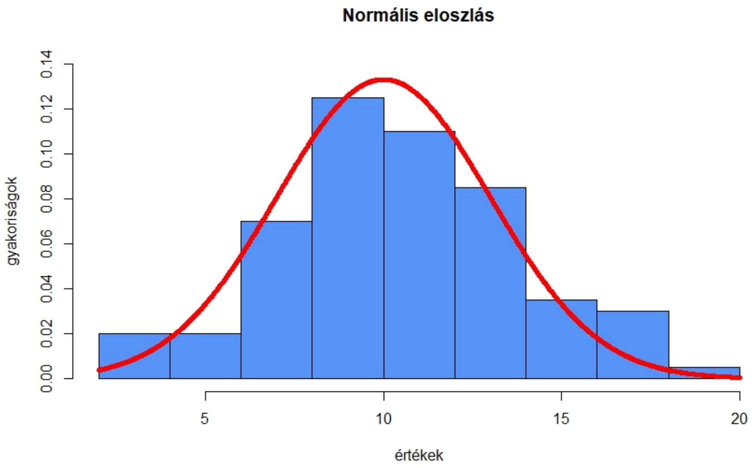
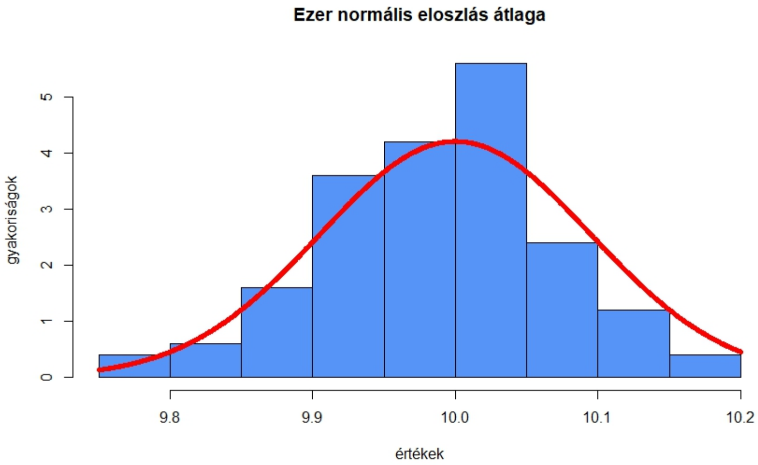
Normális eloszlás



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

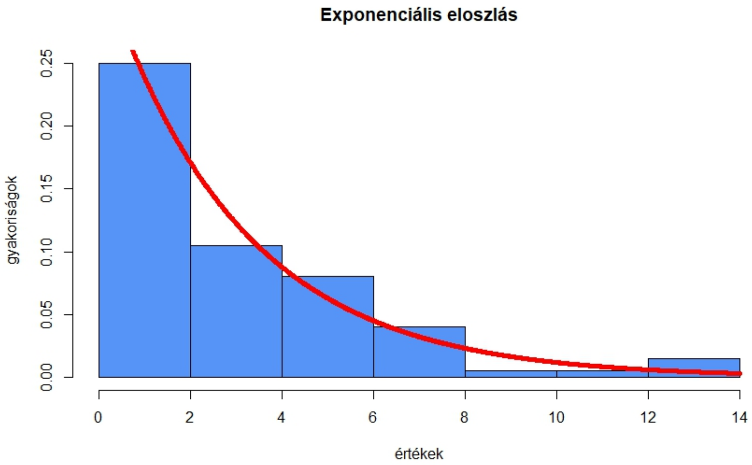
- $X + b$ eloszlása normális, $m_1 + b$ várható értékkel és σ szórással;
- aX eloszlása normális am_1 várható értékkel és $|a|\sigma$ szórással;
- $X + Y$ eloszlása normális, $m_1 + m_2$ várható értékkel és $\sqrt{\sigma_1^2 + \sigma_2^2}$ szórással.

