

A kovariancia (6. előadás)

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor X és Y **kovarianciája**:

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))).$$

A kovariancia (6. előadás)

Legyenek X és Y olyan valószínűségi változók, melyeknek szórása létezik. Ekkor X és Y **kovarianciája**:

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))).$$

- **A kovariancia kiszámítása:**

$$\text{cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y).$$

- **Szimmetria.** $\text{cov}(X, Y) = \text{cov}(Y, X)$.
- **Kapcsolat a szórásnégyzettel.** $\text{cov}(X, X) = D^2(X)$.
- **Függetlenséggel való kapcsolat.** Ha az X és Y valószínűségi változók **függetlenek**, akkor $\text{cov}(X, Y) = 0$. Azaz

$$X, Y \text{ függetlenek} \quad \Rightarrow \quad \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Fordítva nem igaz: $\text{cov}(X, Y) = 0$ esetén nem biztos, hogy X és Y függetlenek. Ha $\text{cov}(X, Y) = 0$, akkor azt mondjuk, hogy X és Y **korrelálatlanok**.

A kovariancia tulajdonságai

- Konstanssal való kovariancia. $\text{cov}(X, c) = 0$, ha c valós szám.
- **Linearitás.** Egyrészt

$$\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z),$$

másrészt tetszőleges $c \in \mathbb{R}$ valós számra

$$\text{cov}(c \cdot X, Y) = c \cdot \text{cov}(X, Y).$$

- **Összeg szórásnégyzete.** $D^2(X + Y) = D^2(X) + D^2(Y) + 2\text{cov}(X, Y)$. Ezért:
 X, Y függetlenek $\Rightarrow D^2(X + Y) = D^2(X) + D^2(Y)$.

Továbbá

$$D^2\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n D^2(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j).$$

- Különbség szórásnégyzete. $D^2(X - Y) = D^2(X) + D^2(Y) - 2\text{cov}(X, Y)$.

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X ;
- a B újságból eladott példányok száma Y .
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y -é 180.
- Az A újság ára 300 forint, a B -é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája?

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X ;
- a B újságból eladott példányok száma Y .
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y -é 180.
- Az A újság ára 300 forint, a B -é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája? Azaz mennyi $\text{cov}(X + Y, 300X + 400Y)$?

A számolás előtt: **pozitív**, **negatív** vagy **0** kovarianciára tippelnénk?

Kovariancia: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X ;
- a B újságból eladott példányok száma Y .
- Tegyük fel, hogy X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, Y -é 180.
- Az A újság ára 300 forint, a B -é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a kovarianciája? Azaz mennyi $\text{cov}(X + Y, 300X + 400Y)$?

A számolás előtt: **pozitív**, **negatív** vagy **0** kovarianciára tippelnénk?

Mivel **minél nagyobb** a példányszám, „**valószínűleg**” **annál nagyobb a bevétel**, **pozitív** kovarianciára számíthatunk.

Kovariancia: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y -é 180. Ekkor az **eladott példányok számának** és a **bevételnek** a kovarianciája:

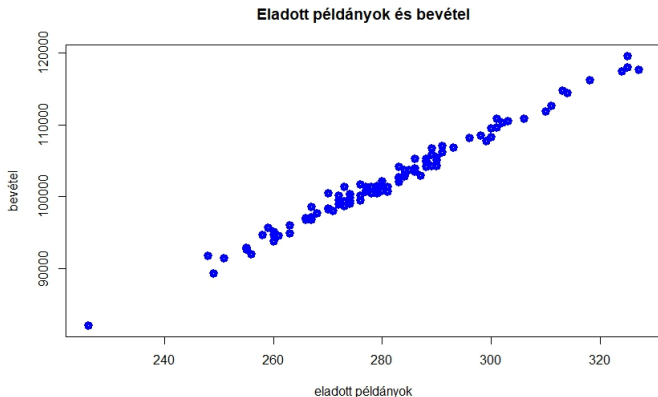
$$\begin{aligned}\text{cov}(X + Y, 300X + 400Y) &\stackrel{(a)}{=} \text{cov}(X, 300X) + \text{cov}(X, 400Y) + \\ &\quad + \text{cov}(Y, 300X) + \text{cov}(Y, 400Y) = \\ &\stackrel{(a,b)}{=} 300 \cdot \text{cov}(X, X) + 400 \cdot \text{cov}(Y, Y) = \\ &\stackrel{(b)}{=} 300D^2(X) + 400D^2(Y) = \\ &\stackrel{(c)}{=} 300 \cdot 100 + 400 \cdot 180 = \mathbf{102000},\end{aligned}$$

