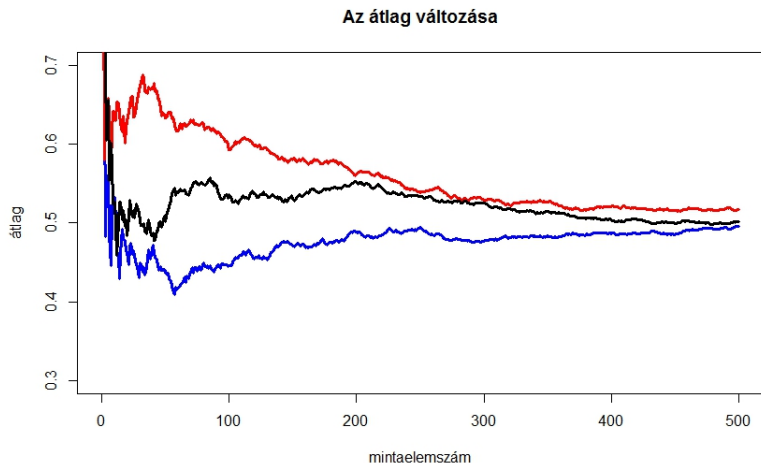


Az átlag konvergenciája (12. előadás)



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

Konvergenciafajták

Definíció

A Z_1, Z_2, \dots , valószínűségi változókból álló sorozat **sztochasztikusan konvergál** az Z valószínűségi változóhoz, ha minden $\varepsilon > 0$ -ra

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \rightarrow 0$$

teljesül $n \rightarrow \infty$ esetén.

Konvergenciafajták

Definíció

A Z_1, Z_2, \dots , valószínűségi változókból álló sorozat **sztochasztikusan konvergál** az Z valószínűségi változóhoz, ha minden $\varepsilon > 0$ -ra

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \rightarrow 0$$

teljesül $n \rightarrow \infty$ esetén.

Definíció

A Z_1, Z_2, \dots , valószínűségi változókból álló sorozat **1 valószínűséggel** konvergál az Z valószínűségi változóhoz, ha

$$\mathbb{P}(\omega \in \Omega : Z_n(\omega) \rightarrow Z(\omega) \text{ } n \rightarrow \infty \text{ esetén}) = 1.$$

Konvergenciafajták

Definíció

A Z_1, Z_2, \dots , valószínűségi változókból álló sorozat **sztochasztikusan konvergál** az Z valószínűségi változóhoz, ha minden $\varepsilon > 0$ -ra

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \rightarrow 0$$

teljesül $n \rightarrow \infty$ esetén.

Definíció

A Z_1, Z_2, \dots , valószínűségi változókból álló sorozat **1 valószínűséggel konvergál** az Z valószínűségi változóhoz, ha

$$\mathbb{P}(\omega \in \Omega : Z_n(\omega) \rightarrow Z(\omega) \text{ } n \rightarrow \infty \text{ esetén}) = 1.$$

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

Lehetséges, hogy $Z_n \rightarrow Z$ sztochasztikusan, de Z_n nulla valószínűséggel konvergens.

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)

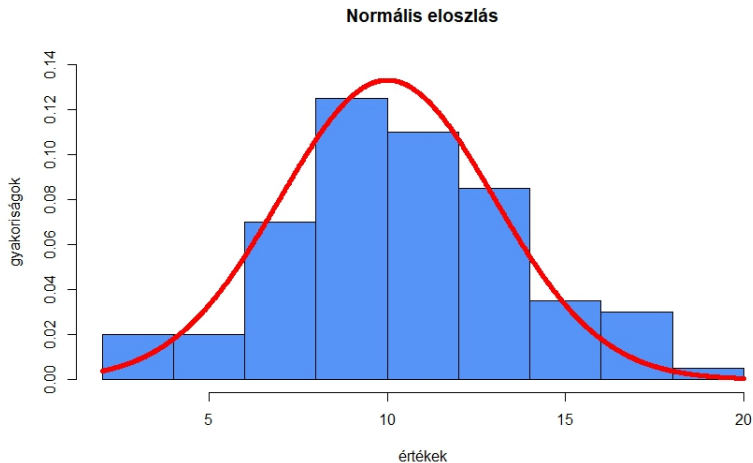
Legyenek X_1, X_2, \dots valószínűségi változók, melyek függetlenek és azonos eloszlásúak. Tegyük fel még, hogy $m = \mathbb{E}(X_1) < \infty$. Ekkor

