

Eloszlásfüggvény (9. előadás)

Definíció (Eloszlásfüggvény)

Legyen $X : \Omega \rightarrow \mathbb{R}$ valószínűségi változó. Ekkor X eloszlásfüggvénye az alábbi $F : \mathbb{R} \rightarrow [0, 1]$ függvény:

$$F(t) = \mathbb{P}(X \leq t)$$

minden $t \in \mathbb{R}$ valós számra.

Definíció (Abszolút folytonosság és sűrűségfüggvény)

Az X valószínűségi változó **abszolút folytonos**, ha van olyan $f : \mathbb{R} \rightarrow \mathbb{R}$ függvény, melyre

$$\mathbb{P}(X \leq t) = \int_{-\infty}^t f(s) ds$$

teljesül minden $t \in \mathbb{R}$ számra. Ilyenkor az f függvényt az X valószínűségi változó **sűrűségfüggvényének** nevezzük.

Eloszlásfüggvény (9. előadás)

Definíció (Abszolút folytonosság és sűrűségfüggvény)

Az X valószínűségi változó **abszolút folytonos**, ha van olyan $f : \mathbb{R} \rightarrow \mathbb{R}$ függvény, melyre

$$\mathbb{P}(X \leq t) = \int_{-\infty}^t f(s) ds$$

teljesül minden $t \in \mathbb{R}$ számra. Ilyenkor az f függvényt az X valószínűségi változó **sűrűségfüggvényének** nevezzük.

Állítás

Legyen az X abszolút folytonos valószínűségi változó, melynek sűrűségfüggvénye f . Ekkor tetszőleges $a < b$ számokra teljesül, hogy

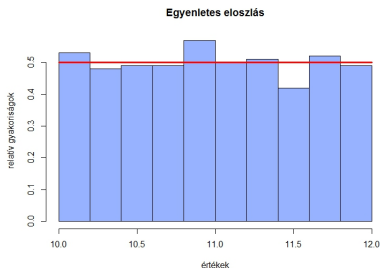
$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(s) ds.$$

Egyenletes eloszlás (uniform distribution)

Definíció (Egyenletes eloszlás)

Legyenek $a < b$ valós számok. Azt mondjuk, hogy az X valószínűségi változó **egyenletes eloszlású** az $[a, b]$ intervallumon, ha sűrűségfüggvénye

$$f(s) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq s \leq b; \\ 0, & \text{különben.} \end{cases}$$



$U(10, 12)$ sűrűségfüggvény és 500 elemű minta hisztogramja

Az egyenletes eloszlás várható értéke

Az $[a, b]$ intervallumon egyenletes eloszlás sűrűségfüggvénye:

$$f(s) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq s \leq b; \\ 0, & \text{különben.} \end{cases}$$

Ha X egyenletes eloszlású az $[a, b]$ intervallumon, akkor várható értéke

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} s \cdot f(s) ds = \int_a^b s \cdot \frac{1}{b-a} ds = \frac{1}{b-a} \left[\frac{s^2}{2} \right]_{s=a}^b = \\ &= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}, \end{aligned}$$

hiszen az x függvény primitív függvénye $\frac{x^2}{2}$, és $b^2 - a^2 = (b-a)(b+a)$.

Az egyenletes eloszlás szórása

Az $[a, b]$ intervallumon egyenletes eloszlás sűrűségfüggvénye:

$$f(s) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq s \leq b; \\ 0, & \text{különben.} \end{cases}$$

Ha X egyenletes eloszlású az $[a, b]$ intervallumon, akkor várható értéke

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{-\infty}^{\infty} s^2 \cdot f(s) ds = \int_a^b s^2 \cdot \frac{1}{b-a} ds = \frac{1}{b-a} \left[\frac{s^3}{3} \right]_{s=a}^b \\ &= \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}, \end{aligned}$$

hiszen az x^2 függvény primitív függvénye $\frac{x^3}{3}$, és $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$.

