Valószínűségi vektorváltozó (5. előadás)

Definíció

Az

$$\underline{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

függvény **valószínűségi vektorváltozó**, *ha* $X_1, X_2, \ldots, X_n$ *valószínűségi változók.*

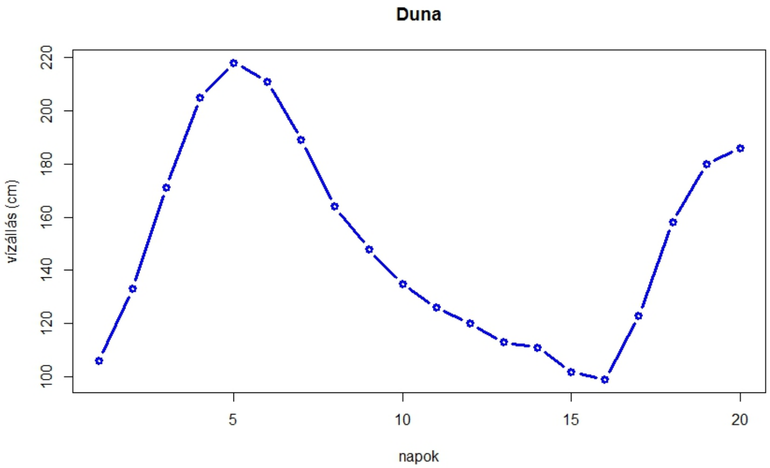
Ha $\underline{X}$ *valószínűségi vektorváltozó, akkor az* X_i *valószínűségi változó eloszlását az* $\underline{X}$ i. **peremeloszlásának** *nevezzük.*

Az $\underline{X}$ *valószínűségi vektorváltozó* **diszkrét**, *ha értékkészlete véges vagy megszámlálhatóan végtelen.*

Példa. 1000 embert megkérdezünk a havi jövedelméről. Legyen X_i az i. megkérdezett jövedelme. Ekkor $(X_1, X_2, \ldots, X_{1000})$ valószínűségi vektorváltozó.

Y_i: a Duna vízállása az i. napon $(i = 1, 2, \ldots, 20)$.

Valószínűségi vektorváltozó: példa



$X_1 = 106, X_2 = 133, \ldots, X_{20} = 186$ (az adatok forrása: Országos Vízjelző Szolgálat)

Együttes eloszlásfüggvény

Definíció

Az $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó együttes eloszlásfüggvénye az* $F : \mathbb{R}^n \to [0,1]$ *függvény, melyre*

$$F(\underline{t}) = F(t_1, \ldots, t_n) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n), \text{ ha } (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

Például: egy véletlenszerűen választott embert megkérdezünk a havi jövedelméről (X_1), a havi kiadásairól (X_2), és az életkoráról (X_3). Ekkor (X_1, X_2, X_3) valószínűségi vektorváltozó, és ha eloszlásfüggvénye F, akkor például

$$F(200000, 150000, 40) = \mathbb{P}(X_1 \leq 200000, X_2 \leq 1500000, X_3 \leq 40)$$

annak valószínűsége, hogy egy véletlenszerűen választott ember havi jövedelme legfeljebb 200000 (forint), havi kiadása legfeljebb 150000 (forint), életkora pedig legfeljebb 40 (év).

Együttes sűrűségfüggvény

Definíció

Az $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó abszolút folytonos, ha van olyan* $f : \mathbb{R}^n \to \mathbb{R}$ *függvény, melyre*

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ *esetén. Ilyenkor az* f *függvényt az* $(X_1, X_2, \ldots, X_n)$ **együttes sűrűségfüggvényének** *nevezzük.*

Együttes sűrűségfüggvény

Definíció

Az $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó abszolút folytonos, ha van olyan* $f : \mathbb{R}^n \to \mathbb{R}$ *függvény, melyre*

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ *esetén. Ilyenkor az* f *függvényt az* $(X_1, X_2, \ldots, X_n)$ **együttes sűrűségfüggvényének** *nevezzük.*

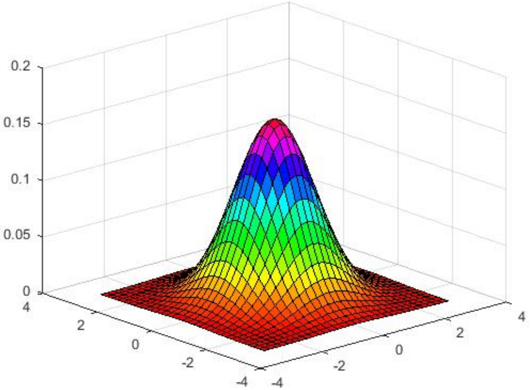
Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Ekkor egy $A \subseteq \mathbb{R}^n$ halmazra

