

## Normális eloszlás: definíció (12. előadás)

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

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

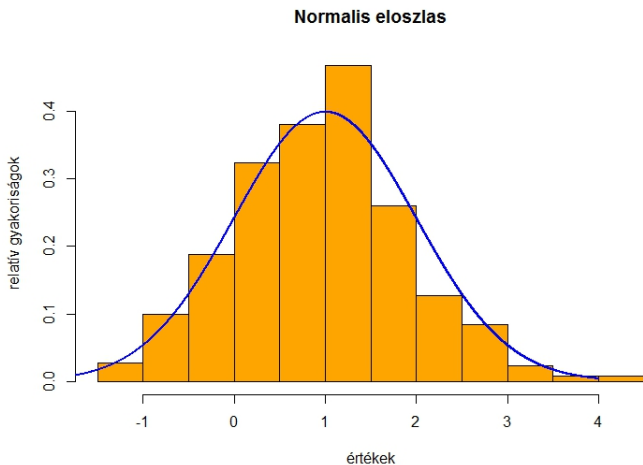
Jelölése:  $Y \sim N(m, \sigma^2)$ .

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

**Standard normális eloszlás:** az  $m = 0$  várható értékű és  $\sigma = 1$  szórájú normális eloszlás. **Eloszlásfüggvénye:**  $\Phi$ , sűrűségfüggvénye  $\varphi$ , ahol

$$\Phi(t) = \int_{-\infty}^t \varphi(x) dx; \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

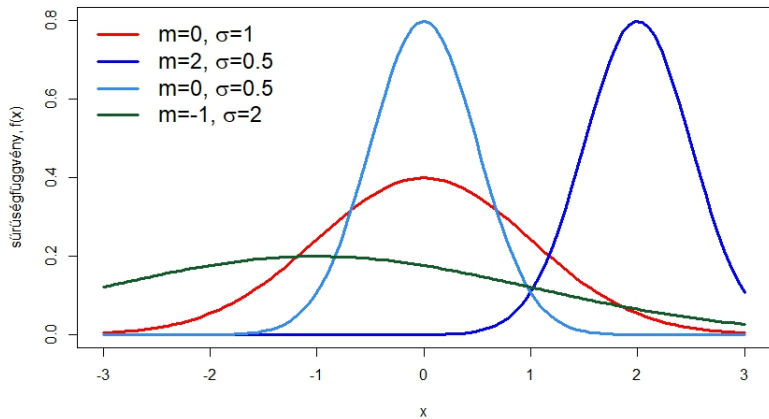
# Normális eloszlás



Normális eloszlás ( $m = 1, \sigma = 1$ ) sűrűségfüggvénye és 500 darab független,  $N(1, 1)$  eloszlású valószínűségi változóból álló minta hisztogramja

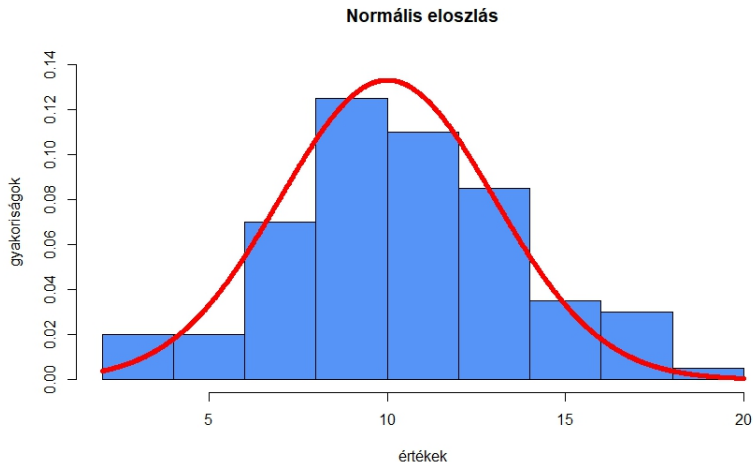
# Normális eloszlás

Normális eloszlások sűrűségfüggvényei



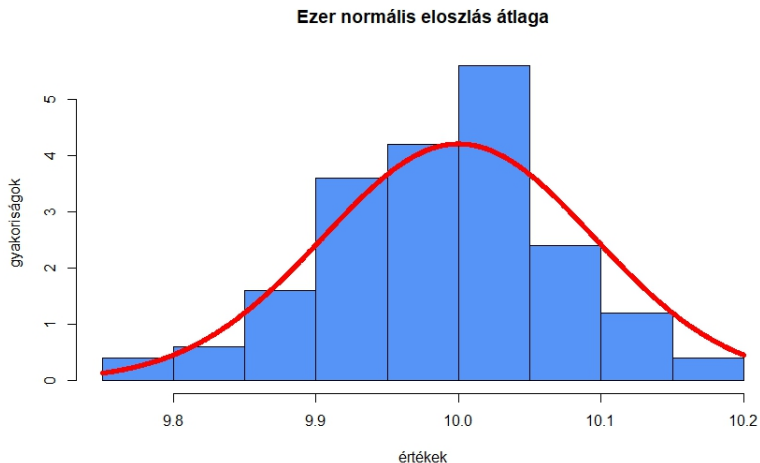
Különböző várható értékű ( $m$ ) és szórású ( $\sigma$ ) normális eloszlások sűrűségfüggvényei

