Sűrűségfüggvény: definíció (11. előadás)

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

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

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

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

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

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

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.

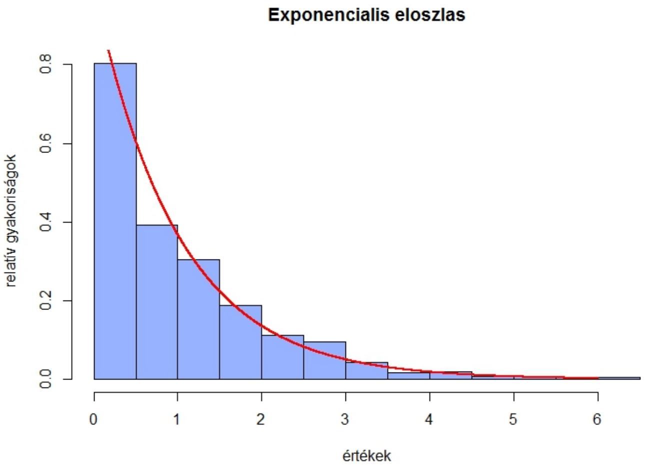
Ha az $f : \mathbb{R} \to \mathbb{R}$ függvény sűrűségfüggvény, akkor

(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.

Exponenciális eloszlás



$\lambda = 1$ paraméterű exponenciális eloszlás sűrűségfüggvénye és 500 darab független, 1 paraméterű exponenciális eloszlású valószínűségi változóból álló minta hisztogramja

Exponenciális eloszlás: definíció és tulajdonságok

Legyen $\lambda > 0$ valós szám. Az X valószínűségi változó **exponenciális eloszlású** λ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ha } x > 0; \\ 0, & \text{különben.} \end{cases}$$

Ekkor

(i) X eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \le t) = \int_{-\infty}^{t} f(x)\,dx = \begin{cases} 1 - e^{-\lambda t}, & \text{ha } t > 0; \\ 0 & \text{különben.} \end{cases}$$

(ii) X várható értéke: $\mathbb{E}(X) = \frac{1}{\lambda}$, szórása: $D(X) = \frac{1}{\lambda}$.

(iii) **Örökifjú tulajdonság.** Legyenek s, t pozitív számok. Ekkor

$$\mathbb{P}(X \ge s + t | X \ge s) = \mathbb{P}(X \ge t).$$

Exponenciális eloszlás: példa

Tegyük fel, hogy egy boltban egy vevő kiszolgálásának ideje (percben számolva) 3 **várható értékű exponenciális eloszlású** valószínűségi változó.

Mennyi a valószínűsége, hogy a vevőt **legalább** 5 **percig** tart kiszolgálni?
Mennyi a valószínűsége, hogy a vevő kiszolgálása **legalább** 2, **de legfeljebb** 4 **percig** tart?

Exponenciális eloszlás: példa

Tegyük fel, hogy egy boltban egy vevő kiszolgálásának ideje (percben számolva) **3 várható értékű exponenciális eloszlású** valószínűségi változó.

Mennyi a valószínűsége, hogy a vevőt **legalább 5 percig** tart kiszolgálni? Mennyi a valószínűsége, hogy a vevő kiszolgálása **legalább 2, de legfeljebb 4 percig** tart? Legyen X a kiszolgálás ideje. Exponenciális eloszlás esetén

$$\mathbb{E}(X) = 1/\lambda, \text{ és most } \mathbb{E}(X) = 3 \Rightarrow \lambda = 1/3.$$

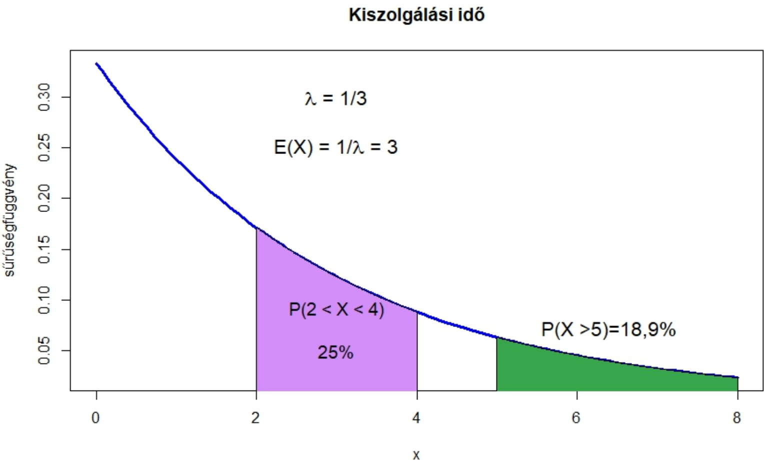
Ezért az eloszlásfüggvény alapján

$$\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X < 5) = 1 - F(5) = 1 - (1 - e^{-\lambda \cdot 5}) = e^{-\lambda \cdot 5} =$$
$$= e^{-5/3} = 18,9\%.$$