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in A) = \int_A f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n.$$

Következmény:

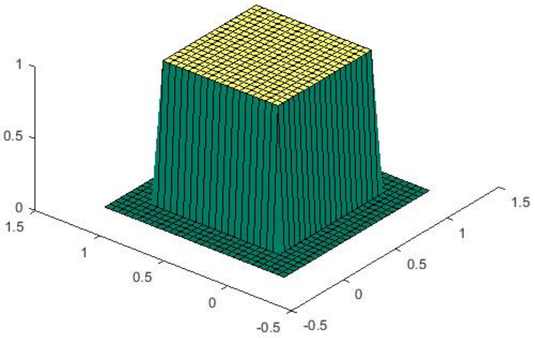
$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n = 1.$$

Kétdimenziós normális eloszlás



Két független standard normális eloszlás együttes sűrűségfüggvénye

Egyenletes eloszlás a négyzeten



A $[0,1] \times [0,1]$ négyzeten egyenletes eloszlás (két független, $[0,1]$-en egyenletes eloszlás) együttes sűrűségfüggvénye

Függetlenség

Definíció (Véges eset)

Azt mondjuk, hogy az $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *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, azaz*

$$F(t_1, t_2, \ldots, t_n) = F_1(t_1) \cdot F_2(t_2) \cdot \ldots F_n(t_n),$$

ahol F_j *az* X_j *valószínűségi változó eloszlásfüggvénye.*

Definíció (Végtelen eset)

Az $X_1, X_2, X_3 \ldots$ *valószínűségi változók függetlenek, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.*

Függetlenség és együttes sűrűségfüggvény

Állítás (Függetlenség és sűrűségfüggvény)

Legyen az $\underline{X} = (X_1, \dots, X_n)$ *valószínűségi vektorváltozó együttes sűrűségfüggvénye* f*, továbbá az* X_j *valószínűségi változó sűrűségfüggvénye* f_j *minden* $j = 1, 2, \dots, n$ *esetén. Ezekkel a jelölésekkel:* $X_1, X_2, \dots, X_n$ *pontosan akkor függetlenek, ha*

$$f(t_1, \dots, t_n) = f_1(t_1) \cdot f_2(t_2) \dots f_n(t_n)$$

teljesül bármely $t_1, t_2, \dots, t_n$ *valós számokra.*

Például: ha X_1, X_2 független standard normális eloszlású valószínűségi változók, akkor együttes sűrűségfüggvényük

$$f(t_1, t_2) = f_1(t_1) f_2(t_2) = \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_1^2}{2} \right) \right) \cdot \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_2^2}{2} \right) \right)$$
$$= \frac{1}{2\pi} \exp\left(-\frac{t_1^2 + t_2^2}{2} \right).$$

Ez csak a (t_1, t_2) pont origótól mért távolságától függ.

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Hogyan kapható meg például az első peremeloszlás, azaz X_1 sűrűségfüggvénye?

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Hogyan kapható meg például az első peremeloszlás, azaz X_1 sűrűségfüggvénye?

Állítás

Tegyük fel, hogy az $\underline{X} = (X_1, \ldots, X_n)$ valószínűségi változók együttes sűrűségfüggvénye f. Ekkor az X_j valószínűségi változó sűrűségfüggvénye (melyet f_j-vel jelölünk), azaz a j. peremsűrűségfüggvény így kapható meg f-ből:

$$f_j(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(s_1, \ldots, s_{j-1}, t, s_{j+1}, \ldots, s_n) ds_1 \ldots ds_{j-1} ds_{j+1} \ldots ds_n.$$

Speciálisan $n = 2$-re:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Többdimenziós normális eloszlás

Definíció

Az $(X_1, X_2, \ldots, X_n)$ *valószínűségi vektorváltozó n-dimenziós normális eloszlású, ha tetszőleges* $a_1, a_2, \ldots, a_n$ *valós számokra az*

$$\sum_{j=1}^{n} a_j X_j = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

valószínűségi változó normális eloszlású.