Az egyenletes eloszlás szórása

Az $[a, b]$ intervallumon egyenletes eloszlás sűrűségfüggvénye:

$$f(s) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq s \leq b; \\ 0, & \text{különben.} \end{cases}$$

Ha X egyenletes eloszlású az $[a, b]$ intervallumon, akkor várható értéke

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{-\infty}^{\infty} s^2 \cdot f(s) ds = \int_a^b s^2 \cdot \frac{1}{b-a} ds = \frac{1}{b-a} \left[\frac{s^3}{3} \right]_{s=a}^b \\ &= \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}, \end{aligned}$$

hiszen az x^2 függvény primitív függvénye $\frac{x^3}{3}$, és $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$.

$$\begin{aligned} D^2(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 \\ &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{a^2 - 2ab + b^2}{12}. \end{aligned}$$

Normális eloszlás

Definíció (Normális eloszlás)

Legyen m valós, σ pedig pozitív szám. Azt mondjuk, hogy az Y valószínűségi változó **normális eloszlású** m várható értékkel és σ^2 szórásnégyzettel, ha sűrűségfüggvénye

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (x \in \mathbb{R}).$$

Jelölése: $Y \sim N(m, \sigma^2)$.

Normális eloszlás

Definíció (Normális eloszlás)

Legyen m valós, σ pedig pozitív szám. Azt mondjuk, hogy az Y valószínűségi változó **normális eloszlású** m várható értékkel és σ^2 szórásnégyzettel, ha sűrűségfüggvénye

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (x \in \mathbb{R}).$$

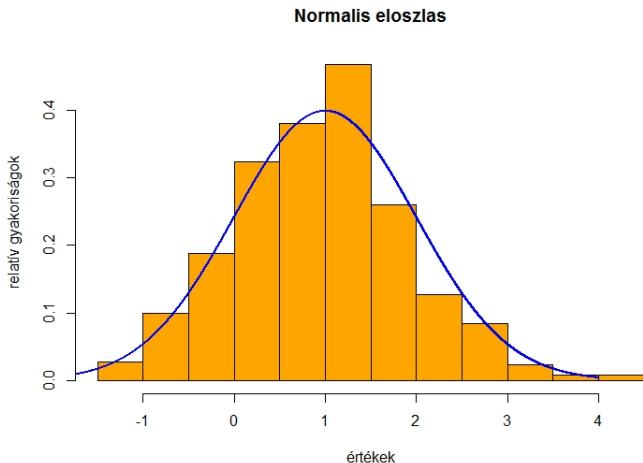
Jelölése: $Y \sim N(m, \sigma^2)$.

Ha $Y \sim N(m, \sigma^2)$, akkor $\mathbb{E}(Y) = m$, $D(Y) = \sigma$.

Standard normális eloszlás: $m = 0$ várható értékű és $\sigma = 1$ szórájú normális eloszlás. Eloszlásfüggvénye: Φ , sűrűségfüggvénye:

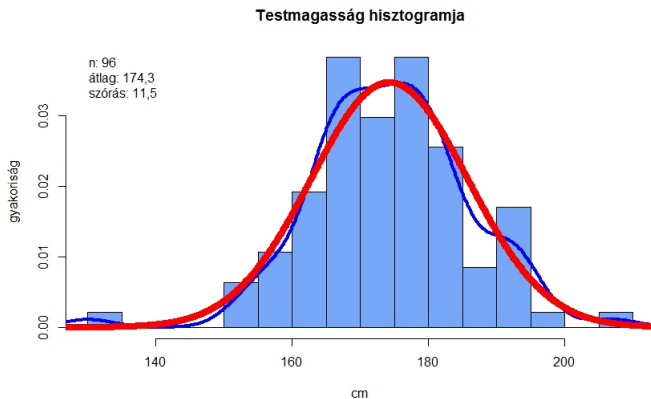
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Normális eloszlás