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

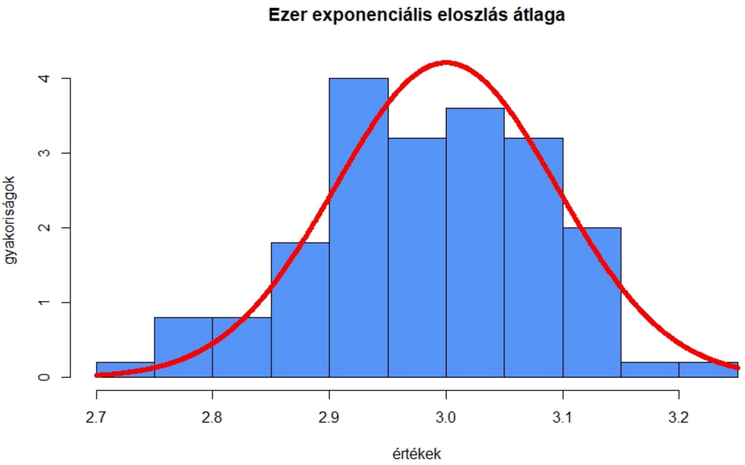
$$\frac{X_1 + \ldots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

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 $N(3, 3/\sqrt{1000})$ normális eloszlás sűrűségfüggvénye ($\overline{x} = 2,98, s_n^* = 0,098, \sigma/\sqrt{n} = 0,095$)

Házi feladat december 4., 10:15-ig

Válasszunk egy n pozitív egészt (legalább 100), és $\lambda > 0$-t.

Sorsoljunk $X_1, \ldots, X_n$ és $Y_1, \ldots, Y_n$ valószínűségi változókat, melyek mind függetlenek, exponenciális eloszlásúak λ paraméterrel.

```
> n=500
```

```
> lambda=3
```

```
> x <- rexp(n, lambda)
```

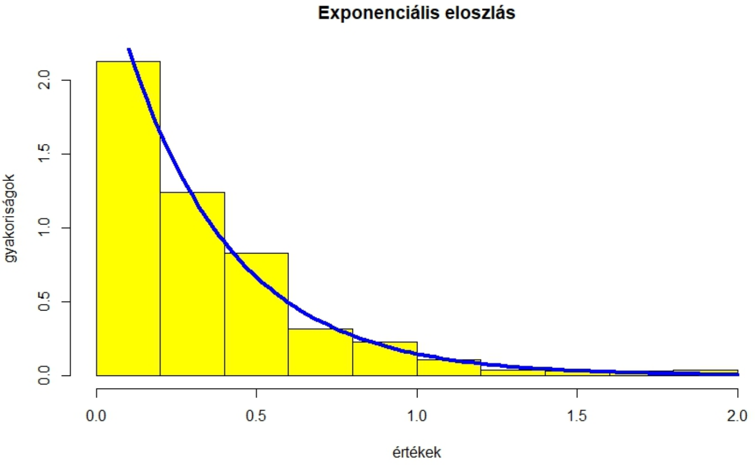
```
> y <- rexp(n, lambda)
```

(a) Készítsünk hisztogramot az $X_1, X_2, \ldots, X_n$ értékekből, és hasonlítsuk ezt össze a λ paraméterű exponenciális eloszlás sűrűségfüggvényével.

```
> hist(x, col="yellow", main="Exponenciális eloszlás",
xlab="értékek", ylab="gyakoriságok", freq=F)
```

```
> curve(dexp(x, rate=lambda), from=0, to=2, add=TRUE, lwd="4",
col="blue")
```

Házi feladat december 4., 10:15-ig



Exponenciális eloszlás $\lambda = 3$ paraméterrel: 500 elemű minta hisztogramja és sűrűségfüggvény: $f(x) = 3\exp(-3x)\mathbb{I}(x > 0)$

Házi feladat december 4., 10:15-ig

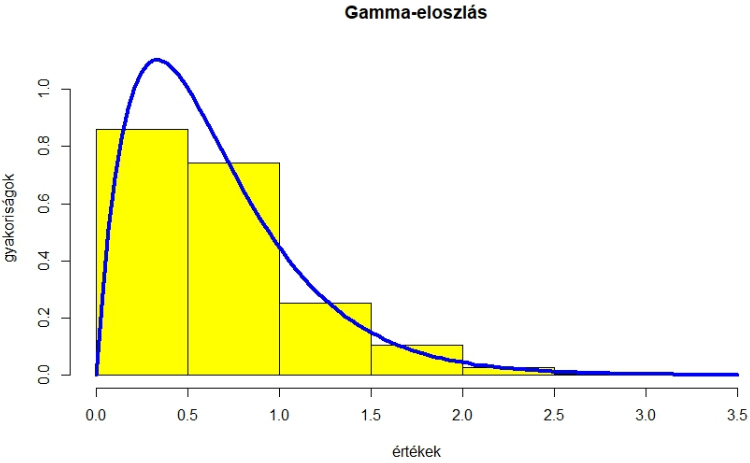
Válasszunk egy n pozitív egészt (legalább 100), és $\lambda > 0$-t.

