Binomiális eloszlás: definíció (5. előadás)

- n független kísérletet végzünk;
- mindegyik p valószínűséggel sikerül;
- X a sikeres kísérletek száma.

Az X valószínűségi változó **binomiális eloszlású** n renddel és p paraméterrel, ha lehetséges értékei:

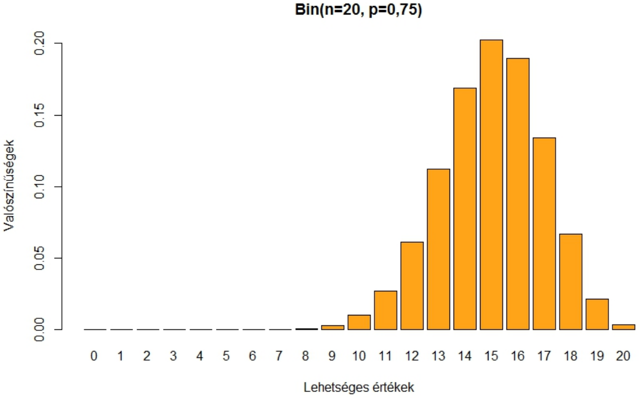
$$0, 1, 2, \ldots, n,$$

és minden $0 \leq k \leq n$ egészre

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

($n \geq 1$ egész, $0 < p < 1$.) Jelölés: $\mathrm{Bin}(n, p)$.

Példa: binomiális eloszlás



Binomiális eloszlás, $n = 20$, $p = 0,75$. Vízszintes tengely: lehetséges értékek, azaz $k = 0, 1, \ldots, 20$, oszlopok magassága: a $\mathbb{P}(X = k)$ valószínűségek.

Példa: binomiális eloszlás

Egy felmérésben $n = 1500$ embert kérdezünk meg. Egy adott kérdésre minden résztvevő **egymástól függetlenül** $p = 0,8$ valószínűséggel válaszol. Jelölje X, hogy hányan válaszoltak erre a kérdésre. Ekkor

- X **binomiális eloszlású** $n = 1500$ renddel és $p = 0,8$ paraméterrel.
- Tetszőleges $0 \le k \le 1500$ esetén

$$\mathbb{P}(k \text{ válasz}) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} =$$
$$= \binom{1500}{k} 0,8^k \cdot 0,2^{1500-k}.$$

- Például annak valószínűsége, hogy pontosan $k = 1200$-an válaszolnak a kérdésre:

$$\mathbb{P}(1200 \text{ válasz}) = \mathbb{P}(X = 1200) = \binom{1500}{1200} 0,8^{1200} \cdot 0,2^{300} = 2,57\%.$$

A binomiális eloszlás várható értéke és szórása

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

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

akkor X **várható értéke**, illetve **szórása**:

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

Példa: binomiális eloszlás

Egy felmérésben $n = 1500$ embert kérdezünk meg. Egy adott kérdésre minden résztvevő **egymástól függetlenül** $p = 0,8$ valószínűséggel válaszol. Jelölje X, hogy hányan válaszoltak erre a kérdésre. Ekkor

- A válaszadók számának **várható értéke**:

$$\mathbb{E}(X) = np = 1500 \cdot 0,8 = 1200.$$

- A válaszadók számának **szórása**:

$$D(X) = \sqrt{np(1-p)} = \sqrt{1500 \cdot 0,8 \cdot 0,2} = 15,5.$$

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**:

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

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

$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

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**:

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

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

$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

$$= \frac{100000 \cdot 99999 \cdot \ldots (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx$$

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**:

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

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

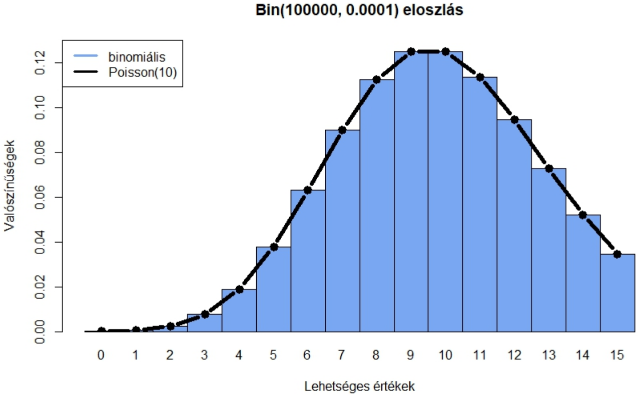
$$\mathbb{P}(X = k) = \binom{100000}{k} \cdot 0,0001^k \cdot 0,9999^{100000-k} =$$

$$= \frac{100000 \cdot 99999 \cdot \ldots (100001 - k)}{k!} \cdot 0,0001^k \cdot 0,9999^{100000-k} \approx$$

$$\approx \frac{100000^k \cdot 0,0001^k}{k!} \left(1 - \frac{10}{100000}\right)^{100000} \approx \frac{10^k}{k!} e^{-10},$$