Hasonlóképpen

$$\mathbb{P}(2 \leq X \leq 4) = \mathbb{P}(X \leq 4) - \mathbb{P}(X \leq 2) = F(4) - F(2) =$$
$$= (1 - e^{-4/3}) - (1 - e^{-2/3}) = e^{-2/3} - e^{-4/3} = 25\%.$$

Exponenciális eloszlás: példa



Annak valószínűsége, hogy a kiszolgálás legalább 5 percig tart, illetve hogy legalább 2, de legalább 4 percig, ha a kiszolgálási idő 3 várható értékű exponenciális eloszlású valószínűségi változó

Normális eloszlás: definíció

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**

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

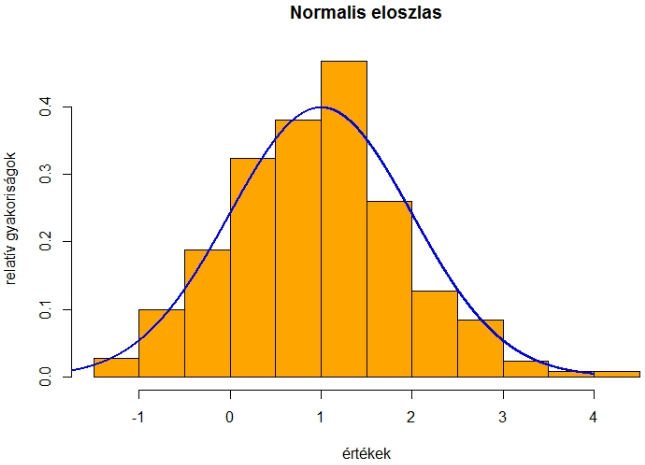
Jelölése: $Y \sim N(m, \sigma^2)$.

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

Standard normális eloszlás: az $m = 0$ várható értékű és $\sigma = 1$ szórású normális eloszlás. **Eloszlásfüggvénye:** Φ, sűrűségfüggvénye φ, ahol

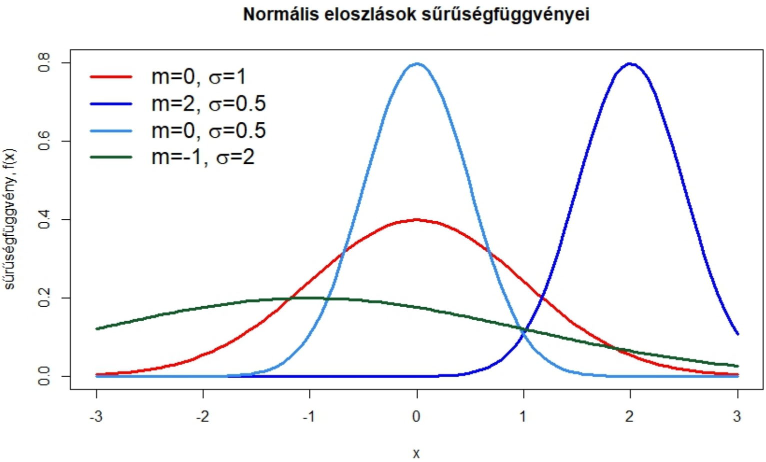
$$\Phi(t) = \int_{-\infty}^{t} \varphi(x)\, dx; \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right).$$

Normális eloszlás



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

Normális eloszlás



Különböző várható értékű (m) és szórású (σ) normális eloszlások sűrűségfüggvényei

További nevezetes eloszlások

Az alábbi eloszlások többek között **statisztikai alkalmazásokban** fordulnak elő:

- Pareto-eloszlás: végtelen momentumokkal rendelkező eloszlások modellezésére (például jövedelmek, kárnagyságok)
- t-eloszlás: például két eloszlás **várható értékének** összehasonlítására
- F-eloszlás: például két eloszlás **szórásának** összehasonlítására
- χ^2-eloszlás: például annak eldöntésére, hogy két tulajdonság között van-e **összefüggés**
- gamma-eloszás: nemnegatív valószínűségi változók modellezésére
- beta-eloszlás: $[0, 1]$-értékű valószínűségi változók modellezésére

Pareto-eloszlás

Az X valószínűségi változó Pareto-eloszlású, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\beta}{\alpha^{\beta+1}} \cdot x^{-\beta}; & \text{ha } x \leq \alpha \\ 0, & \text{ha } x < \alpha. \end{cases}$$

Itt $\alpha > 0, \beta > 1$ rögzített számok. Ekkor X eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 1 - \left(\frac{t}{\alpha}\right)^{-\beta+1}; & \text{ha } t \leq \alpha \\ 0, & \text{ha } t < \alpha. \end{cases}$$

Pareto-eloszlás

Az X valószínűségi változó Pareto-eloszlású, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\beta}{\alpha^{\beta+1}} \cdot x^{-\beta}; & \text{ha } x \leq \alpha \\ 0, & \text{ha } x < \alpha. \end{cases}$$

Itt $\alpha > 0, \beta > 1$ rögzített számok. Ekkor X eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 1 - \left(\frac{t}{\alpha}\right)^{-\beta+1}; & \text{ha } t \leq \alpha \\ 0, & \text{ha } t < \alpha. \end{cases}$$