ahol felhasználtuk, hogy (a) a kovariancia **lineáris**;

(b) **független** valószínűségi változók kovarianciája **0**, illetve egy valószínűségi változó saját magával vett kovarianciája a szórásnégyzete;

(c) egy λ paraméterű **Poisson-eloszlású** valószínűségi változó **szórásnégyzete** λ .

Kovariancia: példa



A bevétel ($300X + 400Y$) és az eladott példányszám ($X + Y$) együttes előfordulása $n = 100$ független megfigyelésből. Kovariancia: $\text{cov}(X + Y, 300X + 400Y) = 102000$.

Korrelációs együttható: bevezetés

- A **kovariancia** bevezetésének célja, hogy két valószínűségi változó közötti **összefüggés erősségét** tudjuk mérni.
- A korábbi példában: a példányszám és a bevétel kovarianciája **102000** volt.
- Viszont ha a bevételt nem forintban, hanem ezer forintos egységben mérjük:

X : példányszám Y : bevétel forintban Z : bevétel ezer forintban,

akkor

$$\text{cov}(X, Z) = \text{cov}\left(X, \frac{Y}{1000}\right) = \frac{\text{cov}(X, Y)}{1000} = \mathbf{102}.$$

Vagyis a kovariancia a **mértékegységtől függ** \Rightarrow hasznos egy olyan mennyiség, ami szintén az összefüggés erősségét méri, de a mértékegység választásától függetlenül.

- Ilyen lesz a **korrelációs együttható**.

Korrelációs együttható: definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója**:

$$R(X, Y) = \begin{cases} \frac{\text{cov}(X, Y)}{D(X)D(Y)}, & \text{ha } D(X) > 0, D(Y) > 0; \\ 0, & \text{ha } D(X) = 0 \text{ vagy } D(Y) = 0. \end{cases}$$

Korrelációs együttható: definíció

Legyenek X és Y olyan valószínűségi változók, melyek szórásnégyzete létezik. Ekkor X és Y **korrelációs együtthatója**:

$$R(X, Y) = \begin{cases} \frac{\text{cov}(X, Y)}{D(X)D(Y)}, & \text{ha } D(X) > 0, D(Y) > 0; \\ 0, & \text{ha } D(X) = 0 \text{ vagy } D(Y) = 0. \end{cases}$$

- **Lehetséges értékek.** A korrelációs együttható értéke mindig -1 és 1 között esik:

$$|R(X, Y)| \leq 1.$$

- **Lineáris összefüggés.** Legyen $a > 0$ valós szám, b tetszőleges valós szám. Ekkor

$$R(X, aX + b) = 1 \quad \text{és} \quad R(X, -aX + b) = -1.$$

- Tegyük fel, hogy $|R(X, Y)| = 1$. Ekkor léteznek olyan a és b valós számok, hogy az $Y = aX + b$ egyenlet 1 valószínűséggel teljesül.

Korrelációs együttható: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X ;
- a B újságból eladott példányok száma Y .
- Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y -é 180.
- Az A újság ára 300 forint, a B -é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételnek** a korrelációs együtthatója?

Korrelációs együttható: példa

Egy üzletben az A és B újság forgalmát figyelik.

- Az A újságból egy nap alatt eladott példányok száma X ;
- a B újságból eladott példányok száma Y .
- Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y -é 180.
- Az A újság ára 300 forint, a B -é 400.