# Normális eloszlás



Száz független normális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény ( $m = 10, \sigma = 3, \bar{x} = 9,88, s_n^* = 2,58$ )

# Normális eloszlások átlaga



Százelemű minta az alábbi eloszlásból:  $n = 1000$  független normális eloszlású ( $m = 10, \sigma = 3$ ) valószínűségi változó átlaga és az  $N(10, 9/1000)$  normális eloszlás sűrűségfüggvénye ( $\bar{x} = 9,99, s_n^* = 0,084, \sigma/\sqrt{n} = 0,095$ )

# Normális eloszlások átlaga

Legyenek  $X, Y$  függetlenek, normális eloszlásúak:  $X \sim N(m_1, \sigma_1^2)$ ,  $Y \sim N(m_2, \sigma_2^2)$ .  
Ekkor a következők igazak:

- $X + b$  eloszlása normális,  $m_1 + b$  várható értékkel és  $\sigma$  szórással;
- $aX$  eloszlása normális  $am_1$  várható értékkel és  $|a|\sigma$  szórással;
- $X + Y$  eloszlása normális,  $m_1 + m_2$  várható értékkel és  $\sqrt{\sigma_1^2 + \sigma_2^2}$  szórással.

Emlékeztető:  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ , és ha  $X$  és  $Y$  függetlenek, akkor  $D^2(X + Y) = D^2(X) + D^2(Y)$ .

# Normális eloszlások átlaga

Legyenek  $X, Y$  függetlenek, normális eloszlásúak:  $X \sim N(m_1, \sigma_1^2)$ ,  $Y \sim N(m_2, \sigma_2^2)$ .  
Ekkor a következők igazak:

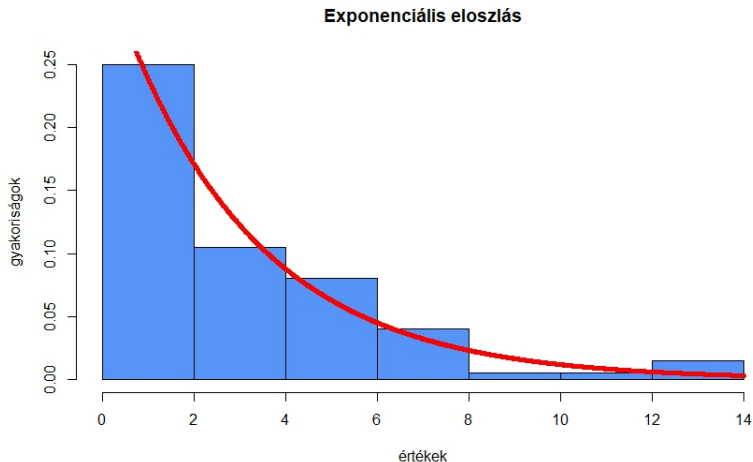
- $X + b$  eloszlása normális,  $m_1 + b$  várható értékkel és  $\sigma$  szórással;
- $aX$  eloszlása normális  $am_1$  várható értékkel és  $|a|\sigma$  szórással;
- $X + Y$  eloszlása normális,  $m_1 + m_2$  várható értékkel és  $\sqrt{\sigma_1^2 + \sigma_2^2}$  szórással.

Emlékeztető:  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ , és ha  $X$  és  $Y$  függetlenek, akkor  $D^2(X + Y) = D^2(X) + D^2(Y)$ .

Ebből következik: ha  $X_1, \dots, X_n$  független normális eloszlásúak  $m$  várható értékkel és  $\sigma$  szórással, akkor

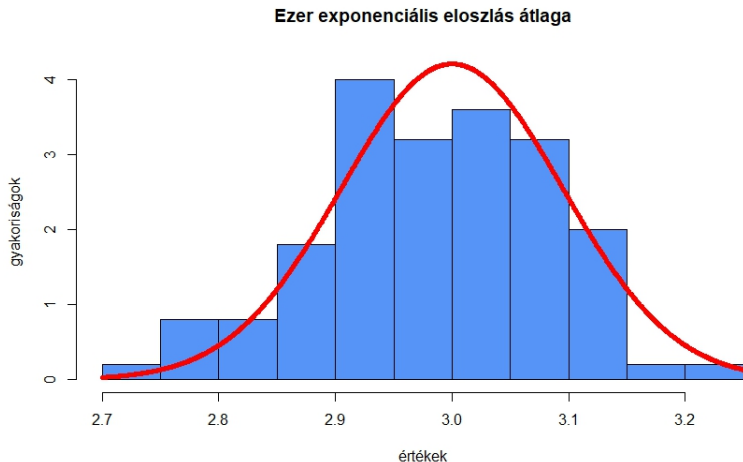
$$\frac{X_1 + \dots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