Sorsoljunk $X_1, \ldots, X_n$ és $Y_1, \ldots, Y_n$ valószínűségi változókat, melyek mind függetlenek, exponenciális eloszlásúak λ paraméterrel.

(b) Készítsünk hisztogramot az $X_1 + Y_1, X_2 + Y_2, \ldots, X_n + Y_n$ értékekből, és hasonlítsuk ezt össze az $a = 2$ rendű és λ paraméterű gamma-eloszlás sűrűségfüggvényével.

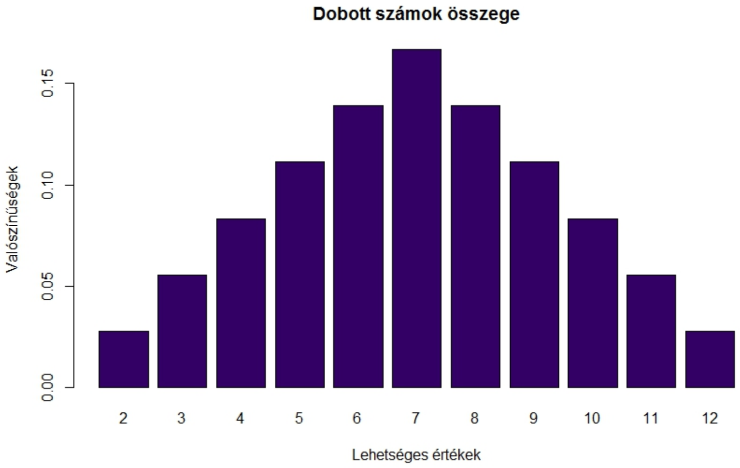
```
> n=500

> lambda=3

> x <- rexp(n, lambda)

> y <- rexp(n, lambda)

> hist(x+y, col="yellow", main="Gamma-eloszlás",
xlab="értékek", ylab="gyakoriságok", freq=F, ylim=c(0,1.1))

> curve(dgamma(x,shape=2, rate=lambda), from=0,
to=3.5, add=TRUE, lwd="4", col="blue")
```

Házi feladat december 4., 10:15-ig



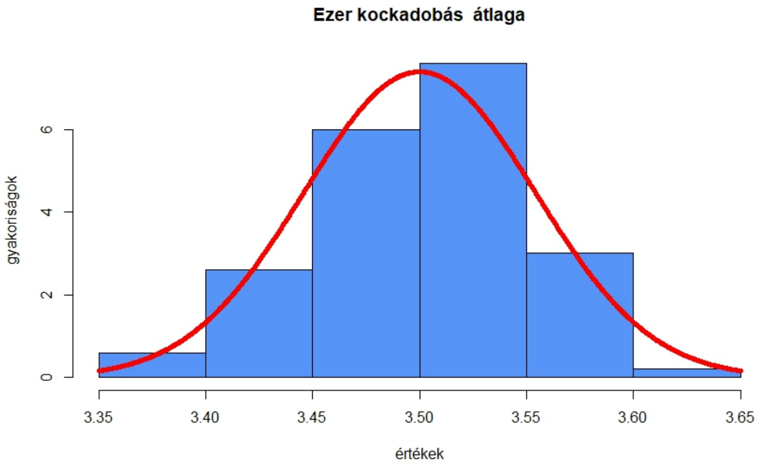
Exponenciális eloszlások összege: gamma-eloszlás $a = 2$ renddel és $\lambda = 3$ paraméterrel: 500 elemű minta hisztogramja és sűrűségfüggvény: $f(x) = 3^2 \cdot x \cdot e^{-3x} \mathbb{I}(x > 0)$

Két kockadobás összege



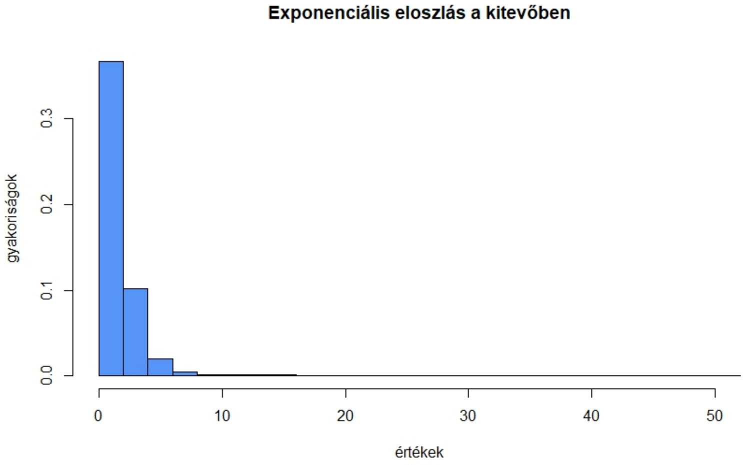
Két szabályos kockadobás összegének eloszlása