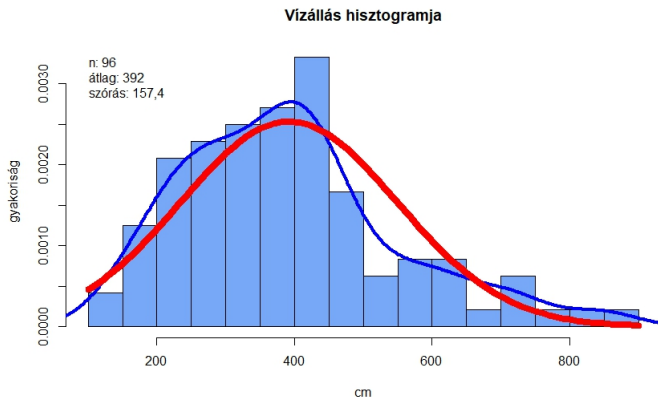
Normális eloszlás ($m = 1, \sigma = 1$) sűrűségfüggvénye és 500 elemű minta hisztogramja

Testmagasság



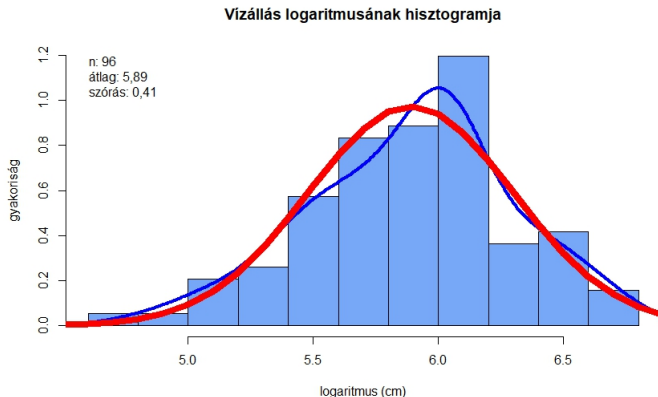
Testmagasság histogramja $n = 96$ elemű mintából, és az $\bar{X} = 174,3$ várható értékű és $s_n^* = 11,5$ szórású normális eloszlás sűrűségfüggvénye (pirossal)

A Duna vízállása



A Duna havi legnagyobb vízállásának histogramja (2002–2009, $n = 96$, forrás: Országos Vízelző Szolgálat), és az $\bar{X} = 392$ várható értékű és $s_n^* = 157,4$ szórású normális eloszlás sűrűségfüggvénye (piros): nem illeszkedik jól a normális eloszlás

A Duna vízállása



A Duna havi legnagyobb vízállásának **logaritmusának** hisztogramja (2002–2009, $n = 96$, forrás: Országos Vízeljáró Szolgálat), és az $\bar{X} = 5,89$ várható értékű és $s_n^* = 0,41$ szórású normális eloszlás sűrűségfüggvénye: jobban illeszkedik, a vízállás logaritmusa lehet normális eloszlású (vagyis a vízállás lehet **lognormális eloszlású**)

Standard normális eloszlás

Az X valószínűségi változó **standard normális eloszlású**, ha sűrűségfüggvénye

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Ekkor eloszlásfüggvénye Φ , azaz

$$\Phi(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X < t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Standard normális eloszlás

Az X valószínűségi változó **standard normális eloszlású**, ha sűrűségfüggvénye

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Ekkor eloszlásfüggvénye Φ , azaz

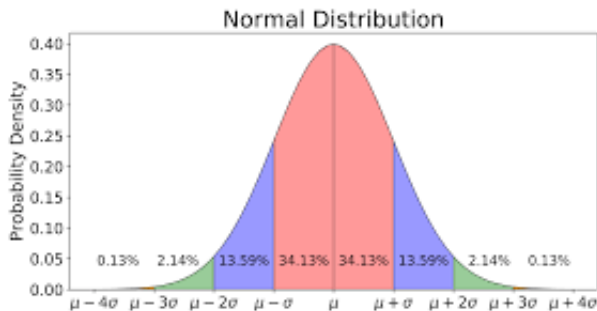
$$\Phi(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X < t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Továbbá