# Exponenciális eloszlás



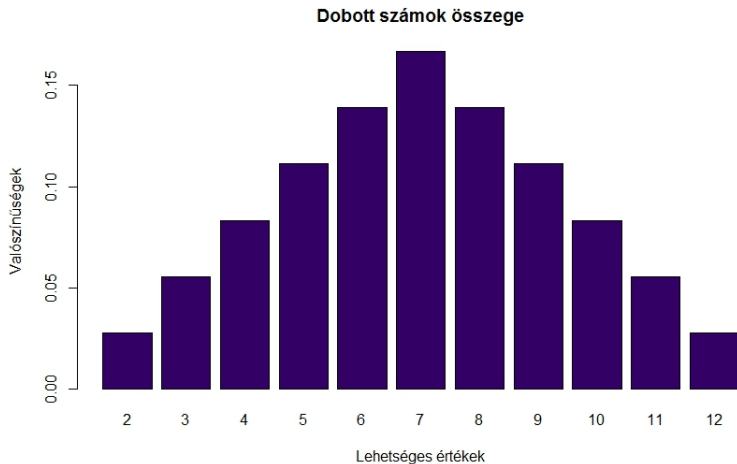
Száz független  $\lambda = 1/3$  paraméterű exponenciális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz  $e^{-1/3}/3$  ( $\mathbb{E}(X) = D(X) = 3, \bar{x} = 3,03, s_n^* = 2,89$ )

# Exponenciális eloszlások átlaga



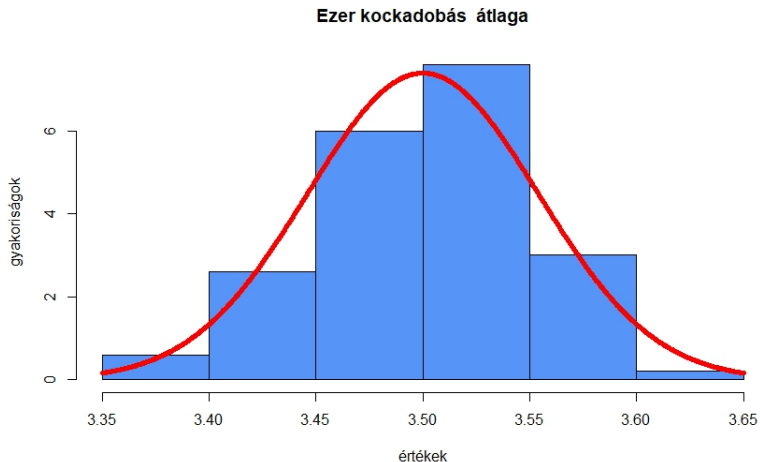
Százelemű minta az alábbi eloszlásból:  $n = 1000$  független exponenciális eloszlású ( $\lambda = 1/3$ ) valószínűségi változó átlaga, és az  $N(3, 9/1000)$  normális eloszlás sűrűségfüggvénye ( $\bar{x} = 2,98, s_n^* = 0,098, \sigma/\sqrt{n} = 0,095$ )

# Két kockadobás összege



Két szabályos kockadobás összegének eloszlása

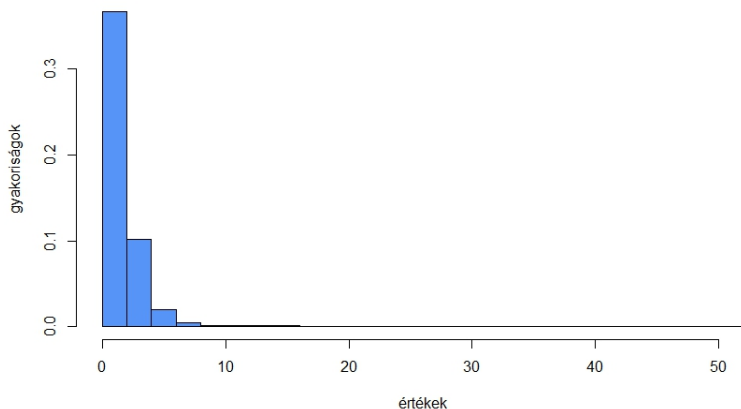
# Kockadobások átlaga



Százelemű minta az alábbi eloszlásból:  $n = 1000$  független szabályos kockadobás átlaga, és az  $N(3,5, D^2(X_1)/1000)$  normális eloszlás sűrűségfüggvénye ( $\bar{x} = 3,501, s_n^* = 0,098, \sigma/\sqrt{n} = 0,051$ )