Többdimenziós normális eloszlás

Definíció

Az $(X_1, X_2, \ldots, X_n)$ *valószínűségi vektorváltozó n-dimenziós normális eloszlású, ha tetszőleges* $a_1, a_2, \ldots, a_n$ *valós számokra az*

$$\sum_{j=1}^{n} a_j X_j = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

valószínűségi változó normális eloszlású.

- Abból, hogy $X_1, X_2, \ldots, X_n$ normális eloszlásúak, nem következik, hogy $(X_1, X_2, .$ normális eloszlású.
- Ha $X_1, X_2, \ldots, X_n$ **függetlenek** és normális eloszlásúak, akkor $(X_1, X_2, \ldots, X_n)$ többdimenziós normális eloszlású.
- Ha $\underline{X} = (X_1, X_2, \ldots, X_n)$ többdimenziós normális eloszlású, és A tetszőleges $k \times n$ méretű mátrix, akkor az $A\underline{X}$ valószínűségi vektorváltozó k-dimenziós normális eloszlású.

A kovariancia

Definíció (Kovariancia)

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor az X és Y kovarianciája:

$$\operatorname{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\big].$$

Legyenek $X, Y, Z, X_1, \ldots, X_n$ olyan valószínűségi változók, melyek szórása létezik. Ekkor a következők teljesülnek.

- **A kovariancia kiszámítása:**

$$\operatorname{cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y).$$

- Szimmetria. $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$.
- Kapcsolat a szórásnégyzettel. $\operatorname{cov}(X, X) = D^2(X)$.

A kovariancia tulajdonságai

- Konstanssal való kovariancia. $\operatorname{cov}(X, c) = 0$, ha $c \in \mathbb{R}$.
- **Linearitás.** Egyrészt
$$\operatorname{cov}(X + Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z),$$
másrészt tetszőleges $c \in \mathbb{R}$ számra
$$\operatorname{cov}(cX, Y) = c \cdot \operatorname{cov}(X, Y).$$
- **Függetlenséggel való kapcsolat.** Ha az X és Y valószínűségi változók függetlenek, akkor $\operatorname{cov}(X, Y) = 0$. Fordítva nem igaz: lehet nem független esetben is 0 a kovariancia.
- **Összeg szórásnégyzete.** $D^2(X + Y) = D^2(X) + D^2(Y) + 2\operatorname{cov}(X, Y)$. Továbbá
$$D^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D^2(X_i) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j).$$
- Különbség szórásnégyzete. $D^2(X - Y) = D^2(X) + D^2(Y) - 2\operatorname{cov}(X, Y)$.

Kovariancia: példa.

Példa. Legyen X Poisson-eloszlású valószínűségi változó 2 paraméterrel. A linearitás, a szórásnégyzettel való kapcsolat és a konstanssal való kovariancia alapján:

$$\begin{aligned} \operatorname{cov}(X+3, 2\cdot X) &= 2\operatorname{cov}(X+3, X) = 2\operatorname{cov}(X, X) + 2\operatorname{cov}(3, X) = \\ &= 2D^2(X) = 2\cdot 2 = 4. \end{aligned}$$

Kovariancia: példa.

Példa. Legyen X Poisson-eloszlású valószínűségi változó 2 paraméterrel. A linearitás, a szórásnégyzettel való kapcsolat és a konstanssal való kovariancia alapján:

$$\mathrm{cov}(X+3, 2\cdot X) = 2\mathrm{cov}(X+3, X) = 2\mathrm{cov}(X, X) + 2\mathrm{cov}(3, X) =$$
$$= 2D^2(X) = 2\cdot 2 = 4.$$

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é 400. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a kovarianciája?

Kovariancia: példa.

Példa. Legyen X Poisson-eloszlású valószínűségi változó 2 paraméterrel. A linearitás, a szórásnégyzettel való kapcsolat és a konstanssal való kovariancia alapján:

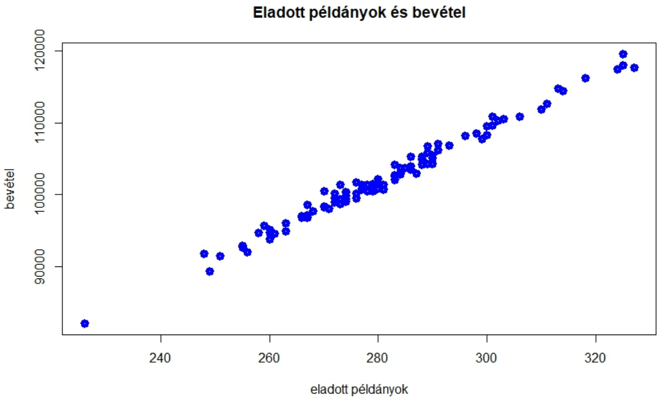
$$\begin{aligned} \mathrm{cov}(X+3, 2\cdot X) &= 2\mathrm{cov}(X+3, X) = 2\mathrm{cov}(X,X) + 2\mathrm{cov}(3,X) = \\ &= 2D^2(X) = 2\cdot 2 = 4. \end{aligned}$$

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é 400. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a kovarianciája?

$$\begin{aligned} \mathrm{cov}(X+Y, 300X+400Y) &= \mathrm{cov}(X,300X) + \mathrm{cov}(X,400Y) + \mathrm{cov}(Y,300X) + \\ &\quad + \mathrm{cov}(Y,400Y) = 300\cdot\mathrm{cov}(X,X) + 400\cdot\mathrm{cov}(Y,Y) \\ &= 300D^2(X) + 400D^2(Y) = \\ &= 300\cdot 100 + 400\cdot 180 = 102000, \end{aligned}$$

ahol felhasználtuk a linearitást, azt, hogy függetlenség esetén 0 a kovariancia, illetve a Poisson-eloszlás tulajdonságait.

Kovariancia: példa



A bevétel $(300X + 400Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 102000.

Korrelációs együttható

Definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója***:*

$$R(X,Y) = \frac{\operatorname{cov}(X,Y)}{D(X)D(Y)},$$

ha $D(X) > 0, D(Y) > 0$ és 0, ha $D(X) = 0$ vagy $D(Y) = 0$.

Korrelációs együttható

Definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója**:

$$R(X, Y) = \frac{\mathrm{cov}(X, Y)}{D(X)D(Y)},$$

ha $D(X) > 0, D(Y) > 0$ és 0, ha $D(X) = 0$ vagy $D(Y) = 0$.

Állítás

Legyenek X és Y olyan valószínűségi változók, melyek szórása létezik.

(i) *Ekkor teljesül, hogy*

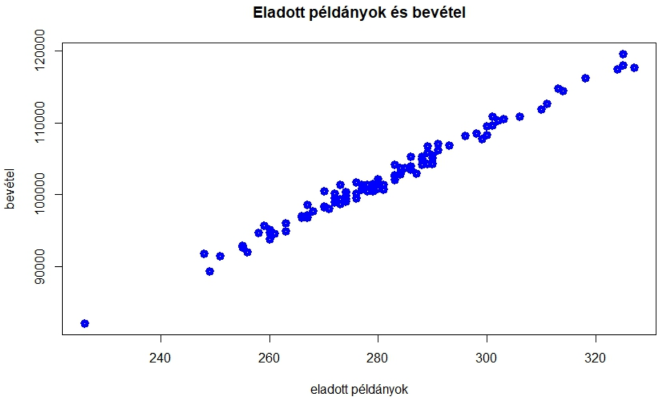
$$|R(X, Y)| \leq 1.$$

(ii) *Legyen $a > 0$ valós szám, b tetszőleges valós szám. Ekkor*

$$R(X, aX + b) = 1 \text{ és } R(X, -aX + b) = -1.$$

(iii) *Tegyük fel, hogy $|R(X, Y)| = 1$. Ekkor léteznek olyan a és b valós számok, hogy az $Y = aX + b$ egyenlet 1 valószínűséggel teljesül.*

Korrelációs együttható: példa



A bevétel $(300X + 400Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 102000, korrelációs együttható: $0,9915$.

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é 400. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$
\begin{aligned}
\mathrm{cov}(X+Y, 300X+400Y) &= 300 \cdot 100 + 400 \cdot 180 = 102000; \\
D(X+Y) &= \sqrt{D^2(X) + D^2(Y)} = \sqrt{100+180} = 16,73; \\
D(300X+400Y) &= \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} \\
&= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = 6148,17; \\
R(X+Y, 300X+400Y) &= \frac{\mathrm{cov}(X+Y, 300X+400Y)}{D(X+Y)D(300X+400Y)} = \frac{102000}{16,73 \cdot 6148,17} \\
&= 0,9915.
\end{aligned}
$$