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = \\ &= \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(b) - \Phi(a). \end{aligned}$$

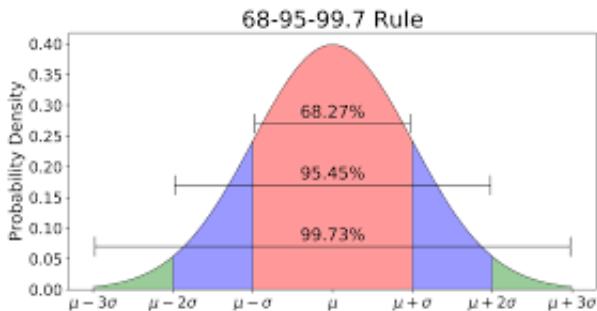
A Φ függvény az R -ben: pnorm

Normális eloszlás



A normális eloszlás sűrűségfüggvénye és annak valószínűsége, hogy a valószínűségi változó az $(m-4\sigma, m-3\sigma)$, $(m-3\sigma, m-2\sigma)$, \dots , $(m+3\sigma, m+4\sigma)$ intervallumokba esik (forrás: towardsdatascience.com)

Normális eloszlás



A normális eloszlás sűrűségfüggvénye és annak valószínűsége, hogy a valószínűségi változó az $(\mu - 3\sigma, \mu + 3\sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$, $(\mu - \sigma, \mu + \sigma)$ intervallumokba esik (forrás: towardsdatascience.com)

A normális eloszlás tulajdonságai

Tegyük fel, hogy Y normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, azaz $Y \sim N(m, \sigma^2)$. Ekkor tetszőleges $a \leq b$ valós számokra

$$\mathbb{P}(Y < b) = \mathbb{P}(Y \leq b) = \mathbb{P}\left(\frac{Y - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = \Phi\left(\frac{b - m}{\sigma}\right)$$

A normális eloszlás tulajdonságai

Tegyük fel, hogy Y normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, azaz $Y \sim N(m, \sigma^2)$. Ekkor tetszőleges $a \leq b$ valós számokra

$$\mathbb{P}(Y < b) = \mathbb{P}(Y \leq b) = \mathbb{P}\left(\frac{Y - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = \Phi\left(\frac{b - m}{\sigma}\right)$$

$$\mathbb{P}(Y > a) = \mathbb{P}(Y \geq a) = 1 - \Phi\left(\frac{a - m}{\sigma}\right)$$

A normális eloszlás tulajdonságai

Tegyük fel, hogy Y normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, azaz $Y \sim N(m, \sigma^2)$. Ekkor tetszőleges $a \leq b$ valós számokra

$$\mathbb{P}(Y < b) = \mathbb{P}(Y \leq b) = \mathbb{P}\left(\frac{Y - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = \Phi\left(\frac{b - m}{\sigma}\right)$$

$$\mathbb{P}(Y > a) = \mathbb{P}(Y \geq a) = 1 - \Phi\left(\frac{a - m}{\sigma}\right)$$

$$\mathbb{P}(a < Y < b) = \mathbb{P}(a \leq Y \leq b) = \Phi\left(\frac{b - m}{\sigma}\right) - \Phi\left(\frac{a - m}{\sigma}\right)$$