# Exponenciális eloszlás a kitevőben

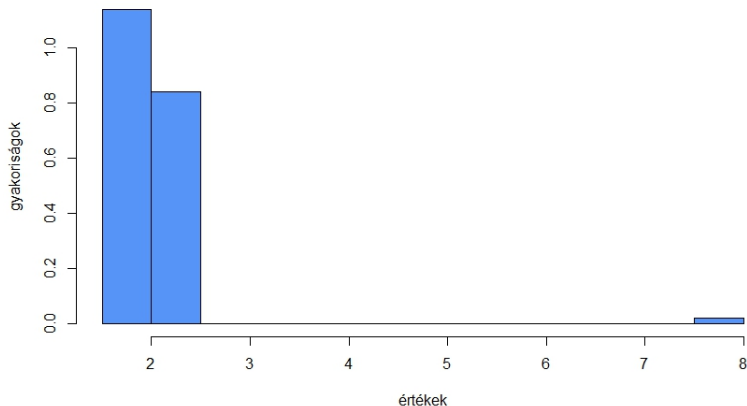
Exponenciális eloszlás a kitevőben



$e^{X_1}, e^{X_2}, \dots, e^{X_{1000}}$  hisztogramja, ahol  $X_i$ -k függetlenek, 2 paraméterű exponenciális eloszlásúak ( $\mathbb{E}(e^{X_1}) = 2, D(e^{X_1}) = \infty, \bar{x} = 1,99, s_n^* = 2,33$ )

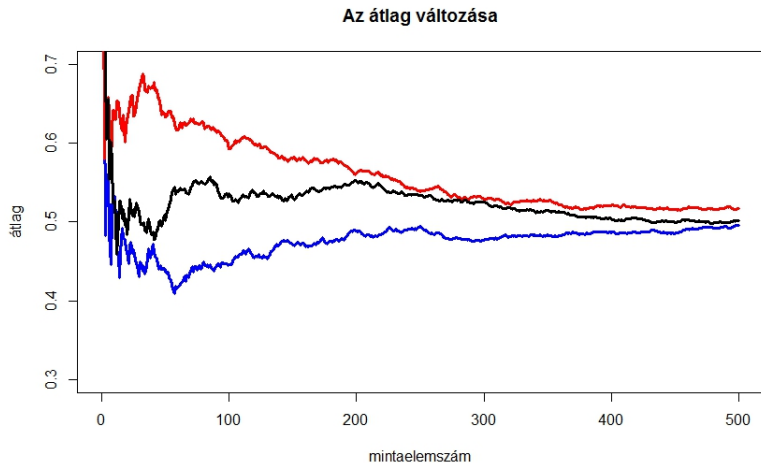
# Exponenciális eloszlás a kitevőben

Ezer exponenciális eloszlás átlaga



Százelemű minta az alábbi eloszlásból:  $e^{X_1}, e^{X_2}, \dots, e^{X_{1000}}$  átlaga, ahol  $X_i$ -k függetlenek, 2 paraméterű exponenciális eloszlásúak. Itt  $e^{X_i}$  várható értéke véges, de szórása végtelen.

# Az átlag konvergenciája



A  $[0, 1]$  intervallumon egyenletes eloszlásból vett minta átlaga  $n = 500$ -ig

# A nagy számok törvényei

## Tétel (A nagy számok gyenge törvénye)

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

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

azaz  $\bar{X}_n \rightarrow \mathbb{E}(X_1)$  sztochasztikusan.

# A nagy számok törvényei

## Tétel (A nagy számok gyenge törvénye)

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

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| > \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

*azaz  $\bar{X}_n \rightarrow \mathbb{E}(X_1)$  sztochasztikusan.*

## Tétel (A nagy számok erős törvénye)

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

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mathbb{E}(X_1) = m$$

*teljesül 1 valószínűséggel  $n \rightarrow \infty$  esetén.*

A második esetben gyengébb feltevésből erősebb állítás következik.

# Centrális határeloszlástétel

## Tétel (Centrális határeloszlástétel)

Legyenek  $X_1, X_2, \dots$  **független azonos eloszlású** valószínűségi változók, melyekre  $\mathbb{E}(X_1) = m$  és  $D(X_1) = \sigma < \infty$ , azaz **szórásuk véges**. Ekkor tetszőleges  $t$  valós számra

$$\mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma\sqrt{n}} \leq t\right) \rightarrow \mathbb{P}(Z \leq t) \quad (n \rightarrow \infty),$$

ahol  $Z$  standard normális eloszlású, azaz

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

Ezt úgy is fogalmazhatjuk, hogy

$$\frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma\sqrt{n}} \rightarrow N(0, 1)$$

teljesül  $n \rightarrow \infty$  esetén eloszlásban. Azonos eloszlású:  $\mathbb{P}(X_i \leq t) = P(X_j \leq t)$  minden  $i, j$  párra és  $t$  valós számra

# Centrális határeloszlástétel

Legyenek  $X_1, X_2, \dots$  független azonos eloszlású valószínűségi változók, melyekre  $\mathbb{E}(X_1) = m$  és  $D(X_1) = \sigma < \infty$ . Ekkor

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma \sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

A határértéket  $\Phi(b) - \Phi(a) = \mathbb{P}(a \leq Y \leq b)$  alakban is írhatjuk, ahol  $Y \sim N(0, 1)$ .