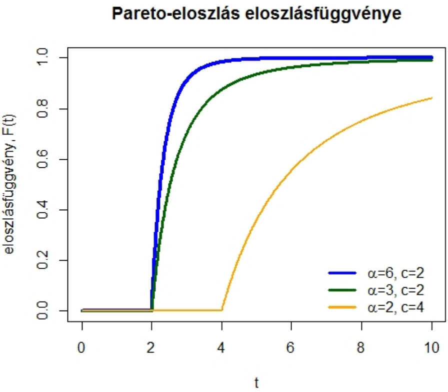
Az X valószínűségi változó k. momentuma:

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k \cdot f(x)\, dx = \frac{\beta}{\alpha^{\beta+1}} \int_{-\infty}^{\infty} x^{k-\beta}\, dx < \infty \Leftrightarrow k - \beta < -1.$$

Tehát a Pareto-eloszlásnak csak $\beta-1$-nél kisebb k-ra véges a k. momentuma.

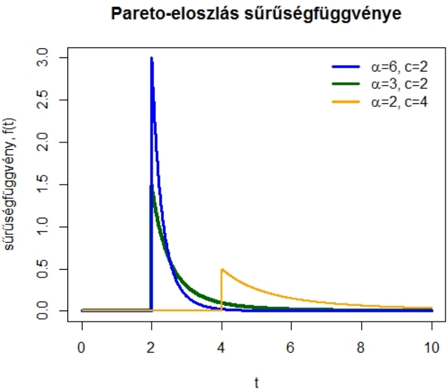
Például ha $\beta = 2,5$, akkor a várható érték létezik és véges, de szórása nem létezik.

Pareto-eloszlás



Különböző paraméterű Pareto-eloszlások eloszlásfüggvénye

Pareto-eloszlás



Különböző paraméterű Pareto-eloszlások sűrűségfüggvénye

t-eloszlás

Legyenek $X_1, X_2, \ldots, X_f$ és Y független standard normális eloszlású valószínűségi változók. Ekkor a

$$Z = \frac{Y}{\sqrt{(X_1^2 + X_2^2 + \ldots + X_f^2)/f}}$$

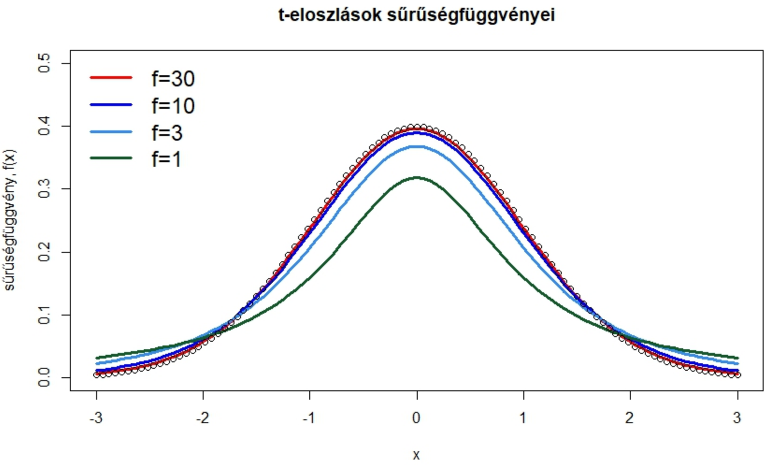
valószínűségi változó eloszlását f szabadsági fokú ***t*-eloszlásnak** (vagy Student-eloszlásnak) nevezzük.

Az $f = 1$ szabadsági fokú t-eloszlás, vagyis Y/X eloszlása a **Cauchy-eloszlás**. Ennek sűrűségfüggvénye:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

A Cauchy-eloszlásnak sem várható értéke, sem szórása nem létezik: $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$ nem értelmezhető, mert $\int_0^{\infty} \frac{1}{x}\, dx$ integrál nem véges.

A t-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú t-eloszlások sűrűségfüggvényei. A pöttyözött vonal a standard normális eloszlás sűrűségfüggvényét jelöli, ez közel van a t-eloszlás sűrűségfüggvényéhez, ha f nagy.

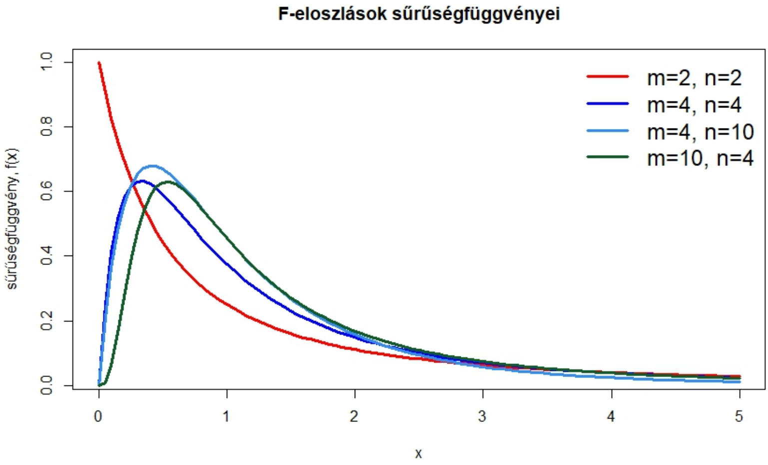
F-eloszlás

Legyenek m, n pozitív egészek, $X_1, \ldots, X_m$, $Y_1, Y_2, \ldots, Y_n$ pedig független standard normális eloszlású valószínűségi változók. Ekkor az