Mennyi az összesen **eladott példányok számának** és az ezekből származó **bevételek** a korrelációs együtthatója?

$$\begin{aligned} R(X + Y, 300X + 400Y) &= \frac{\text{cov}(X + Y, 300X + 400Y)}{D(X + Y)D(300X + 400Y)} = \\ &= \frac{102000}{D(X + Y)D(300X + 400Y)} \end{aligned}$$

a korábbi számolás alapján, így a szórásokat kell meghatároznunk.

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y -é 180. Ekkor az **eladott példányok számának** szórása:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y -é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = \mathbf{16,73}.$$

A bevétel szórása:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y -é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = \mathbf{16,73}.$$

A bevétel szórása:

$$\begin{aligned} D(300X + 400Y) &= \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = \mathbf{6148,17}. \end{aligned}$$

Ezek alapján a korrelációs együttható:

Korrelációs együttható: példa

X és Y **függetlenek**, **Poisson-eloszlásúak**, X paramétere 100, az Y -é 180. Ekkor az **eladott példányok számának** szórása:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = \mathbf{16,73}.$$

A bevétel szórása:

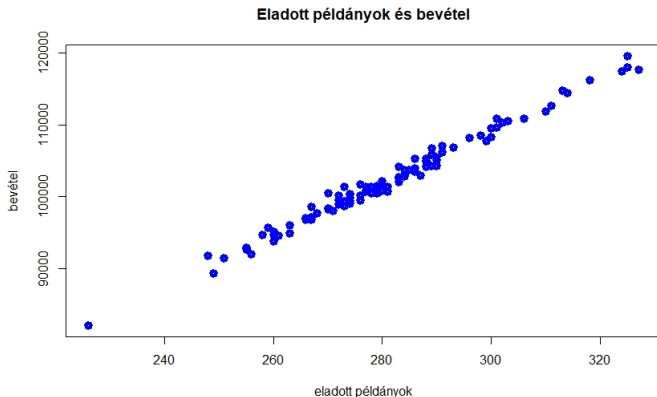
$$\begin{aligned} D(300X + 400Y) &= \sqrt{300^2 D^2(X) + 400^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 400^2 \cdot 180} = \mathbf{6148,17}. \end{aligned}$$

Ezek alapján a korrelációs együttható:

$$\begin{aligned} R(X + Y, 300X + 400Y) &= \frac{\text{cov}(X + Y, 300X + 400Y)}{D(X + Y)D(300X + 400Y)} = \\ &= \frac{\mathbf{102000}}{16,73 \cdot 6148,17} = \mathbf{0,9915}. \end{aligned}$$

A korrelációs együttható lehetséges legnagyobb értéke **1**, így ez **erős pozitív korrelációt** jelent.

Korrelációs együttható: példa



A bevétel ($300X + 400Y$) és az eladott példányszám ($X + Y$) együttes előfordulása $n = 100$ független megfigyelésből. Kovariancia: **102000**, korrelációs együttható: **0,9915**.

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X , a B újságból eladott példányok száma Y . Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y -é 180. Az A újság ára 300 forint, a B -é **4000**. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$\text{cov}(X + Y, 300X + 4000Y) = 300 \cdot 100 + 4000 \cdot 180 = 750000;$$

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = 16,73;$$

$$\begin{aligned} D(300X + 4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 4000^2 \cdot 180} = 53749,42; \end{aligned}$$

$$\begin{aligned} R(X + Y, 300X + 4000Y) &= \frac{\text{cov}(X + Y, 300X + 4000Y)}{D(X + Y)D(300X + 4000Y)} = \frac{750000}{16,73 \cdot 53749,42} \\ &= 0,083. \end{aligned}$$

Korrelációs együttható: példa.

