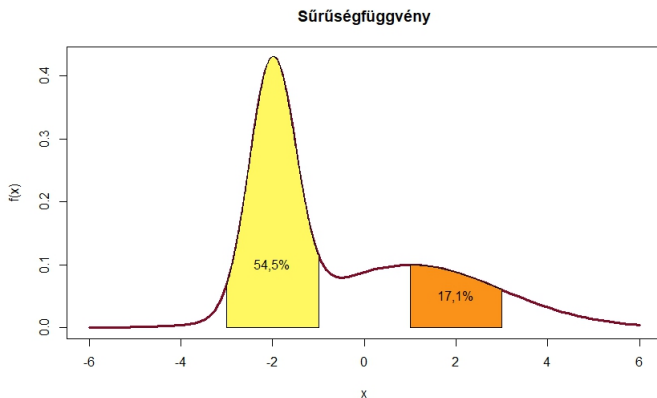


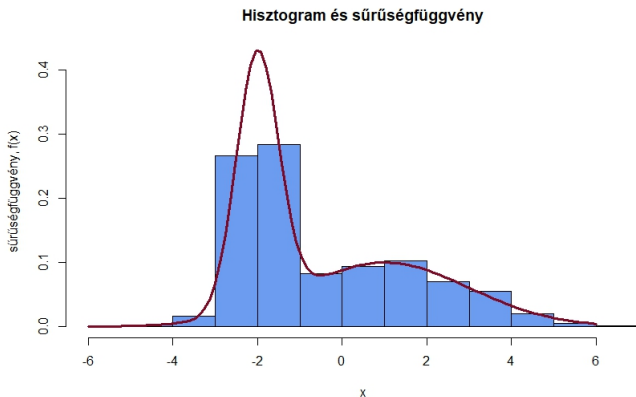
## Sűrűségfüggvény (2. előadás)



Ha az  $X$  valószínűségi változó sűrűségfüggvénye  $f$  (ami most az ábrán látható függvény):  $\mathbb{P}(-3 \leq X \leq -1) = \int_{-3}^{-1} f(x) dx = \mathbf{54,5\%}$ ;

$\mathbb{P}(1 \leq X \leq 3) = \int_1^3 f(x) dx = \mathbf{17,1\%}$ .

# Sűrűségfüggvény



Egy sűrűségfüggvény és hozzá tartozó ezer elemű független minta hisztogramja; nagyobb a sűrűségfüggvény  $\rightarrow$  nagyobb a gyakoriság; minta: független valószínűségi változók, melyek mindegyikének  $f$  a sűrűségfüggvénye

## Sűrűségfüggvény: definíció

Az  $X : \Omega \rightarrow \mathbb{R}$  valószínűségi változó **sűrűségfüggvénye** az  $f : \mathbb{R} \rightarrow \mathbb{R}$  függvény, ha

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

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

**Nem minden** valószínűségi változónak van sűrűségfüggvénye, például a diszkrétnek nincs. Ha  $X$ -nek **van sűrűségfüggvénye**, akkor **abszolút folytonos** valószínűségi változónak nevezzük.

Ha az  $X$  valószínűségi változó sűrűségfüggvénye  $f$ , akkor tetszőleges  $a < b$  számokra

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

## A sűrűségfüggvény tulajdonságai

Legyen  $X$  abszolút folytonos valószínűségi változó, melynek  $F$  az eloszlásfüggvénye.

(a) Ha  $f$  az  $X$  sűrűségfüggvénye, akkor minden  $t \in \mathbb{R}$  számra

$$F(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f(x) dx.$$

(b) Az  $f(t) = F'(t)$  függvény (azokra a  $t$ -kre, ahol  $F$  differenciálható) az  $X$  sűrűségfüggvénye.