$$F = \frac{n(X_1^2 + X_2^2 + \ldots + X_m^2)}{m(Y_1^2 + Y_2^2 + \ldots + Y_n^2)}$$

valószínűségi változó eloszlását m, n szabadsági fokú **F-eloszlásnak** nevezzük.

Az F-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú F-eloszlások sűrűségfüggvényei

χ^2-eloszlás

Legyenek $X_1, X_2, \ldots, X_q$ független standard normális eloszlású valószínűségi változók. Az

$$Y = X_1^2 + X_2^2 + \ldots + X_q^2$$

valószínűségi változó eloszlását q szabadsági fokú **χ^2-eloszlásnak** nevezzük. Ennek sűrűségfüggvénye:

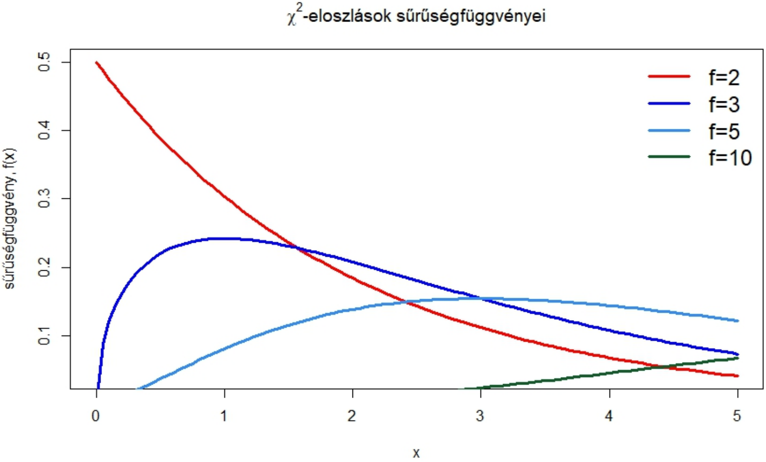
$$f(t) = \begin{cases} \frac{t^{q/2-1}}{2^{q/2}\Gamma(q/2)} e^{-t/2}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Gamma-függvény. Ha $a > 0$ pozitív szám, legyen

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Parciális integrálással belátható, hogy $\Gamma(a) = (a-1)\Gamma(a-1)$ minden $a > 1$-re, és így $\Gamma(n) = (n-1)!$, ha n pozitív egész.

A χ^2-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú χ^2-eloszlások sűrűségfüggvényei

Gamma-eloszlás

Gamma-függvény. Ha $a > 0$ pozitív szám, legyen

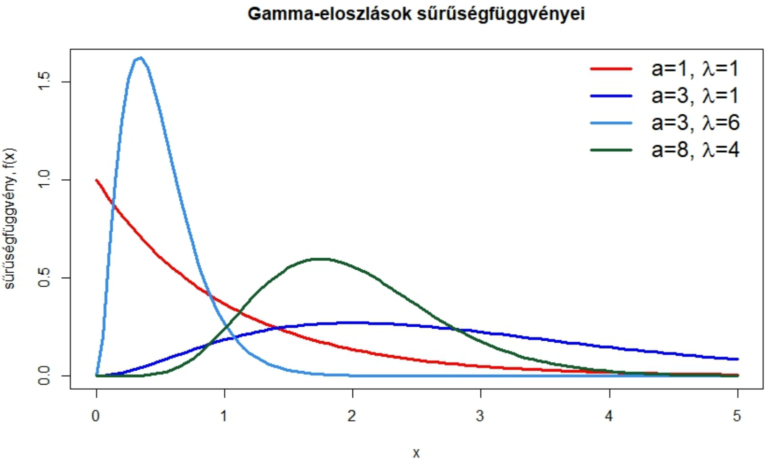
$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Parciális integrálással belátható, hogy $\Gamma(a) = (a-1)\Gamma(a-1)$ minden $a > 1$-re, és így $\Gamma(n) = (n-1)!$, ha n pozitív egész.