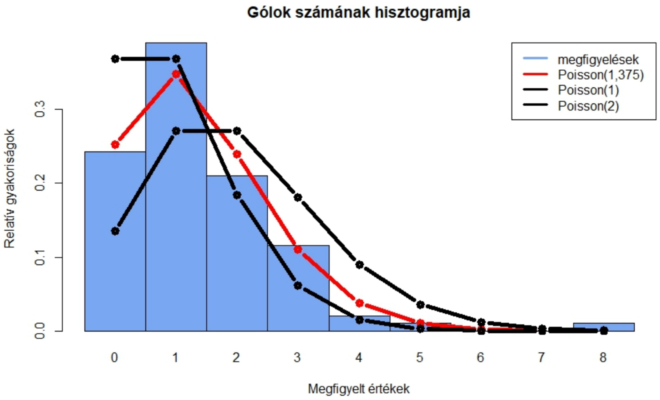
felhasználva, hogy $\lim_{n\to\infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$ tetszőleges $x > 0$-ra.

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.

Poisson-eloszlás: definíció

- ha n független kísérletből mindegyik p valószínűséggel sikeres, ahol n „nagy” és p „kicsi”: $\lambda = np$ a sikeres kísérletek számának várható értéke;
- annak valószínűsége, hogy pontosan k kísérlet sikeres, $\frac{\lambda^k}{k!}e^{-\lambda}$-val közelíthető (a számolás és az ábra alapján);
- ez a gyakran használt Poisson-eloszlás **ritkán bekövetkező események számának** modellezésére.

Definíció

Legyen $\lambda > 0$. Az X valószínűségi változó λ **paraméterű Poisson-eloszlású**, *ha lehetséges értékei:*

$$k = 0, 1, 2, \ldots, \text{ és ekkor } \mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

A λ paraméterű Poisson-eloszlás **várható értéke** és **szórása**:

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

Poisson-eloszlás: példa

Az X valószínűségi változó λ **paraméterű Poisson-eloszlású**, ha lehetséges értékei:

$$k = 0, 1, 2, \ldots, \text{ és ekkor } \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

A λ paraméterű Poisson-eloszlás **várható értéke** és **szórása**:

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

Példa. Tegyük fel, hogy egy városban az egy nap alatt bekövetkező autóbalesetek száma Poisson-eloszlású, és **várható értéke** $3,61$. Ekkor az egy nap alatt bekövetkező autóbalesetek számának **szórása**:

$$D(X) = \sqrt{3,61} = 1,9.$$

Annak valószínűsége, hogy pontosan 5 baleset lesz:

$$\mathbb{P}(X = 5) = \frac{3,61^5}{5!} e^{-5} = 14\%.$$

Binomális és Poisson-eloszlás az R-ben

- ha X binomiális eloszlású n renddel és p paraméterrel, akkor

$$\mathbb{P}(X=k): \qquad \mathtt{dbinom}(\mathtt{k}, \mathtt{size}=\mathtt{n}, \mathtt{prob}=\mathtt{p})$$

  és

$$\mathbb{P}(X\leq k): \qquad \mathtt{pbinom}(\mathtt{k}, \mathtt{size}=\mathtt{n}, \mathtt{prob}=\mathtt{p}).$$

- Továbbá az
  `minta<-rbinom(r, size=n, prob=p)`
  eredményeképpen a „minta” vektorba r darab *független* adott binomiális eloszlású valószínűségi változó kerül.
- ha X Poisson-eloszlású λ paraméterrel, akkor

$$\mathbb{P}(X=k): \qquad \mathtt{dpois}(\mathtt{k}, \mathtt{lambda}=\lambda)$$

  és

$$\mathbb{P}(X\leq k): \qquad \mathtt{ppois}(\mathtt{k}, \mathtt{lambda}=\lambda).$$

- Továbbá az
  `minta<-rpois(r, lambda=`λ`)`
  eredményeképpen a „minta” vektorba r darab *független* λ paraméterű eloszlású valószínűségi változó kerül.