Ha az  $f : \mathbb{R} \rightarrow \mathbb{R}$  függvény sűrűségfüggvény, akkor

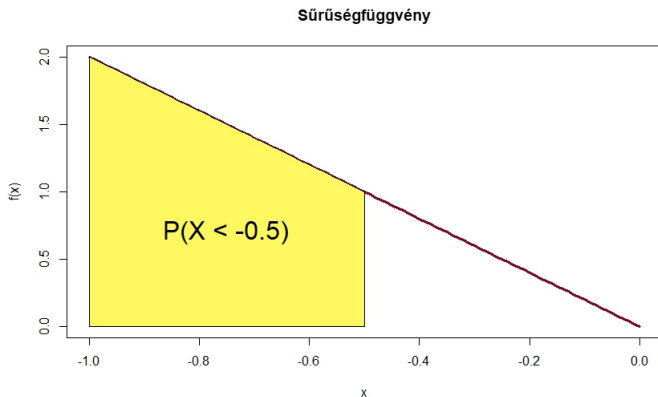
- (i)  $f(x) \geq 0$  teljesül „majdnem minden”  $x \in \mathbb{R}$ -re (például véges vagy megszámlálható sok kivétel lehetséges).
- (ii)  $\int_{-\infty}^{\infty} f(x) dx = 1$ .

*Fordítva:* ha  $f$  teljesíti ezt a két tulajdonságot, akkor van olyan valószínűségi változó, aminek  $f$  a sűrűségfüggvénye.

## Sűrűségfüggvény: példa

Legyen az  $X$  valószínűségi változó sűrűségfüggvénye  $f(x) = 2|x|$ , ha  $-1 < x < 0$ , és 0 különben. Mennyi  $X$  eloszlásfüggvényének értéke a  $-1/2$  helyen?

## Sűrűségfüggvény: példa



Ha  $X$  sűrűségfüggvénye  $f$ , akkor

$$F(-1/2) = \mathbb{P}(X \leq -1/2) = \int_{-\infty}^{-1/2} f(x) dx.$$

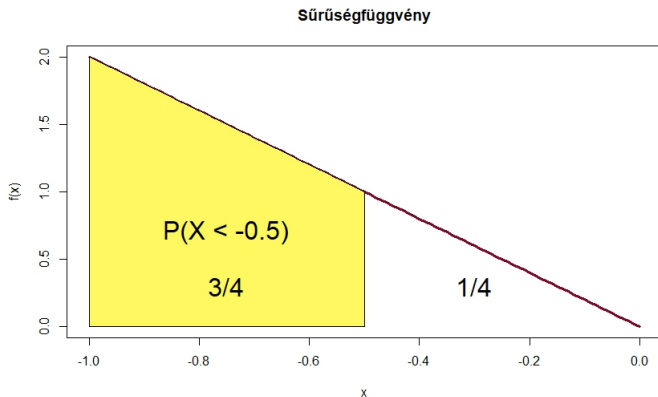
## Sűrűségfüggvény: példa

Legyen az  $X$  valószínűségi változó sűrűségfüggvénye  $f(x) = 2|x|$ , ha  $-1 < x < 0$ , és 0 különben. Mennyi  $X$  eloszlásfüggvényének értéke a  $-1/2$  helyen?

Felhasználva az **eloszlásfüggvény** és a **sűrűségfüggvény** definícióját, illetve hogy  $x \leq -1$  esetén  $f(x) = 0$ , azt kapjuk, hogy

$$\begin{aligned} F(-1/2) &= \mathbb{P}(X \leq -1/2) = \int_{-\infty}^{-1/2} f(x) dx = \int_{-1}^{-1/2} f(x) dx = \\ &= \int_{-1}^{-1/2} 2|x| dx = - \int_{-1}^{-1/2} 2x dx = - [x^2]_{x=-1}^{x=-1/2} = \\ &= - \left( \frac{1}{4} - 1 \right) = \frac{3}{4}. \end{aligned}$$