# Centrális határeloszlástétel

Legyenek  $X_1, X_2, \dots$  független azonos eloszlású valószínűségi változók, melyekre  $\mathbb{E}(X_1) = m$  és  $D(X_1) = \sigma < \infty$ . Ekkor

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

A határértéket  $\Phi(b) - \Phi(a) = \mathbb{P}(a \leq Y \leq b)$  alakban is írhatjuk, ahol  $Y \sim N(0, 1)$ .

Így is átfogalmazható a tétel állítása:

$$\mathbb{P}(nm + a\sigma\sqrt{n} \leq X_1 + X_2 + \dots + X_n < nm + b\sigma\sqrt{n}) \rightarrow \Phi(b) - \Phi(a).$$

Ez azt jelenti, hogy az  $\bar{X}_n$  átlag eloszlása közel van egy  $m$  várható értékű,  $\sigma/\sqrt{n}$  szórású normális eloszláshoz.

# Centrális határeloszlástétel

Legyenek  $X_1, X_2, \dots$  **független** azonos eloszlású valószínűségi változók, melyekre  $\mathbb{E}(X_1) = m$  és  $D(X_1) = \sigma < \infty$ , azaz **szórásuk véges**. Ekkor

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left( a \leq \frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma \sqrt{n}} < b \right) &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = \\ &= \Phi(b) - \Phi(a) = \mathbb{P}(a \leq Z \leq b), \end{aligned}$$

ahol  $Z \sim N(0, 1)$  **standard normális eloszlású**. Tovább alakítva:

# Centrális határeloszlástétel

Legyenek  $X_1, X_2, \dots$  **független** azonos eloszlású valószínűségi változók, melyekre  $\mathbb{E}(X_1) = m$  és  $D(X_1) = \sigma < \infty$ , azaz **szórásuk véges**. Ekkor

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left( a \leq \frac{X_1 + X_2 + \dots + X_n - n \cdot m}{\sigma \sqrt{n}} < b \right) &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = \\ &= \Phi(b) - \Phi(a) = \mathbb{P}(a \leq Z \leq b), \end{aligned}$$

ahol  $Z \sim N(0, 1)$  **standard normális eloszlású**. Tovább alakítva:

$$\mathbb{P}(nm + a\sigma\sqrt{n} \leq X_1 + X_2 + \dots + X_n < nm + b\sigma\sqrt{n}) \rightarrow \mathbb{P}(a \leq Z \leq b).$$

Ha  $n$ -nel osztunk, hogy az átlag jelenjen meg:

$$\mathbb{P} \left( m + a \frac{\sigma}{\sqrt{n}} \leq \frac{X_1 + X_2 + \dots + X_n}{n} < m + b \frac{\sigma}{\sqrt{n}} \right) \rightarrow \mathbb{P}(a \leq Z \leq b).$$

Vagyis az **átlag eloszlása** „közel van” egy  $m$  várható értékű,  $\frac{\sigma}{\sqrt{n}}$  szórású **normális eloszláshoz**.

## Centrális határeloszlástétel: példa

Legyenek  $X_1, X_2, \dots$  független, 2 várható értékű, exponenciális eloszlású valószínűségi változók. Mi a limesze a  $\mathbb{P}(X_1 + \dots + X_n - 2n < 2\sqrt{n})$  mennyiségnek  $n \rightarrow \infty$  esetén?

## Centrális határeloszlástétel: példa

Legyenek  $X_1, X_2, \dots$  független, 2 várható értékű, exponenciális eloszlású valószínűségi változók. Mi a limesze a  $\mathbb{P}(X_1 + \dots + X_n - 2n < 2\sqrt{n})$  mennyiségnek  $n \rightarrow \infty$  esetén?

Mivel a valószínűségi változók **függetlenek**, **azonos eloszlásúak** és **véges szórásúak**, teljesülnek a centrális határeloszlástétel feltételei. Ezért

$$\mathbb{P}(X_1 + \dots + X_n - 2n < 2\sqrt{n}) = \mathbb{P}\left(\frac{X_1 + \dots + X_n - 2n}{2\sqrt{n}} < 1\right) \rightarrow \Phi(1),$$

ha  $n \rightarrow \infty$ , hiszen  $m = 2$  a várható érték, és mivel az eloszlás exponenciális, a várható érték egyenlő a szórással, így  $\sigma = 2$  a szórás.