Példa. Egy üzletben az A és B újság forgalmát figyelik. Legyen az A újságból egy nap alatt eladott példányok száma X , a B újságból eladott példányok száma Y . Tegyük fel, hogy X és Y függetlenek, Poisson-eloszlásúak, X paramétere 100, Y -é 180. Az A újság ára 300 forint, a B -é **4000**. Mennyi az összesen eladott példányok számának és az ezekből származó bevételnek a korrelációs együtthatója?

$$\text{cov}(X + Y, 300X + 4000Y) = 300 \cdot 100 + 4000 \cdot 180 = 750000;$$

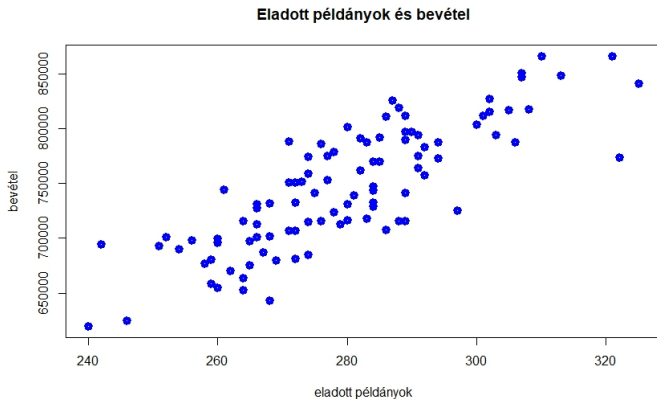
$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{100 + 180} = 16,73;$$

$$\begin{aligned} D(300X + 4000Y) &= \sqrt{300^2 D^2(X) + 4000^2 D^2(Y)} = \\ &= \sqrt{300^2 \cdot 100 + 4000^2 \cdot 180} = 53749,42; \end{aligned}$$

$$\begin{aligned} R(X + Y, 300X + 4000Y) &= \frac{\text{cov}(X + Y, 300X + 4000Y)}{D(X + Y)D(300X + 4000Y)} = \frac{750000}{16,73 \cdot 53749,42} \\ &= 0,083. \end{aligned}$$

A korrelációs együttható értéke majdnem 0, azaz nincs jelentős korreláció az eladott példányok száma és a bevétel között, ha az újságok ára nagyon eltérő.

Korrelációs együttható: példa



A bevétel ($300X + 4000Y$) és az eladott példányszám ($X + Y$) együttes előfordulása $n = 100$ megfigyelésből. Kovariancia: 750000, korrelációs együttható: 0,083.

Az átlag várható értéke és szórása

A statisztikában alapvető kérdés, hogy ha

- **ugyanazt a mérést**
- sokszor, **egymástól függetlenül** megismételjük,
- majd a kapott eredményeket **átlagoljuk**,
- akkor az átlag, mint valószínűségi változó hogyan viselkedik

Vagyis: X_1, X_2, \dots, X_n **független**, **azonos eloszlású** valószínűségi változók, akkor mit mondhatunk az

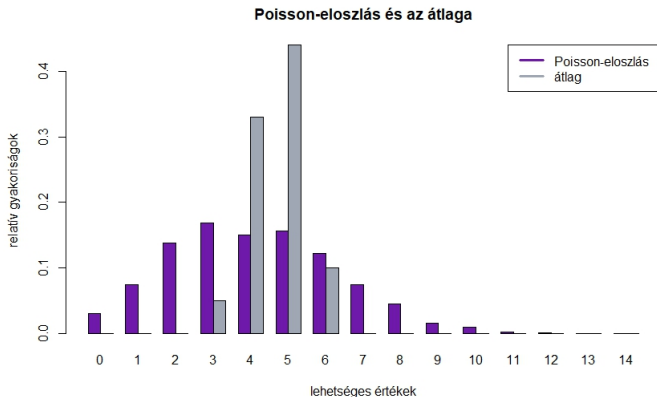
$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

átlagról: mennyi a **várható értéke** és mennyi a **szórása**?

Azonos eloszlás: $\mathbb{P}(X_j \in A) = \mathbb{P}(X_1 \in A)$ tetszőleges j -re és $A \subseteq \mathbb{R}$ „megfelelő” halmazra.