## Sűrűségfüggvény: példa



Ha  $X$  sűrűségfüggvénye  $f$ , akkor

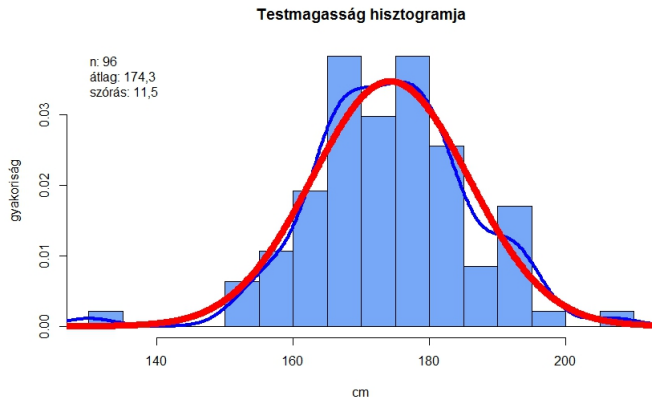
$$F(-1/2) = \mathbb{P}(X \leq -1/2) = \int_{-\infty}^{-1/2} f(x) dx = \frac{3}{4}.$$

## Normális eloszlás: bevezetés

Ha egy, a valóságban megfigyelhető valószínűségi változó eloszlását, pontosabban a **sűrűségfüggvényét** szeretnénk meghatározni:

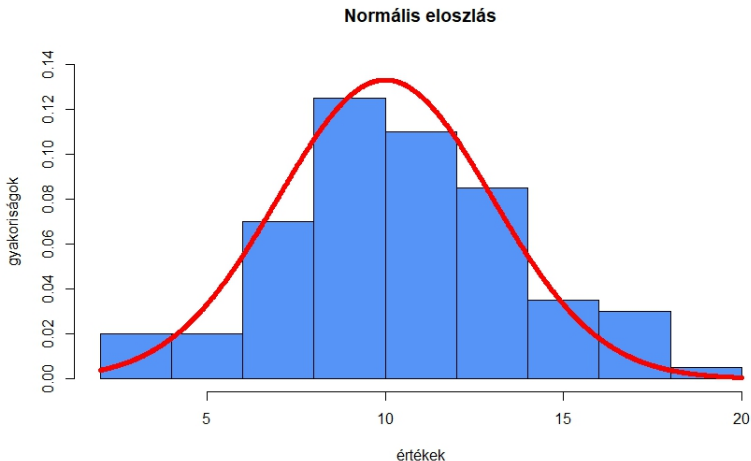
- az adatokból készíthetünk hisztogramot;
- a hisztogram és a sűrűségfüggvény alakja sok független megfigyelés esetén „közel” van egymáshoz;
- megfigyelhetjük, hogy különféle mennyiségek esetén a hisztogramok gyakran **hasonló alakúak** → a gyakran előforduló sűrűségfüggvény-típusokat érdemes külön megérteni;
- az egyik ilyen a **normális eloszlás**, melynek sűrűségfüggvénye az  $e^{-x^2}$  függvényből származtatható
- például különféle *mérési eredmények* (a mérési hibák következtében), illetve élőlények *biológiai jellemzői* gyakran normális eloszlást követnek (például: testmagasság)
- a normális eloszlás a **statisztikában** is kulcsfontosságú