# Konvergenciafajták

## Definíció

A  $Z_1, Z_2, \dots$ , valószínűségi változókból álló sorozat **eloszlásban konvergál** az  $Z$  valószínűségi változóhoz, ha minden olyan  $t$  számra, melyre  $Z$  eloszlásfüggvénye folytonos  $t$ -ben, teljesül, hogy

$$\mathbb{P}(Z_n \leq t) \rightarrow \mathbb{P}(Z \leq t) \quad (n \rightarrow \infty).$$

# Konvergenciafajták

## Definíció

A  $Z_1, Z_2, \dots$ , valószínűségi változókból álló sorozat **eloszlásban konvergál** az  $Z$  valószínűségi változóhoz, ha minden olyan  $t$  számra, melyre  $Z$  eloszlásfüggvénye folytonos  $t$ -ben, teljesül, hogy

$$\mathbb{P}(Z_n \leq t) \rightarrow \mathbb{P}(Z \leq t) \quad (n \rightarrow \infty).$$

Ha  $Z_n \rightarrow Z$  teljesül 1 valószínűséggel, akkor  $Z_n \rightarrow Z$  sztochasztikusan és eloszlásban is.

Lehetséges, hogy  $Z_n \rightarrow Z$  eloszlásban, de  $Z_n$  nem tart  $Z$ -hez sztochasztikusan (és ezért 1 valószínűséggel sem).

# A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember  $p$  valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb 0,01-gyel tér el  $p$ -től, tetszőleges  $p$  esetén legalább 95% legyen?

## A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember  $p$  valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb 0,01-gyel tér el  $p$ -től, tetszőleges  $p$  esetén legalább 95% legyen?

$n$  megkérdezett, mindenki  $p$  valószínűséggel támogatja a pártot

$X$ : a pártot támogatók száma a megkérdezettek között

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95$$

teljesüljön minden  $0 \leq p \leq 1$ -re.

## A centrális határeloszlástétel alkalmazása

$n$  megkérdezett, mindenki  $p$  valószínűséggel támogatja a pártot

$X$ : a pártot támogatók száma a megkérdezettek között,

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95$$

teljesüljön minden  $0 \leq p \leq 1$ -re, ahol  $X = \sum_{j=1}^n X_j$ , az  $X_j$ -k függetlenek, és

$$\mathbb{P}(X_j = 1) = 1 - \mathbb{P}(X_j = 0) = p; \quad \mathbb{E}(X_j) = p; \quad D(X_j) = \sqrt{p(1-p)}.$$

$$\begin{aligned} \mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) &= \mathbb{P}\left(\left|\frac{\sum_{j=1}^n X_j - np}{\sqrt{np(1-p)}}\right| \leq \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &\approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1. \end{aligned}$$

## A centrális határeloszlástétel alkalmazása

$n$  megkérdezett, mindenki  $p$  valószínűséggel támogatja a pártot

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0,95$$

teljesüljön minden  $0 \leq p \leq 1$ -re. Vagyis mivel  $p(1-p) \leq 1/4$ :

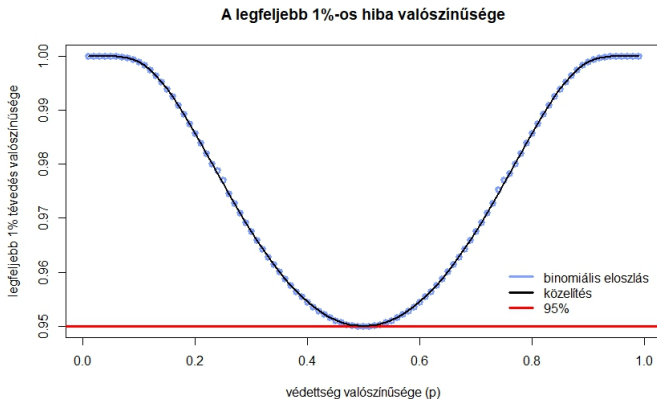
$$\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,975;$$

$$\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}} \geq \Phi^{-1}(0,975) = \text{qnorm}(0,975) = 1,96;$$

$$n \geq p(1-p) \cdot 1,96^2 \cdot \frac{1}{0,01^2};$$

$$n \geq \frac{1}{4} \cdot 1,96^2 \cdot \frac{1}{0,01^2} = 9607.$$

# A hiba valószínűsége



A hiba valószínűsége a  $p$  függvényében

# A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember  $p$  valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb  $0,01$ -gyel tér el  $p$ -től, tetszőleges  $p$  esetén legalább  $95\%$  legyen?

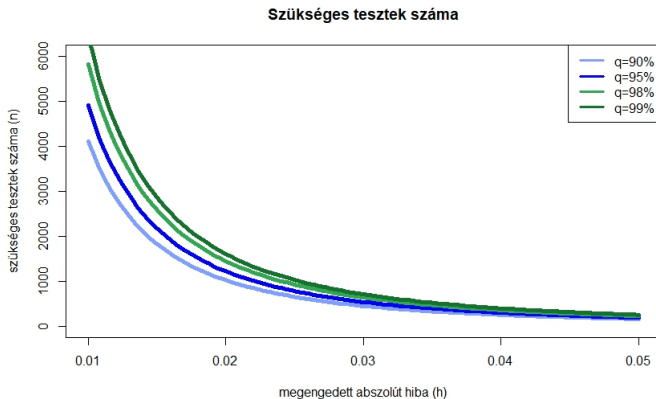
- Csebisev-egyenlőtlenséggel:  $n \geq 50000$  biztosan elég
- centrális határeloszlástétellel közelítve:  $n \geq 9607$  elég

# A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember  $p$  valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb 0,01-gyel tér el  $p$ -től, tetszőleges  $p$  esetén legalább 95% legyen?