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é 400. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$\begin{aligned}
\operatorname{cov}(X+Y, 300X+400Y) &= 300 \cdot 100 + 400 \cdot 180 = 102000; \\
D(X+Y) &= \sqrt{D^2(X) + D^2(Y)} = \sqrt{100+180} = 16,73; \\
D(300X+400Y) &= \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} \\
&= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = 6148,17; \\
R(X+Y, 300X+400Y) &= \frac{\operatorname{cov}(X+Y, 300X+400Y)}{D(X+Y)D(300X+400Y)} = \frac{102000}{16,73 \cdot 6148,17} \\
&= 0,9915.
\end{aligned}$$

A korrelációs együttható értéke majdnem 1, azaz erős pozitív korreláció van az eladott példányok száma és a bevétel között.

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é **4000**. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$
\begin{aligned}
\mathrm{cov}(X+Y, 300X+4000Y) &= 300\cdot 100 + 4000\cdot 180 = 750000;\\
D(X+Y) &= \sqrt{D^2(X)+D^2(Y)} = \sqrt{100+180} = 16,73;\\
D(300X+4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} =\\
&= \sqrt{300^2\cdot 100 + 4000^2\cdot 180} = 53749,42;\\
R(X+Y, 300X+4000Y) &= \frac{\mathrm{cov}(X+Y, 300X+4000Y)}{D(X+Y)D(300X+4000Y)} = \frac{750000}{16,73\cdot 53749,42}\\
&= 0,083.
\end{aligned}
$$

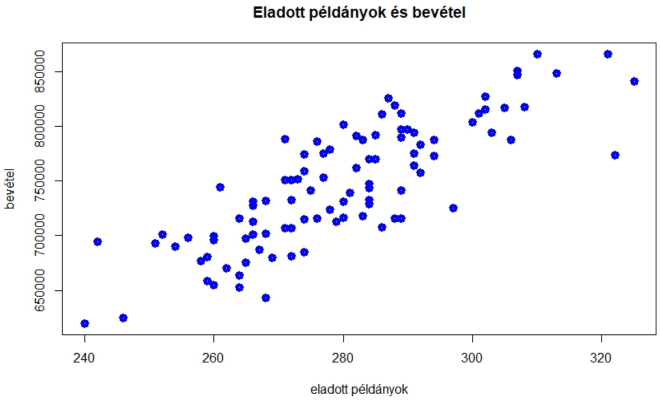
Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X, a B újságból eladott példányok száma Y. Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y-é 180. Az A újság ára 300 forint, a B-é **4000**. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$
\begin{aligned}
\operatorname{cov}(X+Y, 300X+4000Y) &= 300\cdot 100 + 4000\cdot 180 = 750000;\\
D(X+Y) &= \sqrt{D^2(X)+D^2(Y)} = \sqrt{100+180} = 16,73;\\
D(300X+4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} =\\
&= \sqrt{300^2\cdot 100 + 4000^2\cdot 180} = 53749,42;\\
R(X+Y, 300X+4000Y) &= \frac{\operatorname{cov}(X+Y, 300X+4000Y)}{D(X+Y)D(300X+4000Y)} = \frac{750000}{16,73\cdot 53749,42}\\
&= 0,083.
\end{aligned}
$$

A korrelációs együttható értéke majdnem 0, azaz nincs jelentős korreláció az eladott példányok száma és a bevétel között, ha az újságok ára nagyon eltérő.

Korrelációs együttható: példa



A bevétel $(300X + 4000Y)$ és az eladott példányszám $(X + Y)$ együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 750000, korrelációs együttható: 0,083.

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- vitorlázással töltött idő és egészség:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- vitorlázással töltött idő és egészség:

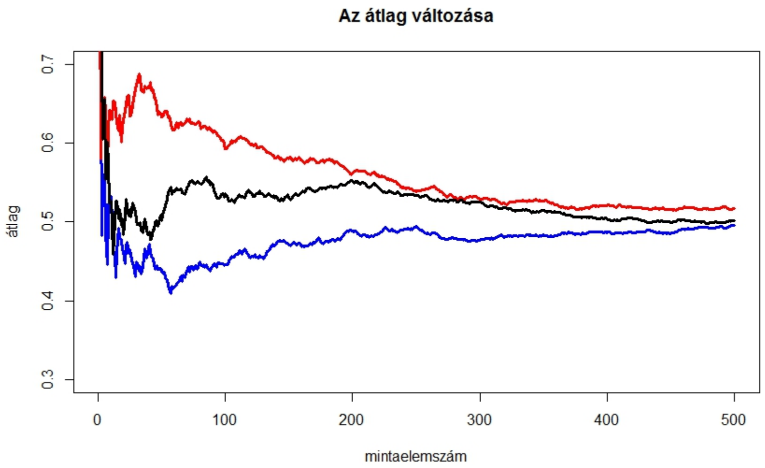
  ha van is pozitív korreláció, **nem biztos, hogy van ok-okozati összefüggés**, a vitorlázás összefügg az anyagi helyzettel, ami az egészséggel, illetve aki beteg, kevesbé tud vitorlázni, vagyis csak a vitorlázástól nem biztos, hogy egészséges lesz valaki
- USA által tudományra és technológiára költött pénz és öngyilkosságok:

Korreláció és ok-okozat

- napsütéses órák száma és hőmérséklet: pozitív korreláció, **van ok-okozati összefüggés**
- napsütéses órák száma és hómennyiség: negatív korreláció, van ok-okozati összefüggés
- anyagi helyzet és iskolai végzettség: van pozitív korreláció, de mindkét irányban lehet ok-okozati összefüggés
- vitorlázással töltött idő és egészség:

  ha van is pozitív korreláció, **nem biztos, hogy van ok-okozati összefüggés**, a vitorlázás összefügg az anyagi helyzettel, ami az egészséggel, illetve aki beteg, kevesbé tud vitorlázni, vagyis csak a vitorlázástól nem biztos, hogy egészséges lesz valaki
- USA által tudományra és technológiára költött pénz és öngyilkosságok: nagyon erős pozitív korreláció ($R = 0,9979$), de **feltehetően nincs ilyen erős ok-okozati összefüggés**
  (forrás és további példák: http://tylervigen.com/spurious-correlations)

Az átlag konvergenciája



A $[0,1]$ intervallumon egyenletes eloszlásból vett mintából az első n elem átlaga $n = 500$-ig

A nagy számok gyenge törvénye

Legyenek $X_1, \ldots, X_n$ független azonos eloszlású véges szórású valószínűségi változók. Legyen $m = \mathbb{E}(X_1)$ és $\sigma = D(X_1)$.

A korábbiak szerint

$$\mathbb{E}(\overline{X}) = m; \qquad D^2(\overline{X}) = \frac{\sigma^2}{n} \to 0 \qquad (n \to \infty).$$

A nagy számok gyenge törvénye

Legyenek $X_1, \dots, X_n$ független azonos eloszlású véges szórású valószínűségi változók. Legyen $m = \mathbb{E}(X_1)$ és $\sigma = D(X_1)$.

A korábbiak szerint

$$\mathbb{E}(\overline{X}) = m; \qquad D^2(\overline{X}) = \frac{\sigma^2}{n} \to 0 \qquad (n \to \infty).$$

A Csebisev-egyenlőtlenség szerint minden $\varepsilon > 0$-ra

$$\mathbb{P}(|\overline{X} - m| > \varepsilon) \leq \frac{D^2(\overline{X})}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \to 0 \quad (n \to \infty).$$

Tehát minden $\varepsilon > 0$-ra

$$\mathbb{P}(|\overline{X} - m| > \varepsilon) \to 0$$

teljesül $n \to \infty$ esetén.

Azaz: $\overline{X} \to m = \mathbb{E}(X_1)$ sztochasztikusan.

A nagy számok törvénye

Tétel (A nagy számok gyenge törvénye)

Legyenek $X_1, X_2, \ldots$ *olyan valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel, hogy* $D(X_1) < \infty$*. Ekkor minden* $\varepsilon > 0$ *esetén*

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

azaz $\overline{X}_n \to \mathbb{E}(X_1)$ *sztochasztikusan.*

A nagy számok törvénye

Tétel (A nagy számok gyenge törvénye)

Legyenek $X_1, X_2, \ldots$ olyan valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel, hogy $D(X_1) < \infty$. Ekkor minden $\varepsilon > 0$ esetén

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

azaz $\overline{X}_n \to \mathbb{E}(X_1)$ sztochasztikusan.

Tétel (A nagy számok erős törvénye)

Legyenek $X_1, X_2, \ldots$ valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel még, hogy $m = \mathbb{E}(X_1) < \infty$. Ekkor

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1) = m$$

teljesül 1 valószínűséggel $n \to \infty$ esetén.

A második esetben gyengébb feltevésből erősebb állítás következik.