# Testmagasság



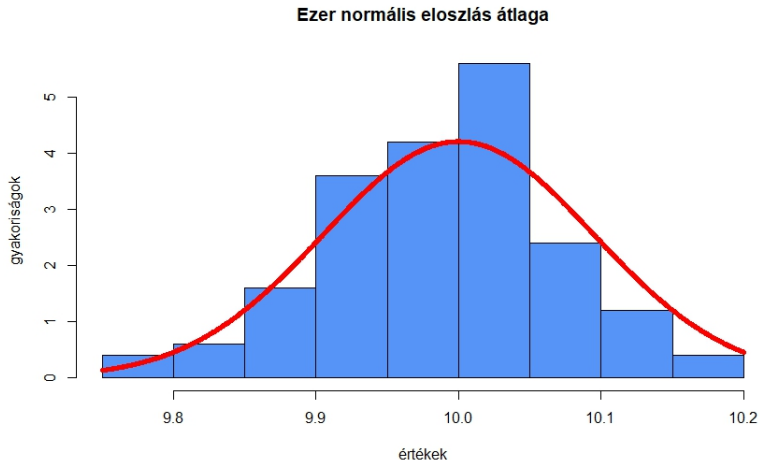
Testmagasság histogramja  $n = 96$  elemű mintából (valós adatokból), és az  $m = \bar{X} = 174,3$  várható értékű és  $\sigma = 11,5$  szórású normális eloszlás sűrűségfüggvénye (pirossal):  $f(x) = \frac{1}{\sqrt{2\pi \cdot 11,5^2}} \exp(-(x - 174,3)^2 / (2 \cdot 11,5^2))$

# Normális eloszlás



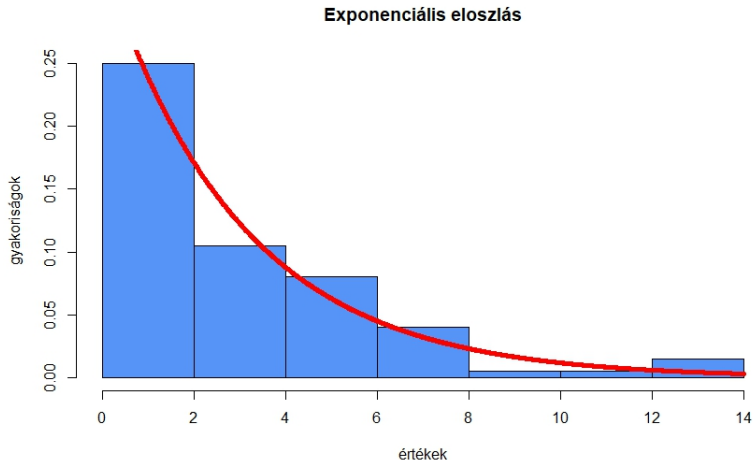
Száz független normális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz  $X_j \sim N(10, 3^2)$  ( $m = 10$ ,  $\sigma = 3$ ,  $\bar{x} = 9,88$ ,  $s_n^* = 2,58$ )

# Normális eloszlások átlaga



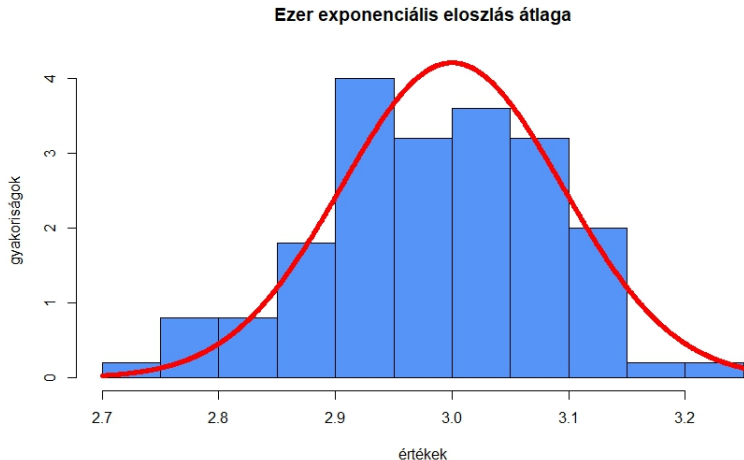
Az átlag eloszlása százelemű mintából, ha  $X_j \sim N(10, 3^2)$  és  $n = 1000$ , továbbá az  $N(10, 3/\sqrt{1000})$  normális eloszlás sűrűségfüggvénye ( $\bar{x} = 9,99, s_n^* = 0,084, \sigma/\sqrt{n} = 0,095$ )

