összeg várható értéke) Ha X, Y valószínűségi változók, és $X, Y, X + Y$ várható értéke létezik, akkor

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- (konstans kiemelése) Ha az X valószínűségi változó várható értéke létezik, és c tetszőleges valós szám, akkor

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

- (szorzat várható értéke független esetben) Ha az X és Y valószínűségi változók **függetlenek**, és $X, Y, X \cdot Y$ várható értéke létezik, akkor

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

- (függvény várható értéke) Ha $g : \mathbb{R} \rightarrow \mathbb{R}$ olyan függvény, hogy $\mathbb{E}(X)$ létezik, akkor

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

ahol az X lehetséges értékei x_1, x_2, \dots

A szórásnégyzet tulajdonságai

- (nemnegativitás) $D^2(X) \geq 0$ és $D(X) \geq 0$ mindig teljesül
- (szorzás és eltolás) ha a, b valós számok, X véges szórású valószínűségi változó, akkor

$$D^2(aX + b) = a^2 D^2(X) \quad \Rightarrow \quad D(aX + b) = |a| D(X).$$

A szórásnégyzet tulajdonságai

- (nemnegativitás) $D^2(X) \geq 0$ és $D(X) \geq 0$ mindig teljesül
- (szorzás és eltolás) ha a, b valós számok, X véges szórású valószínűségi változó, akkor

$$D^2(aX + b) = a^2 D^2(X) \quad \Rightarrow \quad D(aX + b) = |a| D(X).$$

- (összeg szórása független esetben) ha az X, Y valószínűségi változók **függetlenek** és szórásuk létezik, akkor

$$D^2(X + Y) = D^2(X) + D^2(Y) \quad \Rightarrow \quad D(X + Y) = \sqrt{D^2(X) + D^2(Y)}.$$

- van olyan valószínűségi változó, melynek várható értéke véges, de a szórása nem létezik (például: $\mathbb{P}(X = k) = c/k^3$ megfelelő c -vel)

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;
- $D(X + Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;
- $D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(X - Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;
- $D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(X - Y) = \sqrt{D^2(X) + (-1)^2 D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(2X + 3Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;
- $D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(X - Y) = \sqrt{D^2(X) + (-1)^2 D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(2X + 3Y) = \sqrt{2^2 D^2(X) + 3^2 D^2(Y)} = \sqrt{4 \cdot 1 + 9 \cdot 4} = \sqrt{40}$;
- $D(2X - 3Y) =$

A várható érték és szórás kiszámítása

Példa. Legyen X olyan valószínűségi változó, melynek várható értéke 4, szórása 1, az Y valószínűségi változó várható értéke 6, szórása pedig 2. Tegyük fel továbbá, hogy X és Y **függetlenek**. Ekkor

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10$;
- $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = -2$;
- $\mathbb{E}(2X + 3Y) = 2\mathbb{E}(X) + 3\mathbb{E}(Y) = 2 \cdot 4 + 3 \cdot 6 = 26$;
- $\mathbb{E}(2X - 3Y) = 2\mathbb{E}(X) - 3\mathbb{E}(Y) = 2 \cdot 4 - 3 \cdot 6 = -10$;
- $D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(X - Y) = \sqrt{D^2(X) + (-1)^2 D^2(Y)} = \sqrt{1 + 2^2} = \sqrt{5}$;
- $D(2X + 3Y) = \sqrt{2^2 D^2(X) + 3^2 D^2(Y)} = \sqrt{4 \cdot 1 + 9 \cdot 4} = \sqrt{40}$;
- $D(2X - 3Y) = \sqrt{2^2 D^2(X) + (-3)^2 D^2(Y)} = \sqrt{4 \cdot 1 + 9 \cdot 4} = \sqrt{40}$.