A normális eloszlás tulajdonságai

Tegyük fel, hogy Y normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, azaz $Y \sim N(m, \sigma^2)$. Ekkor tetszőleges $a \leq b$ valós számokra

$$\mathbb{P}(Y < b) = \mathbb{P}(Y \leq b) = \mathbb{P}\left(\frac{Y - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = \Phi\left(\frac{b - m}{\sigma}\right)$$

$$\mathbb{P}(Y > a) = \mathbb{P}(Y \geq a) = 1 - \Phi\left(\frac{a - m}{\sigma}\right)$$

$$\mathbb{P}(a < Y < b) = \mathbb{P}(a \leq Y \leq b) = \Phi\left(\frac{b - m}{\sigma}\right) - \Phi\left(\frac{a - m}{\sigma}\right)$$

$$\mathbb{P}(|Y - m| \leq b) = 2\Phi\left(\frac{b}{\sigma}\right) - 1$$

Példa: normális eloszlás

Példa. Tegyük fel, hogy az Y valószínűségi változó normális eloszlású $m = 4$ várható értékkel és $\sigma = 3$ szórással. Ekkor

$$\mathbb{P}(Y \leq 7) = \Phi\left(\frac{7 - m}{\sigma}\right) = \Phi\left(\frac{7 - 4}{3}\right) = \Phi(1) = 0,84.$$

Példa: normális eloszlás

Példa. Tegyük fel, hogy az Y valószínűségi változó normális eloszlású $m = 4$ várható értékkel és $\sigma = 3$ szórással. Ekkor

$$\mathbb{P}(Y \leq 7) = \Phi\left(\frac{7 - m}{\sigma}\right) = \Phi\left(\frac{7 - 4}{3}\right) = \Phi(1) = 0,84.$$

$$\begin{aligned}\mathbb{P}(1 < Y \leq 7) &= \mathbb{P}(Y \leq 7) - \mathbb{P}(Y \leq 1) = \Phi\left(\frac{7 - m}{\sigma}\right) - \Phi\left(\frac{1 - m}{\sigma}\right) \\ &= \Phi\left(\frac{7 - 4}{3}\right) - \Phi\left(\frac{1 - 4}{3}\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0,68,\end{aligned}$$

mert

$$\Phi(-x) = 1 - \Phi(x).$$

A normális eloszlás tulajdonságai

Állítás

Legyen Y normális eloszlású valószínűségi változó m várható értékkel és σ szórással, azaz $Y \sim N(m, \sigma^2)$. Ekkor az $aY + b$ valószínűségi változó normális eloszlású $am + b$ várható értékkel és $a^2\sigma^2$ szórásnégyzettel.

Állítás

Ha Y_1, Y_2 **független, normális eloszlású** valószínűségi változók, akkor $Y_1 + Y_2$ is normális eloszlású, várható értéke $m_1 + m_2$, szórásnégyzete $\sigma_1^2 + \sigma_2^2$ (ahol $Y_j \sim N(m_j, \sigma_j^2)$).

Példa. Ha Y és Z függetlenek, normális eloszlásúak, $Y \sim N(2, 3^2)$ és $Z \sim N(1, 4^2)$, akkor

$$Y + Z \sim N(3, 5^2); \quad Y - Z \sim N(1, 5^2); \quad Y + 3Z \sim N(5, 57).$$

A normális eloszlás tulajdonságai

Állítás

Legyenek Y_1, Y_2, \dots, Y_n **független normális eloszlású** valószínűségi változók, melyek várható értéke m , szórásuk σ . Ekkor az összegük és **az átlaguk is normális eloszlású**, és

$$Y_1 + Y_2 + \dots + Y_n \sim N(nm, n\sigma^2); \quad \bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

A normális eloszlás tulajdonságai

Állítás

Legyenek Y_1, Y_2, \dots, Y_n **független normális eloszlású** valószínűségi változók, melyek várható értéke m , szórásuk σ . Ekkor az összegük és **az átlaguk is normális eloszlású**, és

$$Y_1 + Y_2 + \dots + Y_n \sim N(nm, n\sigma^2); \quad \bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

Példa. Tegyük fel, hogy az emberek testmagassága 176 cm várható értékű és 7 szórású valószínűségi változó. Ekkor

- 100 ember testmagasságának átlaga szintén normális eloszlású, 176 várható értékkel és $7/\sqrt{100} = 0,7$ szórással;
- 10000 ember testmagasságának átlaga normális eloszlású, 176 várható értékkel és $7/\sqrt{10000} = 0,07$ szórással.