# Exponenciális eloszlás



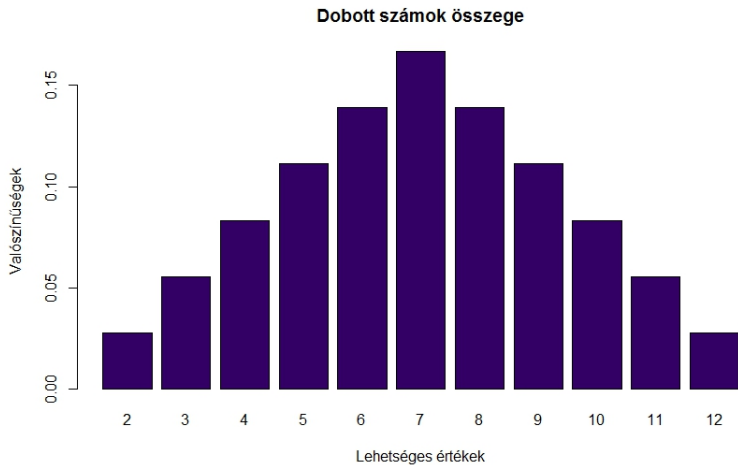
Száz független  $\lambda = 1/3$  paraméterű exponenciális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz  $e^{-1/3}/3$  ( $\mathbb{E}(X) = D(X) = 3, \bar{x} = 3,03, s_n^* = 2,89$ )

# Exponenciális eloszlások átlaga



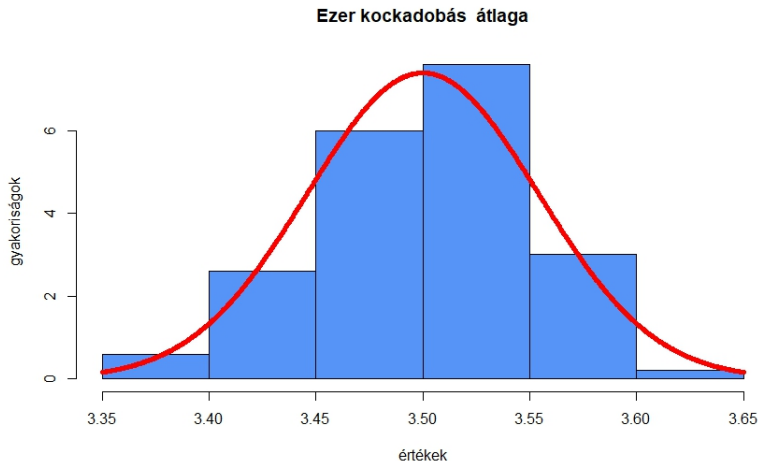
Az átlag eloszlása százelemű mintából, ha  $X_j$  eloszlása **exponenciális**  $\lambda = 1/3$ -dal, és  $n = 1000$ , és az  $N(3, 3/\sqrt{1000})$  **normális** eloszlás sűrűségfüggvénye ( $\bar{x} = 2,98, s_n^* = 0,098$ )

# Két kockadobás összege



Két szabályos kockadobás összegének eloszlása

# Kockadobások átlaga



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

## Normális eloszlás: definíció

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**

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

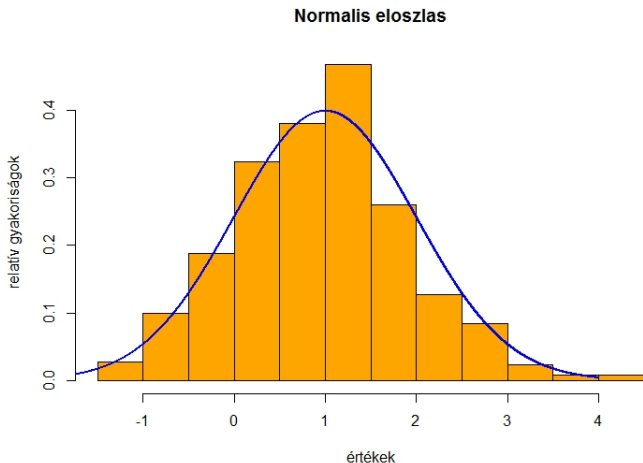
Jelölése:  $Y \sim N(m, \sigma^2)$ .