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

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

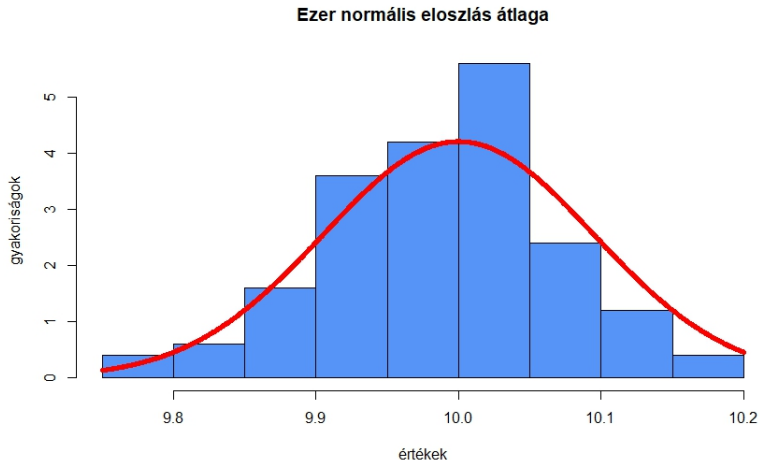
A második esetben gyengébb feltevésből erősebb állítás következik.

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 $N(10, 3/\sqrt{1000})$ 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 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:

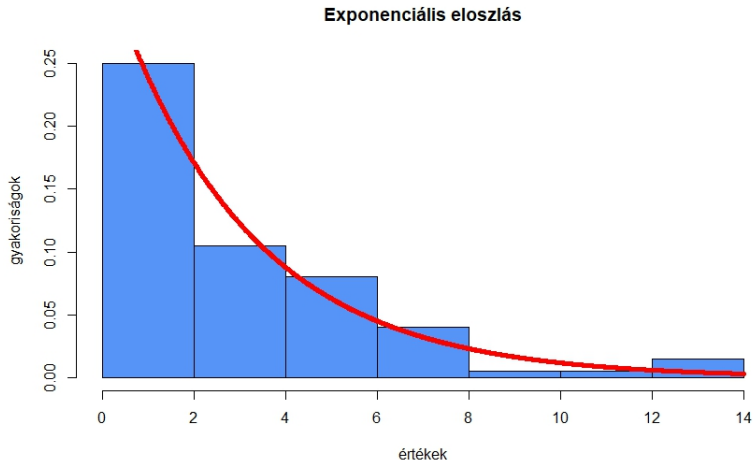
- $X + b$ eloszlása normális, $m_1 + b$ várható értékkel és σ szórással;
- aX eloszlása normális am_1 várható értékkel és $|a|\sigma$ szórással;
- $X + Y$ eloszlása normális, $m_1 + m_2$ várható értékkel és $\sqrt{\sigma_1^2 + \sigma_2^2}$ szórással.

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, \dots, X_n független normális eloszlásúak m várható értékkel és σ szórással, akkor

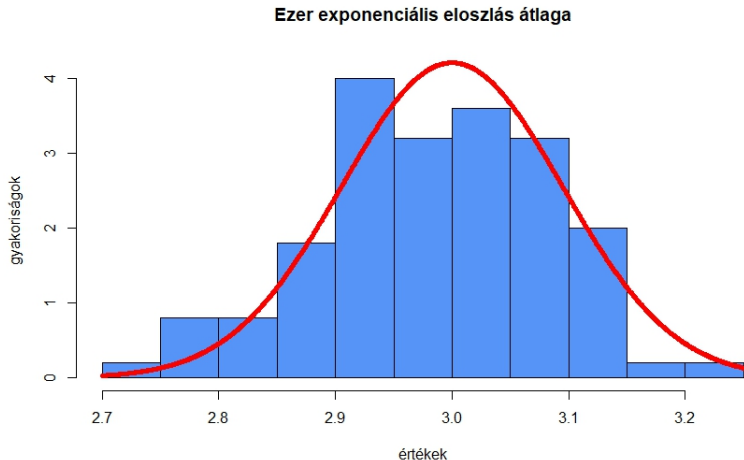
$$\frac{X_1 + \dots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

Exponenciális eloszlás



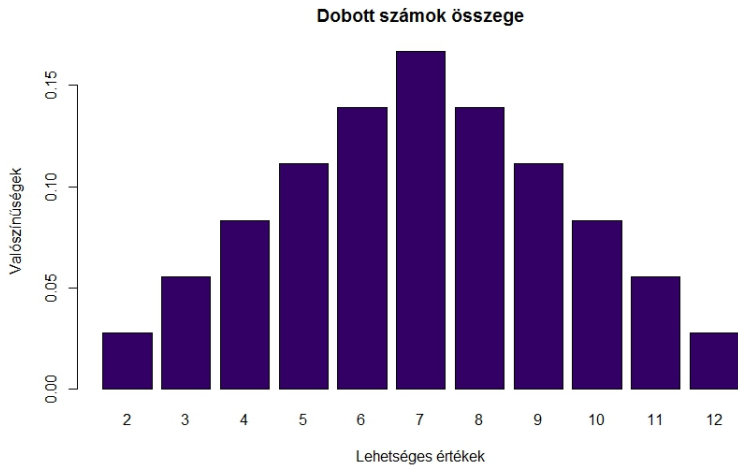
Száz független $\lambda = 1/3$ paraméterű exponenciális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz $e^{-1/3}/3$ ($\mathbb{E}(X) = D(X) = 3, \bar{x} = 3,03, s_n^* = 2,89$)