Legyenek a és λ pozitív számok. Az X valószínűségi változó **gamma-eloszlású** a renddel és λ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

A gamma-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú gamma-eloszlások sűrűségfüggvényei

A gamma-eloszlás tulajdonságai

Az X valószínűségi változó **gamma-eloszlású** a renddel és λ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

- **Kapcsolat az exponenciális eloszlással**: ha $a = 1$, akkor a sűrűségfüggvény $\lambda e^{-\lambda x}$, ha $x \geq 0$, és az exponenciális eloszlást kapjuk vissza.
- **Exponenciális eloszlások összege:** ha $X_1, X_2, \ldots, X_n$ független λ paraméterű exponenciális eloszlású valószínűségi változók, akkor $X_1 + X_2 + \ldots + X_n$ gamma-eloszlású $a = n$ renddel és λ paraméterrel.
- **Kapcsolat a χ^2-eloszlással**: ha $a = q/2$ és $\lambda = 1/2$, akkor a q szabadsági fokú χ^2-eloszlást kapjuk vissza.
- **Várható érték** és **szórás:**

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

Beta-eloszlás

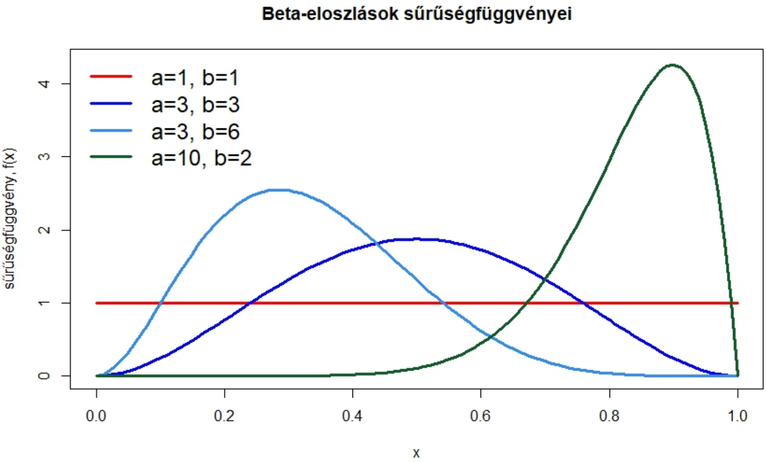
Legyenek $a, b > 1$ számok. Az X valószínűségi változó **beta-eloszlású** a és b paraméterekkel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, & t \in [0,1]; \\ 0, & x < 0 \text{ vagy } x > 1. \end{cases}$$

Ha $X_1, X_2, \ldots, X_n$ független, a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változók, és X_k^* jelöli ebben a mintában a k. legnagyobb számot, akkor X_k^* eloszlása beta-eloszlás $a = k$ és $b = n - k + 1$ paraméterekkel.

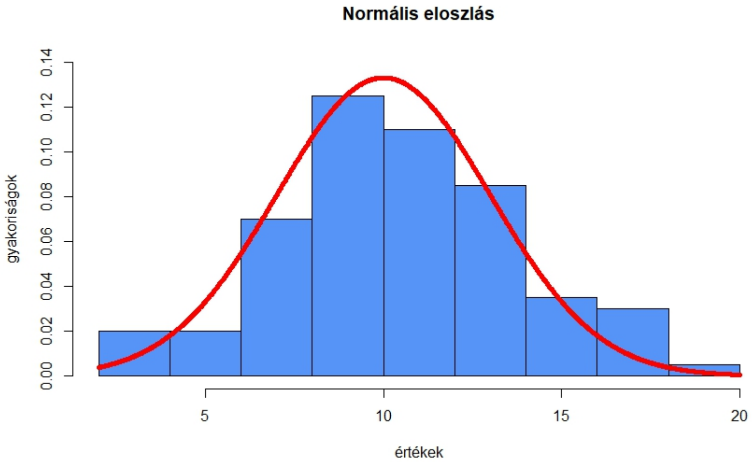
Az $a = 1$ és $b = 1$ választással az **egyenletes eloszlást** kapjuk vissza.

A beta-eloszlás sűrűségfüggvénye



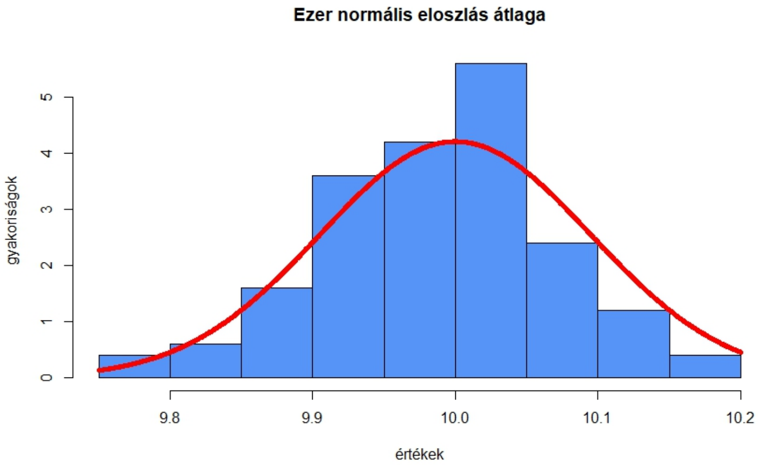
Különböző szabadsági fokú beta-eloszlások sűrűségfüggvényei

Normális eloszlás



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

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 ($\overline{x} = 9,99$, $s_n^* = 0,084$, $\sigma/\sqrt{n} = 0,095$)

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 σ 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:

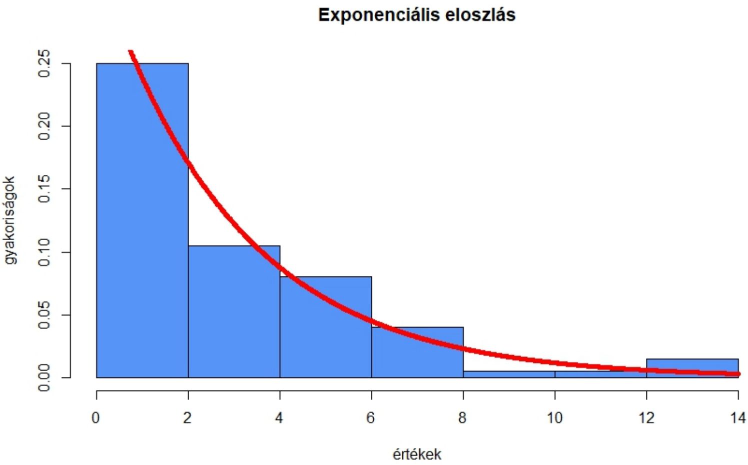
- $X + b$ eloszlása normális, $m_1 + b$ várható értékkel és σ szórással;
- aX eloszlása normális am_1 várható értékkel és $|a|\sigma$ szórással;
- $X + Y$ eloszlása normális, $m_1 + m_2$ várható értékkel és $\sqrt{\sigma_1^2 + \sigma_2^2}$ szórással.

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 $X_1, \ldots, X_n$ független normális eloszlásúak m várható értékkel és σ szórással, akkor

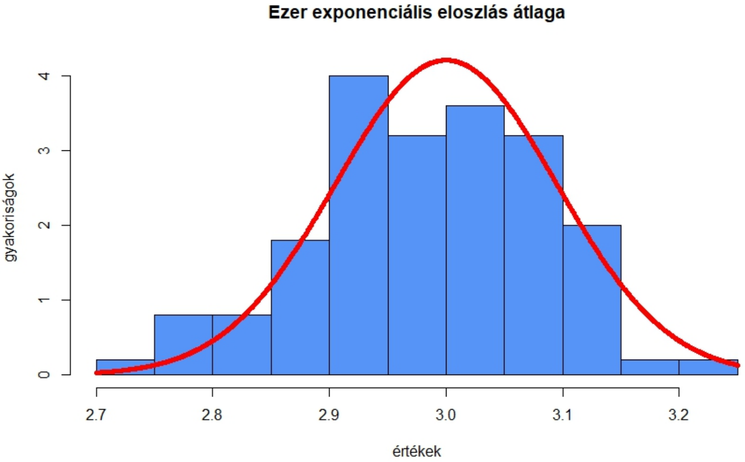
$$\frac{X_1 + \ldots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

Exponenciális eloszlás



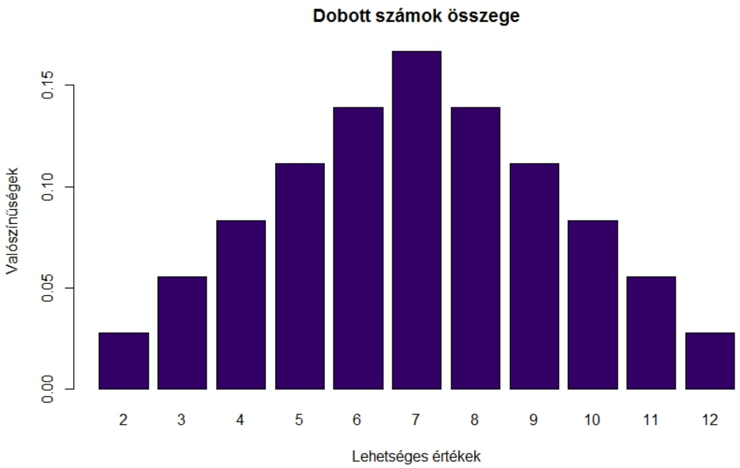
Száz független $\lambda = 1/3$ paraméterű exponencális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz $e^{-1/3}/3$ $(\mathbb{E}(X) = D(X) = 3, \overline{x} = 3,03, s_n^* = 2,89)$

Exponenciális eloszlások átlaga



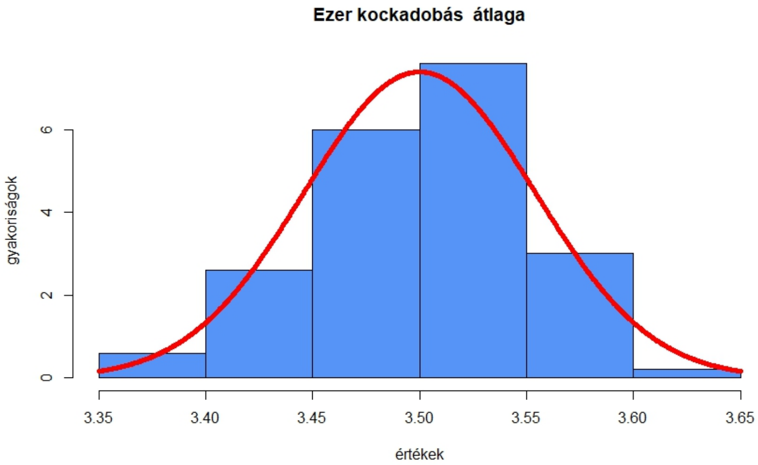
Százelemű minta az alábbi eloszlásból: $n = 1000$ független exponenciális eloszlású ($\lambda = 1/3$) valószínűségi változó átlaga, és az $N(3, 9/1000)$ normális eloszlás sűrűségfüggvénye ($\overline{x} = 2,98$, $s_n^* = 0,098$, $\sigma/\sqrt{n} = 0,095$)