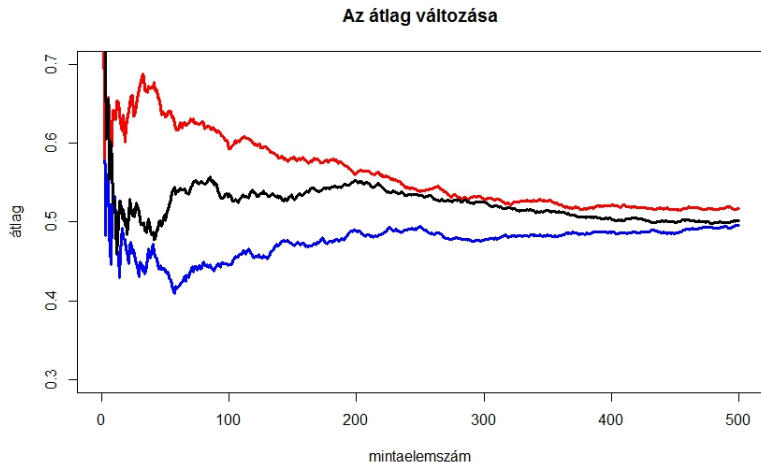
azonos eloszlás \Rightarrow **azonos várható érték, azonos szórás**

Az átlag viselkedése



1000 darab $\lambda = 5$ paraméterű Poisson-eloszlású valószínűségi változó, illetve 100 darab, tíz független, $\lambda = 5$ paraméterű Poisson-eloszlású valószínűségi változó átlagaként előálló megfigyelés hisztogramja → **átlagolásnál a várható érték nem változik**, ez mindkét esetben 5, **a szórás csökken**

Az átlag konvergenciája



Az átlagok sorozata: $[0, 1]$ intervallumon egyenletes eloszlásból vett minta átlagai-nak sorozata $n = 500$ -ig; itt a **várható érték 0,5** \rightarrow az átlagok ehhez esnek egyre közelebb

Az átlag várható értéke

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyek várható értéke $\mathbf{m} = \mathbb{E}(\mathbf{X}_1) < \infty$. Ekkor

$$\mathbb{E}(\bar{\mathbf{X}}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mathbf{m}.$$

Bizonyítás.

Az átlag várható értéke

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyek várható értéke $\mathbf{m} = \mathbb{E}(\mathbf{X}_1) < \infty$. Ekkor

$$\mathbb{E}(\bar{\mathbf{X}}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mathbf{m}.$$

Bizonyítás.

$$\mathbb{E}(\bar{\mathbf{X}}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} \cdot \mathbb{E}(X_1 + \dots + X_n) = \frac{1}{n} \cdot n \cdot \mathbf{m} = \mathbf{m}.$$

Felhasználtuk a várható érték **linearitását**, és hogy csak az eloszlástól függ:

- $\mathbb{E}(cX) = c\mathbb{E}(X)$, ha $c \in \mathbb{R}$;
- $\mathbb{E}(Y + Z) = \mathbb{E}(Y) + \mathbb{E}(Z)$;
- ha Y és Z eloszlása azonos, akkor $\mathbb{E}(Y) = \mathbb{E}(Z) \Rightarrow$ itt mindegyik várható érték m .

Az átlag szórása

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $\sigma = D(\mathbf{X}_1) < \infty$. Ekkor

$$D(\bar{\mathbf{X}}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

Bizonyítás.

Az átlag szórása

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $\sigma = \mathbf{D}(X_1) < \infty$. Ekkor

$$\mathbf{D}(\bar{X}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

Bizonyítás.

$$\mathbf{D}(\bar{X}) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{D(X_1 + \dots + X_n)}{n} = \frac{\sqrt{n\sigma^2}}{n} = \frac{\sigma}{\sqrt{n}}.$$

Felhasználtuk a szórás alábbi tulajdonságait:

- $D(cX) = |c|D(X)$, ha $c \in \mathbb{R}$;
- $D^2(Y + Z) = D^2(Y) + D^2(Z)$, ha Y és Z **függetlenek**;
- ha Y és Z eloszlása megegyezik, akkor $D(Y) = D(Z)$, ezért itt mindegyik szórás σ

A nagy számok erős törvénye

A nagy számok törvényei azt állítják, hogy az **átlagok sorozata** a **várható értékhez** konvergál, megfelelő feltételek mellett.