Exponenciális eloszlások átlaga



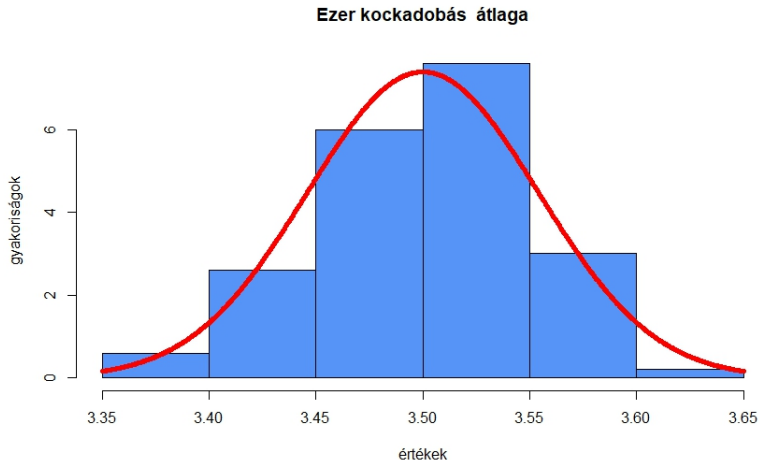
Százelemű minta az alábbi eloszlásból: $n = 1000$ független exponenciális eloszlású ($\lambda = 1/3$) valószínűségi változó átlaga, és az $N(3, 3/\sqrt{1000})$ normális eloszlás sűrűségfüggvénye ($\bar{x} = 2,98, s_n^* = 0,098, \sigma/\sqrt{n} = 0,095$)

Két kockadobás összege



Két szabályos kockadobás összegének eloszlása

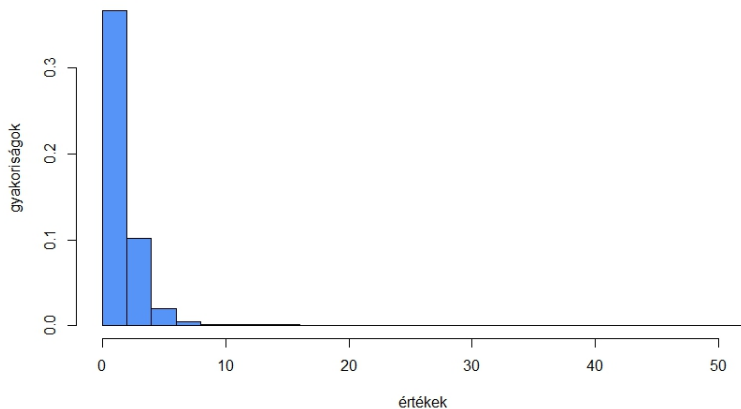
Kockadobások átlaga



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

Exponenciális eloszlás a kitevőben

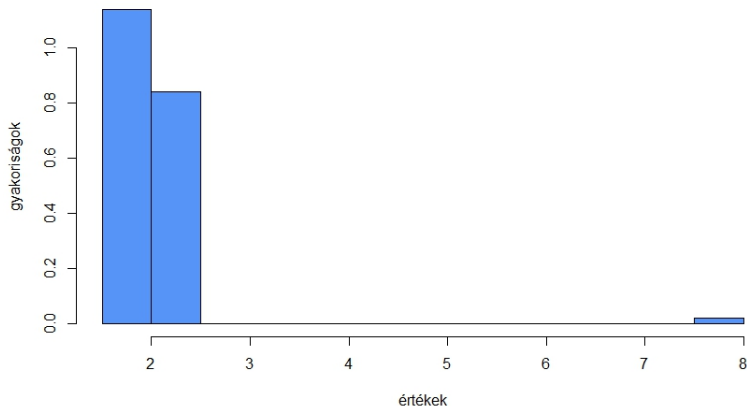
Exponenciális eloszlás a kitevőben



$e^{X_1}, e^{X_2}, \dots, e^{X_{1000}}$ 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.

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, σ/\sqrt{n} szórású normális eloszláshoz.

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 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)