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

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

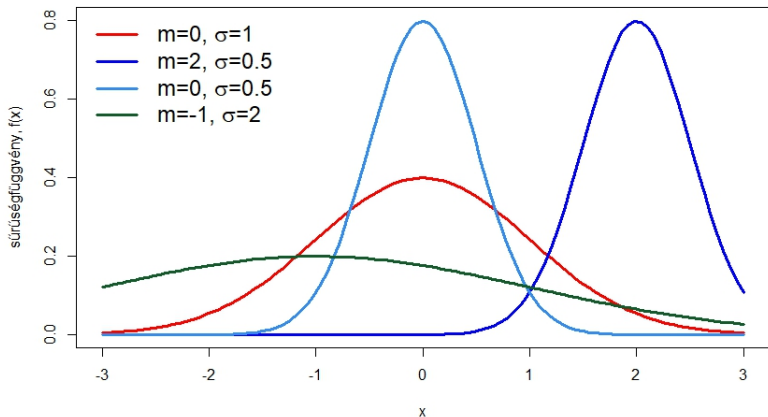
# Normális eloszlás



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

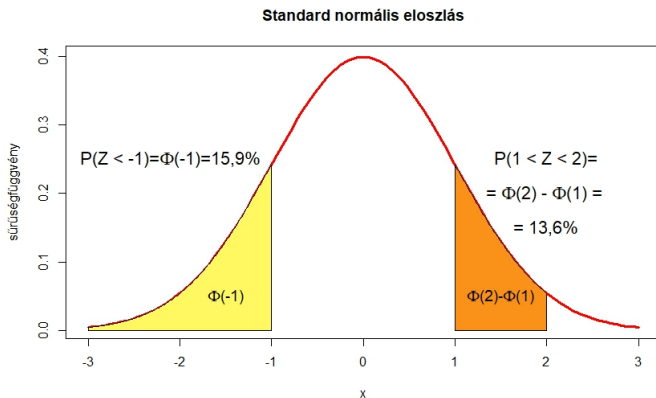
# Normális eloszlás

Normális eloszlások sűrűségfüggvényei



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

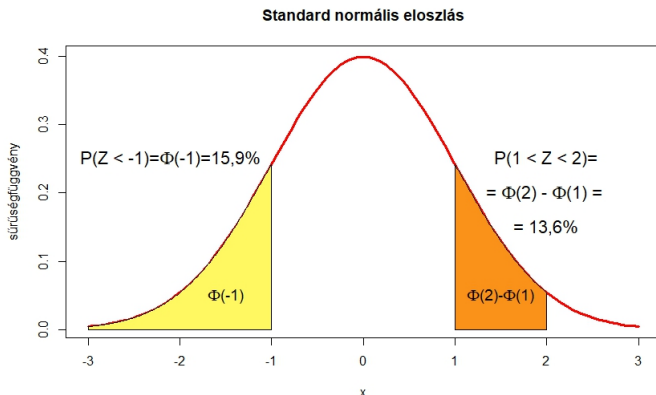
# A $\Phi$ függvény



A  $\Phi$  függvény a standard normális eloszlás eloszlásfüggvénye: ha  $Z \sim N(0, 1)$ , akkor