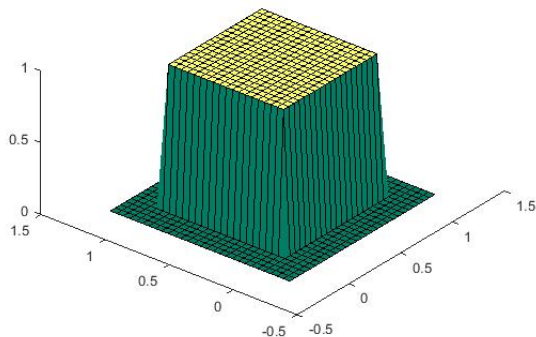
- Csebisev-egyenlőtlenséggel:  $n \geq 50000$  biztosan elég
- centrális határeloszlástétellel közelítve:  $n \geq 9607$  elég
- valójában:  $n = 9607$ ,  $p = 1/2$  esetén 0,94987 adódik a 0,95 helyett
- valójában  $n \geq 9650$  kell (pontos számolással)

# Szükséges mintaelemszám



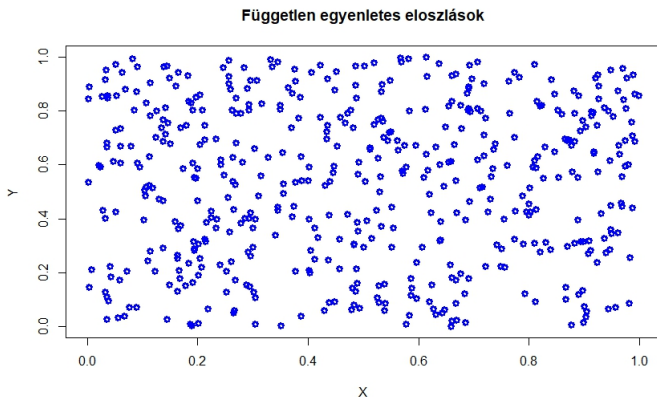
A szükséges mintaelemszám a megengedett hibák függvényében

## Kétdimenziós egyenletes eloszlás



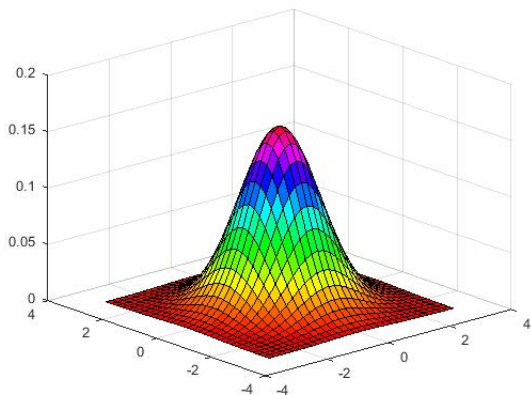
$(X, Y)$  együttes sűrűségfüggvénye, ahol  $X$  és  $Y$  függetlenek és a  $[0, 1]$  intervallumon egyenletes eloszlásúak

# Kétdimenziós egyenletes eloszlás



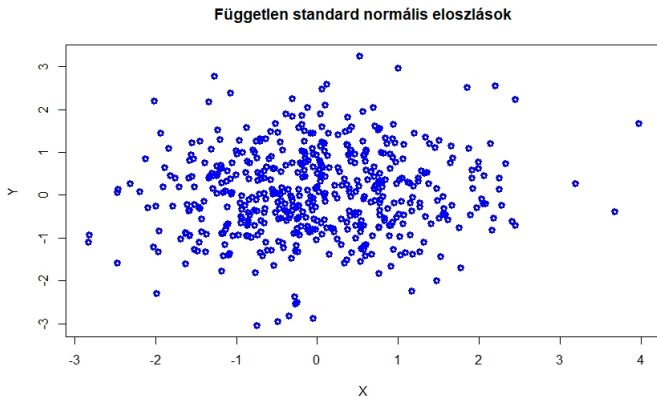
500 darab véletlen pont a síkon, melyek koordinátái függetlenek és a  $[0, 1]$  intervallumon egyenletes eloszlásúak (az együttes sűrűségfüggvény az előző ábrán látható).

## Kétdimenziós normális eloszlás



Két független standard normális eloszlás együttes sűrűségfüggvénye. Azaz:  $(X, Z)$  együttes sűrűségfüggvénye, ahol  $X, Z$  függetlenek,  $N(0, 1)$  eloszlásúak.

# Kétdimenziós normális eloszlás



500 darab véletlen pont a síkon, melyek koordinátái független standard normális eloszlásúak. Ahol nagyobb az együttes sűrűségfüggvény (előző ábra), oda több pont esik.