Házi feladat november 27-ig

Számítógép segítségével (például R-ben) generáljunk 1000 elemű mintát (1000 független valószínűségi változót) normális eloszlásból (a várható érték legyen tetszőlegesen választott szám 2 és 6 között, a szórás pedig szintén tetszőlegesen választott 1 és 3 között).

- 1 Készítsünk hisztogramot az első 10, 100, 1000 mintaelemből (három különböző ábrán).
- 2 Mennyi annak valószínűsége, hogy egy mintaelem értéke 2 és 6 közé esik?
- 3 Ábrázoljuk n függvényében azt, hogy a mintaelemek hányadrésze esik 2 és 6 közé, ha az első n mintaelemet tekintjük ($n = 1, 2, \dots, 1000$).

Házi feladat november 13-ig: megoldás

Tegyük fel, hogy három távoli város mindegyikében a többtől függetlenül minden nap p valószínűséggel van földrengés. Legyen X, Y, Z az, hogy az egyes városokban mostantól hányadik napon lesz először földrengés. Számítsuk ki az alábbi mennyiségeket, és ábrázoljuk őket p függvényében (például R-ben):

❶ $\text{cov}(X + Y, Z - X)$

❷ $R(X + Y, Z - X)$

❸ $R(2X + Y, 2X + Z)$

Az X, Y, Z valószínűségi változók függetlenek, geometriai eloszlásúak p paraméterrel:

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) = \mathbb{P}(Z = k) = (1 - p)^{k-1}p \quad (k = 0, 1, 2, \dots)$$

Házi feladat november 13-ig: megoldás

Az X, Y, Z valószínűségi változók függetlenek, geometriai eloszlásúak p paraméterrel:

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) = \mathbb{P}(Z = k) = (1 - p)^{k-1}p \quad (k = 0, 1, 2, \dots)$$

$$\begin{aligned} \text{cov}(X + Y, Z - X) &= \text{cov}(X, Z) + \text{cov}(X, -X) + \text{cov}(Y, Z) + \text{cov}(Y, -X) \\ &= -D^2(X) = -\frac{1 - p}{p^2}, \end{aligned}$$

felhasználva, hogy a kovariancia lineáris, független valószínűségi változók kovarianciája 0, egy valószínűségi változó saját magával vett kovarianciája a szórásnégyzete, illetve alkalmazva a geometriai eloszlás szórására vonatkozó képletet.

Házi feladat november 13-ig: megoldás

$$\text{cov}(X + Y, Z - X) = -D^2(X) = -\frac{1 - p}{p^2}$$

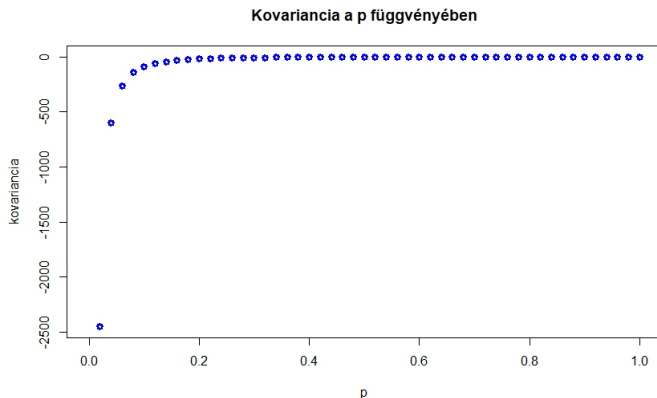
```
kov <- function(p){-(1 - p)/p^2}
```

```
x=seq(from=0, to=1, by=0.02)
```

```
y=kov(x)
```

```
plot(y~x, col="blue", lwd="3", main="Kovariancia a p függvényében", xlab="p",  
ylab="kovariancia")
```