A nagy számok erős törvénye Legyenek X_1, X_2, \dots valószínűségi változók, melyek **függetlenek**, **azonos eloszlásúak**, **véges várható értékűek**. Legyen $m = \mathbb{E}(X_1) < \infty$. 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.

Az átlag viselkedése

Legyenek X_1, X_2, \dots, X_n **független, azonos eloszlású** valószínűségi változók, melyekre

$$\mathbb{E}(X_j) = m, \quad D(X_j) = \sigma.$$

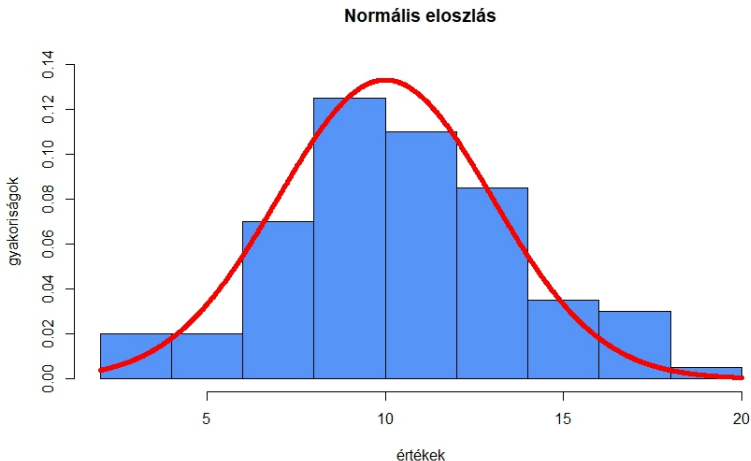
Azt láttuk, hogy ha az **átlaguk** $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$, akkor

$$\mathbb{E}(\bar{X}) = m, \quad D(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Vagyis átlagolásnál a **várható érték nem változik**, a **szórás csökken**.

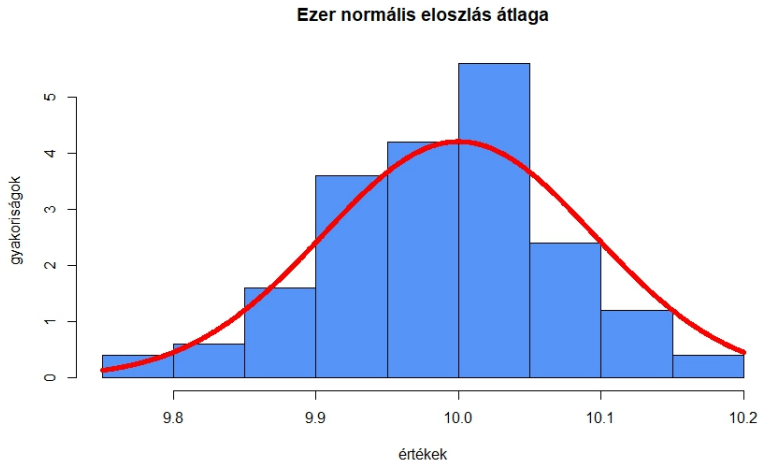
Hogyan függ az átlag eloszlása az X_j valószínűségi változók eloszlásától?

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

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:

- $X + b$ 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.