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

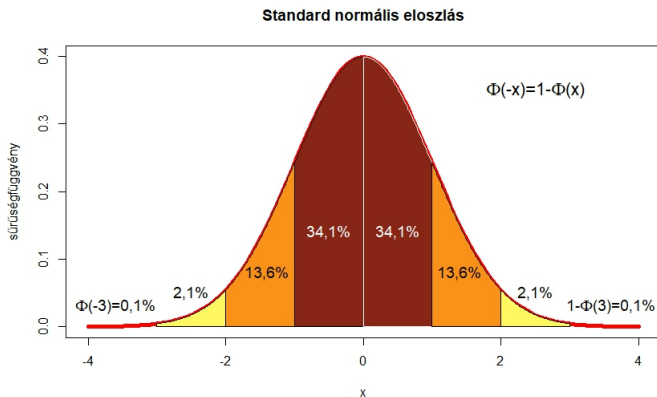
# Standard normális eloszlás



A  $\Phi$  függvény a standard normális eloszlás eloszlásfüggvénye: ha  $Z \sim N(0,1)$ , akkor

$$\mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a) = \int_a^b \varphi(x) dx = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

# Standard normális eloszlás



A  $\Phi$  függvény a standard normális eloszlás eloszlásfüggvénye: ha  $Z \sim N(0,1)$ , akkor

$$\mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a) = \int_a^b \varphi(x) dx = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

# A normális eloszlás tulajdonságai

**Lineáris transzformáció.** Legyen  $Y$  normális eloszlású valószínűségi változó  $m$  várható értékkel és  $\sigma$  szórással, és  $a, b$  valós számok. Ekkor az  $aY + b$  valószínűségi változó normális eloszlású  $am + b$  várható értékkel és  $a^2\sigma^2$  szórásnégyzettel, azaz

$$Y \sim N(m, \sigma^2) \quad \Rightarrow \quad aY + b \sim N(am + b, a^2\sigma^2).$$

**Független összeg.** 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  $Y_1 \sim N(m_1, \sigma_1^2)$  és  $Y_2 \sim N(m_2, \sigma_2^2)$

**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

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

$$Y_1 + Y_2 + \dots + Y_n \sim N(nm, n\sigma^2);$$

$$\frac{Y_1 + Y_2 + \dots + Y_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

# A normális eloszlás tulajdonságai

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

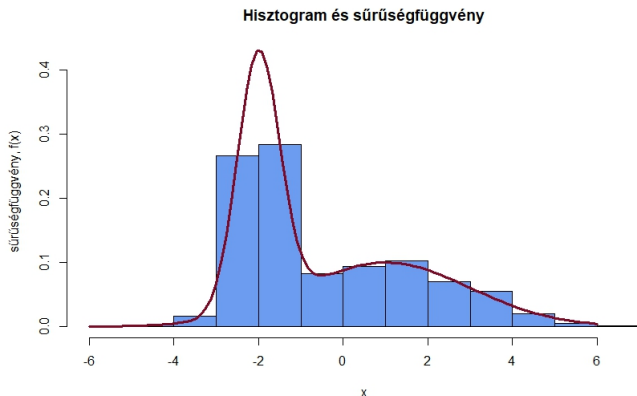
$$Y_1 + Y_2 + \dots + Y_n \sim N(nm, n\sigma^2);$$

$$\frac{Y_1 + Y_2 + \dots + Y_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

**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.

# Sűrűségfüggvény



Egy sűrűségfüggvény és hozzá tartozó ezer elemű független minta hisztogramja – mennyi lehet az  $f$  sűrűségfüggvényű valószínűségi változó **várható értéke** és **szórása**?

## Várható érték és szórás abszolút folytonos esetben

Legyen  $X$  abszolút folytonos valószínűségi változó, melynek sűrűségfüggvénye  $f$ .  
Ekkor  $X$  **várható értéke**:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx,$$

ha ez az integrál létezik és véges.

Tegyük fel, hogy az  $X$  valószínűségi változó abszolút folytonos, sűrűségfüggvénye  $f$ , és  $\mathbb{E}(X^2)$  létezik, azaz az  $\int_{-\infty}^{\infty} x^2 f(x) dx$  integrál véges. Ekkor  $X$  **szórásnégyzete**:

$$D^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

**szórása** pedig

$$D(X) = \sqrt{\mathbb{E}((X - \mathbb{E}(X))^2)} = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}.$$

*A szórás definíciója megegyezik a diszkrét esetben használttal.*