Két kockadobás összege



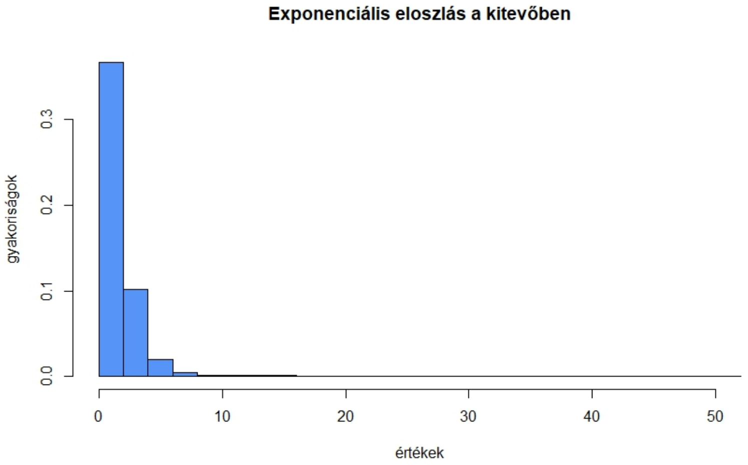
Két szabályos kockadobás összegének eloszlása

Kockadobások átlaga



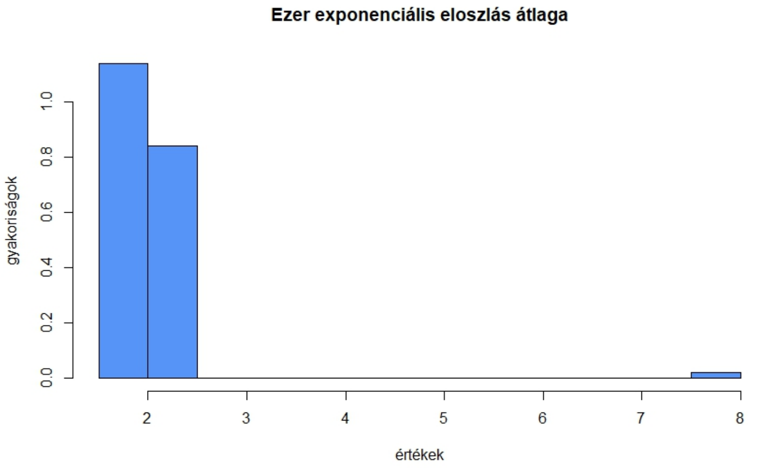
Százelemű minta az alábbi eloszlásból: $n = 1000$ független szabályos kockadobás átlaga, és az $N(3,5, D^2(X_1)/1000)$ normális eloszlás sűrűségfüggvénye ($\overline{x} = 3,501, s_n^* = 0,098, \sigma/\sqrt{n} = 0,051$)

Exponenciális eloszlás a kitevőben



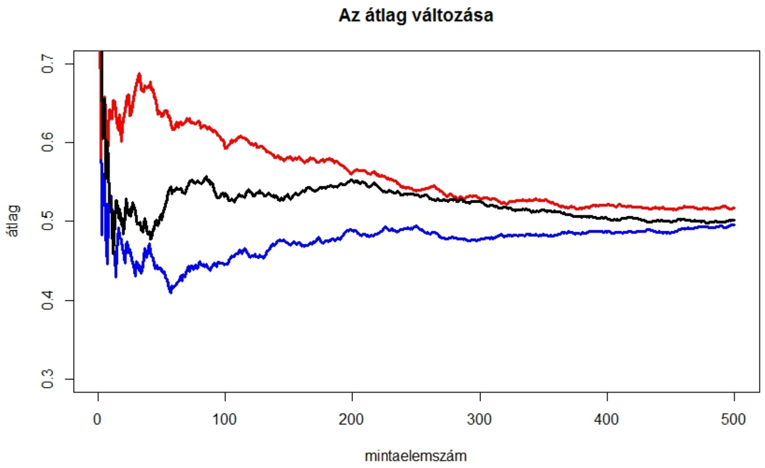
$e^{X_1}, e^{X_2}, \ldots, e^{X_{1000}}$ hisztogramja, ahol X_i-k függetlenek, 2 paraméterű exponenciális eloszlásúak ($\mathbb{E}(e^{X_1}) = 2$, $D(e^{X_1}) = \infty$, $\overline{x} = 1,99$, $s_n^* = 2,33$)

Exponenciális eloszlás a kitevőben



Százelemű minta az alábbi eloszlásból: $e^{X_1}, e^{X_2}, \ldots, e^{X_{1000}}$ 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 $X_1, X_2, \ldots$ *olyan valószínűségi változók, melyek függetlenek és azonosan 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 törvényei

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.