Kockadobások átlaga



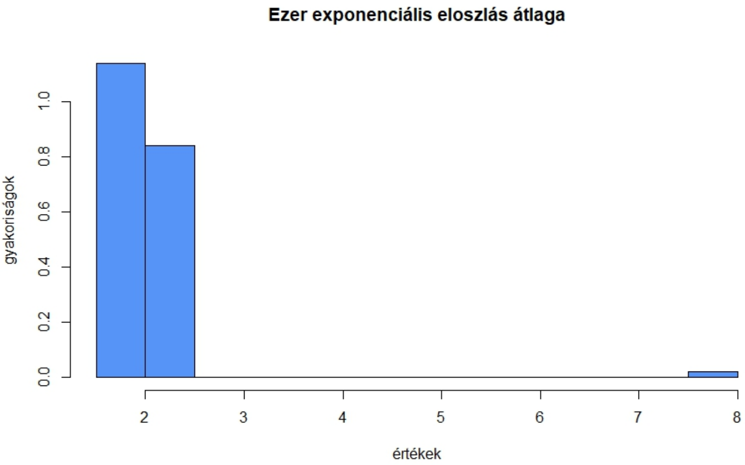
Százelemű minta az alábbi eloszlásból: $n = 1000$ független szabályos kockadobás átlaga, és az $N(3,5, D(X_1)/\sqrt{1000})$ normális eloszlás sűrűségfüggvénye ($\overline{x} = 3,501, s_n^* = 0,098, \sigma/\sqrt{n} = 0,051$)

Exponenciális eloszlás a kitevőben



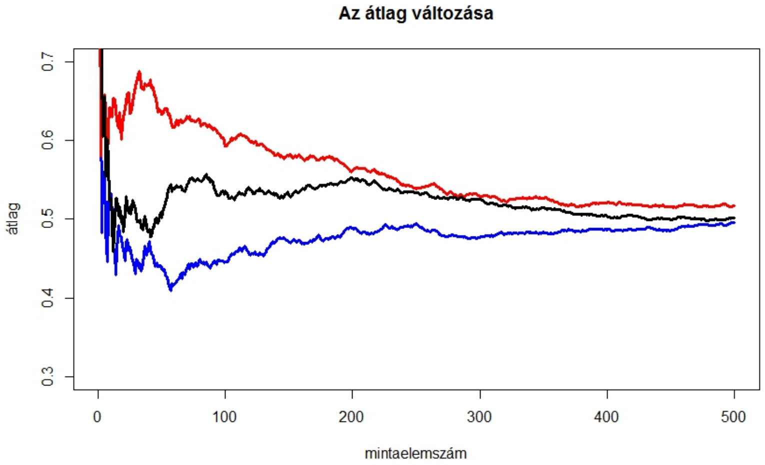
$e^{X_1}, e^{X_2}, \ldots, e^{X_{1000}}$ hisztogramja, ahol X_i-k függetlenek, 2 paraméterű exponenciális eloszlásúak $(\mathbb{E}(e^{X_1}) = 2, D(e^{X_1}) = \infty, \overline{x} = 1,99, s_n^* = 2,33)$

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 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.*

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

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

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

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

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

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

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 $\mathbf{m = 2}$ 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:

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

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

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

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

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 $p(1-p) \leq 1/4$:

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

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

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

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

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)

Házi feladat december 11-ig

Válasszunk egy tetszőleges véges szórású eloszlást, illetve $n \geq 100$ egészt. Legyen $X_1, X_2, \ldots, X_n$ független azonos eloszlású minta a kiválasztott eloszlásból. Az eloszlás várható értéke legyen m, szórása pedig σ.

Tekintsünk az

$$Y_n = \frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}}$$

mennyiséget.

Ezt a kísérletet ismételjük meg 100-szor, és készítsünk hisztogramot a kapott értékekből.

A hisztogramot hasonlítsuk össze a standard normális eloszlás sűrűségfüggvényével.