Példa R-ben

```
> minta=rbinom(100, size=100000, prob=0.0001)
> hist(minta)
> dbinom(8, size=100000, prob=0.0001)
[1] 0.1126013
> minta=rpois(100, lambda=3.61)
> hist(minta)
> mean(minta)
[1] 3.91
> dpois(5, lambda=3.61)
[1] 0.1382139
```

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:	$\binom{9}{k} \cdot \binom{11}{7-k} \longleftarrow$

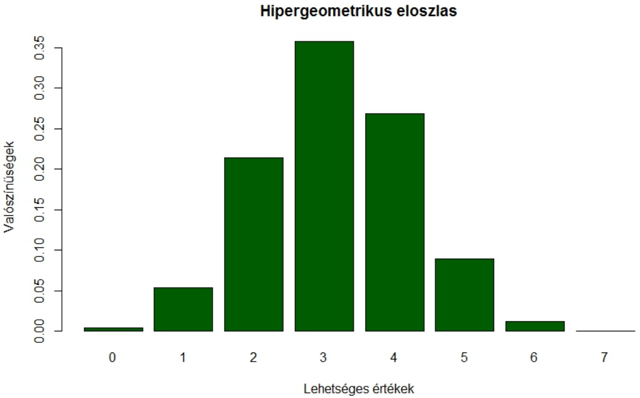
szorzás: bármely balkezes választás bármely jobbkezessel jó

osztás: minden lehetőség egyformán valószínű

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

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

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

- **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:

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

A hipergeometriai eloszlás várható értéke és szórása

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

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

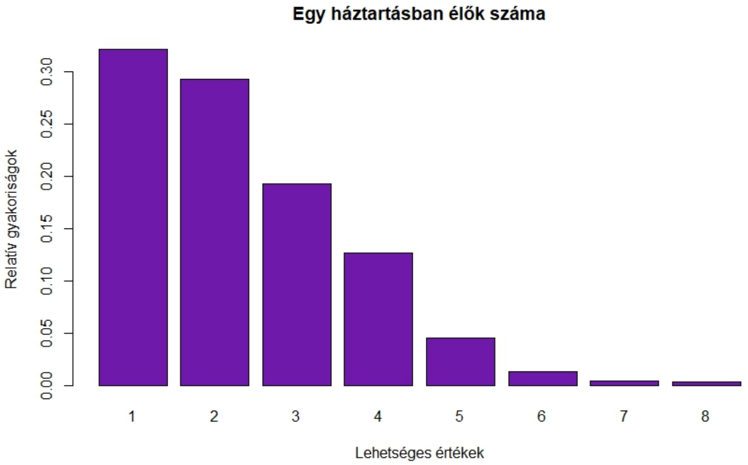
akkor

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

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**:

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

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.

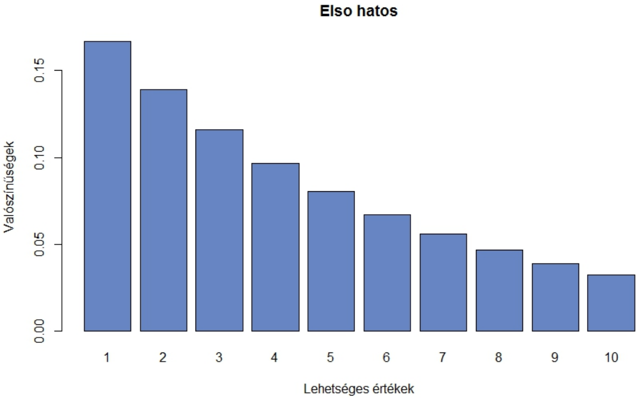
$\mathbb{P}(Y = 1) = \mathbb{P}(\text{az első ember válaszol}) = 0,2;$

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

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

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

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: 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, 0).

Geometriai eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó geometriai eloszlású p paraméterrel, azaz

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

akkor

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

Geometriai eloszlás várható értéke és szórása

Állítás

Ha az X valószínűségi változó geometriai eloszlású p paraméterrel, azaz

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

akkor

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

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

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

Nevezetes diszkrét eloszlások várható értéke és szórása

- Ha X binomiális eloszlású n renddel és p paraméterrel:

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

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

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

- Ha X Poisson-eloszlású λ paraméterrel:

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

- Ha X geometriai eloszlású p paraméterrel:

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

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

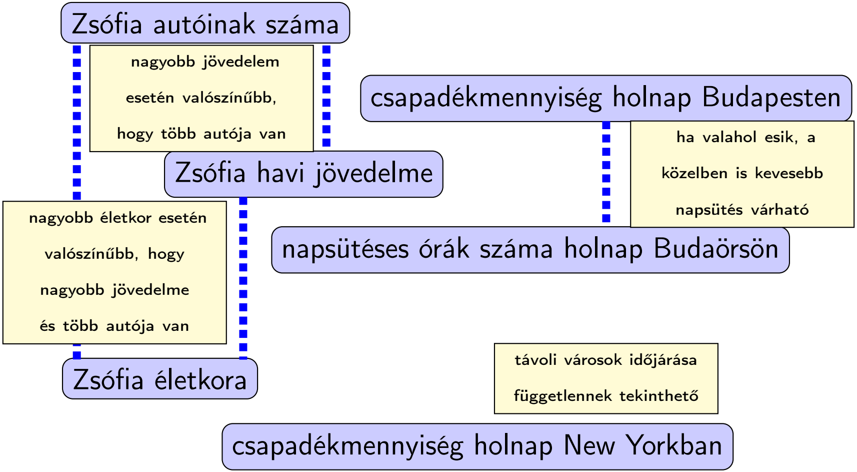
napsütéses órák száma holnap Budaörsön

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

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

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

$$\mathsf{A} : X \leq 5; \qquad \mathsf{B} : Y \leq 5$$

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 $X, Y : \Omega \to \mathbb{R}$ 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**: $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n) = \\ = \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \ldots \mathbb{P}(X_n \leq t_n)$$

  teljesül tetszőleges $t_1, t_2, \ldots, t_n$ valós számokra.

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

Függetlenség diszkrét esetben

Ha a valószínűségi változók **diszkrétek**, 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

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

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 ⇒ 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:

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

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

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

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

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 $g : \mathbb{R} \to \mathbb{R}$ olyan függvény, hogy $\mathbb{E}(X)$ létezik, és az X lehetséges értékei $x_1, x_2, \ldots$, akkor
$$\mathbb{E}(g(X)) = \sum_{i=1}^{\infty} g(x_i)\mathbb{P}(X = x_i).$$

Összeg várható értéke

Állítás

Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor

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

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{k,m} (x_k + y_m)\mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k,m} x_k \mathbb{P}(X = x_k, Y = y_m) + \sum_{k,m} y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k} x_k \mathbb{P}(X = x_k) + \sum_{m} y_m \mathbb{P}(Y = y_m) = \mathbb{E}(X) + \mathbb{E}(Y),
\end{aligned}$$

ahol azt használtuk, hogy az $\{X = x_k, Y = y_m\}$ események kizáróak, uniójuk $\{X = x_k\}$, és hasonlóképpen a másik tagban az Y esetén.

Összeg várható értéke

Állítás

Ha X, Y valószínűségi változók, és $X, Y, X+Y$ várható értéke létezik, akkor

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

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X+Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{k,m} (x_k + y_m)\mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k,m} x_k \mathbb{P}(X = x_k, Y = y_m) + \sum_{k,m} y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&= \sum_{k} x_k \mathbb{P}(X = x_k) + \sum_{m} y_m \mathbb{P}(Y = y_m) = \mathbb{E}(X) + \mathbb{E}(Y),
\end{aligned}$$

ahol azt használtuk, hogy az $\{X = x_k, Y = y_m\}$ események kizáróak, uniójuk $\{X = x_k\}$, és hasonlóképpen a másik tagban az Y esetén.

Következmény: $\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)$.

Szorzat várható értéke független esetben

Állítás

Ha X, Y **független** *valószínűségi változók, és X, Y várható értéke létezik, akkor*

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

Bizonyítás. Legyenek X lehetséges értékei $x_1, x_2, \ldots$, az Y lehetséges értékei $y_1, y_2, \ldots$. Ekkor az $X + Y$ lehetséges értékei $x_k + y_m$ alakúak, és

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{k,m} x_k \cdot y_m \mathbb{P}(X = x_k, Y = y_m) = \\
&\stackrel{(*)}{=} \sum_{k,m} x_k y_m \mathbb{P}(X = x_k) \cdot \mathbb{P}(Y = y_m) \\
&= \left(\sum_k x_k \mathbb{P}(X = x_k) \right) \cdot \left(\sum_m y_m \mathbb{P}(Y = y_m) \right) = \mathbb{E}(X) \cdot \mathbb{E}(Y),
\end{aligned}$$

ahol a $(*)$ lépésben használtuk a függetlenségnek a diszkrét valószínűségi változókra vonatkozó alakját.