A binomiális eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó binomiális eloszlású n renddel és p paraméterrel, azaz

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (k = 0, 1, \dots, n),$$

akkor

$$\mathbb{E}(X) = np; \quad D(X) = \sqrt{np(1 - p)}.$$

Példa. Egy kérdőív egy kérdésére $n = 1000$ megkérdezett közül mindenki a többi-ektől függetlenül $p = 0,65$ valószínűséggel válaszol. Ekkor a válaszadók számának (melyet jelöljünk X -szel) a várható értéke:

$$\mathbb{E}(X) = np = 1000 \cdot 0,65 = 650,$$

míg a szórása:

$$D(X) = \sqrt{np(1 - p)} = \sqrt{1000 \cdot 0,65 \cdot 0,35} = 15,08.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Legyen X binomiális eloszlású n renddel és p paraméterrel. Azaz: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeresek száma. Vezessük be az alábbi indikátorokat $j = 1, 2, \dots, n$ esetén:

$$X_j = \begin{cases} 1 & \text{ha a } j. \text{ kísérlet sikeres;} \\ 0 & \text{ha a } j. \text{ kísérlet nem sikeres.} \end{cases}$$

Ekkor X éppen az indikátorok összege (az egyesek száma):

$$X = X_1 + X_2 + \dots + X_n = \sum_{j=1}^n X_j.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Legyen X binomiális eloszlású n renddel és p paraméterrel. Azaz: n független kísérletet végzünk, mindegyik p valószínűséggel sikerül, X a sikeresek száma. Vezessük be az alábbi indikátorokat $j = 1, 2, \dots, n$ esetén:

$$X_j = \begin{cases} 1 & \text{ha a } j. \text{ kísérlet sikeres;} \\ 0 & \text{ha a } j. \text{ kísérlet nem sikeres.} \end{cases}$$

Ekkor X éppen az indikátorok összege (az egyesek száma):

$$X = X_1 + X_2 + \dots + X_n = \sum_{j=1}^n X_j.$$

Mivel bármely j -re

$$\mathbb{E}(X_j) = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p,$$

ezért

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n) = p + p + \dots + p = np.$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Ugyanazokat az indikátorokat használjuk:

$$X_j = \begin{cases} 1 & \text{ha a } j.\text{kísérlet sikeres;} \\ 0 & \text{ha a } j.\text{kísérlet nem sikeres.} \end{cases}$$

A binomiális eloszlás várható értéke és szórása

Bizonyítás. Ugyanazokat az indikátorokat használjuk:

$$X_j = \begin{cases} 1 & \text{ha a } j.\text{kísérlet sikeres;} \\ 0 & \text{ha a } j.\text{kísérlet nem sikeres.} \end{cases}$$

Most $X_j = X_j^2$, hiszen $0^2 = 0$ és $1^2 = 1$, és már láttuk, hogy $\mathbb{E}(X_j) = p$. Ezért

$$D^2(X_j) = \mathbb{E}(X_j^2) - \mathbb{E}(X_j)^2 = \mathbb{E}(X_j) - \mathbb{E}(X_j)^2 = p - p^2 = p(1 - p),$$

Mivel az X_j indikátorok **függetlenek**, és az összegük X :

$$\begin{aligned} D(X) &= \sqrt{D^2(X_1 + X_2 + \dots + X_n)} = \sqrt{D^2(X_1) + D^2(X_2) + \dots + D^2(X_n)} = \\ &= \sqrt{p(1 - p) + p(1 - p) + \dots + p(1 - p)} = \sqrt{np(1 - p)}. \end{aligned}$$

A hipergeometrikus eloszlás várható értékére vonatkozó összefüggés szintén indikátorokkal igazolható.

Házi feladat október 16., 10:15-ig

Annát minden nap a többitől függetlenül $p = 0,001$ valószínűséggel éri baleset. Legyen X az, hogy holnaptól hányadik nap az első, amikor baleset éri, Y pedig az, hogy az első balesete után hány nap telik el a másodikig. (azaz a második balesete holnaptól $X + Y$ nap múlva lesz.)

- Mennyi $X + Y$ várható értéke?
- Mennyi $X + Y$ szórása?
- Mennyi $X - Y$ várható értéke?
- Mennyi $X - Y$ szórása?