Házi feladat november 13-ig: megoldás



$$\text{cov}(X + Y, Z - X) = -D^2(X) = -\frac{1-p}{p^2} \text{ ábrázolása } p \text{ függvényében}$$

Házi feladat november 13-ig: megoldás

Az X, Y, Z valószínűségi változók függetlenek, geometriai eloszlásúak p paraméterrel.

$$\text{cov}(X + Y, Z - X) = -D^2(X) = -\frac{1 - p}{p^2}$$

Mivel X, Y, Z függetlenek, és szórásnégyzetük megegyezik:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{2D^2(X)} = \sqrt{2} \cdot D(X)$$

$$D(Z - X) = \sqrt{D^2(Z) + D^2(-X)} = \sqrt{2D^2(X)} = \sqrt{2} \cdot D(X)$$

$$R(X + Y, Z - X) = \frac{\text{cov}(X + Y, Z - X)}{D(X + Y)D(Z - X)} = \frac{-D^2(X)}{\sqrt{2} \cdot D(X) \cdot \sqrt{2} \cdot D(X)} = -\frac{1}{2}.$$

Házi feladat november 13-ig: megoldás

Az X, Y, Z valószínűségi változók függetlenek, geometriai eloszlásúak p paraméterrel.

Az előzőhöz hasonlóan

$$\text{cov}(2X + Y, 2X + Z) = \text{cov}(2X, 2X) = 4D^2(X)$$

Mivel X, Y, Z függetlenek, és szórásnégyzetük megegyezik:

$$D(2X + Y) = \sqrt{D^2(2X) + D^2(Y)} = \sqrt{5D^2(X)} = \sqrt{5} \cdot D(X)$$

$$D(2X + Z) = \sqrt{D^2(2X) + D^2(Z)} = \sqrt{5D^2(X)} = \sqrt{5} \cdot D(X)$$

$$R(2X + Y, 2X + Z) = \frac{\text{cov}(2X + Y, 2X + Z)}{D(2X + Y)D(2X + Z)} = \frac{4D^2(X)}{\sqrt{5} \cdot D(X) \cdot \sqrt{5} \cdot D(X)} = \frac{4}{5}$$

Házi feladat november 13-ig: megoldás

$$R(X + Y, Z - X) = -\frac{1}{2}; \quad R(2X + Y, 2X + Z) = \frac{4}{5}.$$

Ezek a korrelációs együtthatók nem függenek p -től, minden p -re ugyanaz az érték adódik.

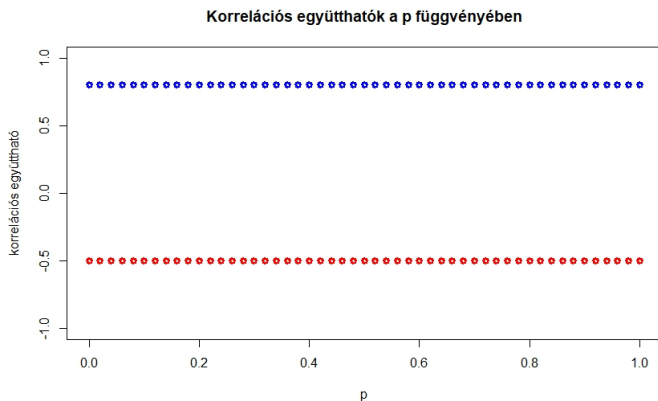
```
z=rep(-0.5, 51)
```

```
u=rep(0.8, 51)
```

```
plot(u~x, col="blue", lwd="3", main="Korrelációs együtthatók a p függvényében", xlab="p", ylab="korrelációs együttható", ylim=c(-1,1), xlim=c(0,1))
```

```
lines(z~x, col="red", lwd="3", type="p")
```

Házi feladat november 13-ig: megoldás



$R(X + Y, Z - X) = -\frac{1}{2}$ (piros) és $R(2X + Y, 2X + Z) = \frac{4}{5}$ (kék) ábrázolása p függvényében.