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*

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

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

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

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

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

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

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

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:

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

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

Centrális határeloszlástétel

Legyenek $X_1, X_2, \ldots$ **független** azonos eloszlású valószínűségi változók, melyekre $\mathbb{E}(\mathrm{X}_1) = \mathrm{m}$ és $D(X_1) = \sigma < \infty$, azaz **szórásuk véges**. Ekkor

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

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

Centrális határeloszlástétel

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

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

$$= \Phi(b) - \Phi(a) = \mathbb{P}(a \le Z \le b),$$

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

$$\mathbb{P}(nm + a\sigma\sqrt{n} \le X_1 + X_2 + \ldots + X_n < nm + b\sigma\sqrt{n}) \to \mathbb{P}(a \le Z \le b).$$

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

$$\mathbb{P}\left(\mathrm{m} + \mathrm{a}\frac{\sigma}{\sqrt{n}} \le \frac{X_1 + X_2 + \ldots + X_n}{n} < \mathrm{m} + \mathrm{b}\frac{\sigma}{\sqrt{n}} \right) \to \mathbb{P}(\mathrm{a} \le Z \le \mathrm{b}).$$

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

Centrális határeloszlástétel: példa

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

Centrális határeloszlástétel: példa

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

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

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

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

Konvergenciafajták

Definíció

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

$$\mathbb{P}(Z_n \leq t) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty).$$

Konvergenciafajták

Definíció

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

$$\mathbb{P}(Z_n \leq t) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty).$$

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

Lehetséges, hogy $Z_n \to Z$ 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:

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

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

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

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \le 0,01\right) = \mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} X_j - np}{\sqrt{np(1-p)}}\right| \le \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

$$\approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.$$

A centrális határeloszlástétel alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

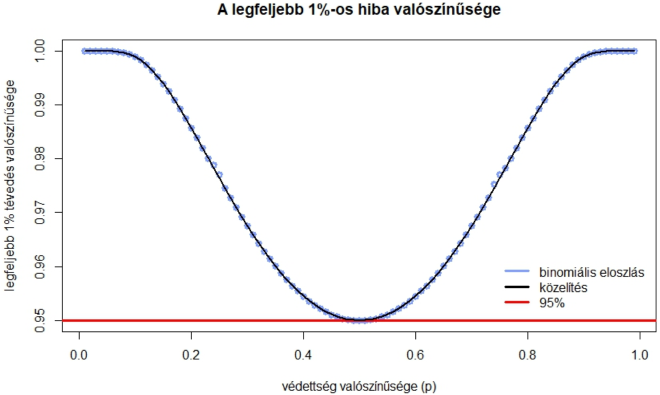
Kell:

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

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

$$\begin{aligned}\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right)&\geq 0,975;\\ \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}&\geq \Phi^{-1}(0,975)=\mathtt{qnorm}(0,975)=1,96;\\ n&\geq p(1-p)\cdot 1,96^2\cdot\frac{1}{0,01^2};\\ n&\geq \frac{1}{4}\cdot 1,96^2\cdot\frac{1}{0,01^2}=9607.\end{aligned}$$

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?

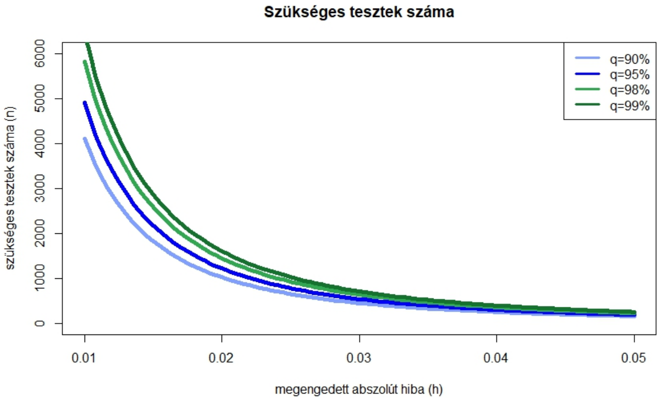
- Csebisev-egyenlőtlenséggel: $n \geq 50000$ biztosan elég
- centrális határeloszlástétellel közelítve: $n \geq 9607$ elég

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

Házi feladat december 11., hétfő, 10:15-ig

Egy koncertre 16000 ember vett jegyet. Mindenki a többiektől függetlenül $1/4$ valószínűséggel vesz egyet a fellépő zenekar pólójából.

A centrális határeloszlástétel alapján közelítve hány pólót kell vinni a koncertre, hogy 95% valószínűséggel jusson mindenkinek, aki szeretne vásárolni?

Számítsuk ki a pontos értéket is az R segítségével.