# Együttes sűrűségfüggvény

## Definíció

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

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

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

# Együttes sűrűségfüggvény

## Definíció

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

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

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

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

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

Következmény:

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

# Peremeloszlások sűrűségfüggvénye

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

## Peremeloszlások sűrűségfüggvénye

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

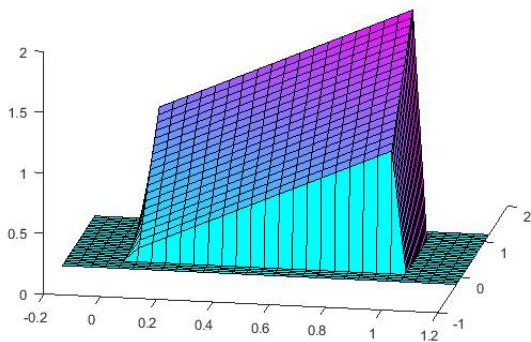
Az  $X_j$  valószínűségi változó sűrűségfüggvénye (melyet  $f_j$ -vel jelölünk), azaz a  $j$ . peremsűrűségfüggvény így kapható meg  $f$ -ből:

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

Speciálisan  $n = 2$ -re:

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

## Együttes sűrűségfüggvény: példa



A  $[0, 1] \times [0, 1]$  négyzeten  $x + y$  alakú együttes sűrűségfüggvény

## Együttes sűrűségfüggvény: példa

Tegyük fel, hogy az  $(X, Y)$  valószínűségi vektorváltozó együttes sűrűségfüggvénye

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

Számítsuk ki  $X$  és  $Y$  korrelációs együtthatóját:

$$R(X, Y) = \frac{\text{cov}(X, Y)}{D(X)D(Y)} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{D(X)D(Y)}.$$

## Együttes sűrűségfüggvény: példa

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

### Állítás

*Legyen az  $(X, Y)$  valószínűségi vektorváltozó együttes sűrűségfüggvénye  $f$ . Ekkor*

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx.$$

## Együttes sűrűségfüggvény: példa

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

### Állítás

Legyen az  $(X, Y)$  valószínűségi vektorváltozó együttes sűrűségfüggvénye  $f$ . Ekkor

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx.$$

$$\begin{aligned} \mathbb{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y) dy dx = \int_0^1 \int_0^1 xy(x + y) dy dx = \\ &= \int_0^1 \int_0^1 x^2 y dy dx + \int_0^1 \int_0^1 xy^2 dy dx = \int_0^1 \frac{x^2}{2} dx + \int_0^1 \frac{x}{3} dx = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}, \end{aligned}$$

felhasználva, hogy  $\int_0^1 x^k dx = [x^{k+1}/(k+1)]_{x=0}^1 = 1/(k+1)$ .

## Együttes sűrűségfüggvény: példa

Az  $(X, Y)$  együttes sűrűségfüggvénye:

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

Az  $X$  valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 x + y dy = x + \frac{1}{2},$$

ha  $0 \leq x \leq 1$ , és 0 különben. Ezért

## Együttes sűrűségfüggvény: példa

Az  $(X, Y)$  együttes sűrűségfüggvénye:

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

Az  $X$  valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 x + y dy = x + \frac{1}{2},$$

ha  $0 \leq x \leq 1$ , és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

## Együttes sűrűségfüggvény: példa

Az  $X$  valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 x + y dy = x + \frac{1}{2},$$

ha  $0 \leq x \leq 1$ , és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_1(x) dx = \int_0^1 x^2 \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^3 + \frac{x^2}{2} dx = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

## Együttes sűrűségfüggvény: példa

Az  $X$  valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 x + y dy = x + \frac{1}{2},$$

ha  $0 \leq x \leq 1$ , és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_1(x) dx = \int_0^1 x^2 \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^3 + \frac{x^2}{2} dx = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

A szimmetria miatt hasonlóképpen:

$$\mathbb{E}(Y) = \frac{7}{12}; \quad \mathbb{E}(Y^2) = \frac{5}{12}.$$

## Együttes sűrűségfüggvény: példa

$$f(x, y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

$$\mathbb{E}(XY) = \frac{1}{3}; \quad \mathbb{E}(X) = \mathbb{E}(Y) = \frac{7}{12}; \quad \mathbb{E}(X^2) = \mathbb{E}(Y^2) = \frac{5}{12}.$$

Az  $X$  és  $Y$  korrelációs együtthatója:

$$\begin{aligned} R(X, Y) &= \frac{\text{cov}(X, Y)}{D(X)D(Y)} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{D(X)D(Y)} = \\ &= \frac{1/3 - (7/12)^2}{\sqrt{5/12 - (7/12)^2} \cdot \sqrt{5/12 - (7/12)^2}} = \frac{1/3 - (7/12)^2}{5/12 - (7/12)^2} = \\ &= -0,091. \end{aligned}$$

Nagyon gyenge negatív korreláció van a két valószínűségi változó között.