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:

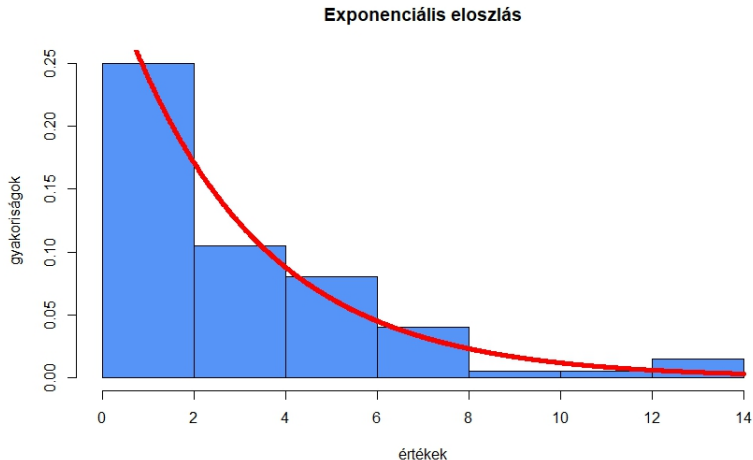
- $X + b$ eloszlása normális, $m_1 + b$ várható értékkel és σ szórással;
- aX eloszlása normális am_1 várható értékkel és $|a|\sigma$ szórással;
- $X + Y$ eloszlása normális, $m_1 + m_2$ várható értékkel és $\sqrt{\sigma_1^2 + \sigma_2^2}$ szórással.

Ebből következik: ha X_1, \dots, X_n **függetlenek**, **normális eloszlásúak** m várható értékkel és σ szórással, akkor

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

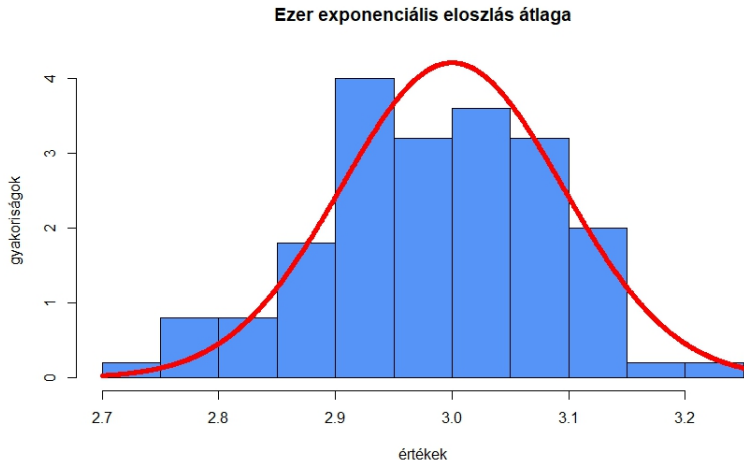
Ennek megfelelően a hisztogram közel volt az ennek megfelelő sűrűségfüggvényhez.

Exponenciális eloszlás



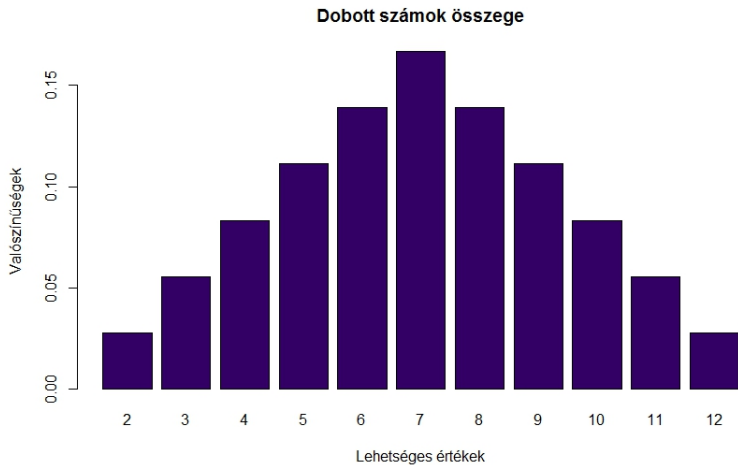
Száz független $\lambda = 1/3$ paraméterű exponenciális eloszlású valószínűségi változó hisztogramja és a sűrűségfüggvény, azaz $e^{-1/3}/3$ ($\mathbb{E}(X) = D(X) = 3, \bar{x} = 3,03, s_n^* = 2,89$)

Exponenciális eloszlások átlaga



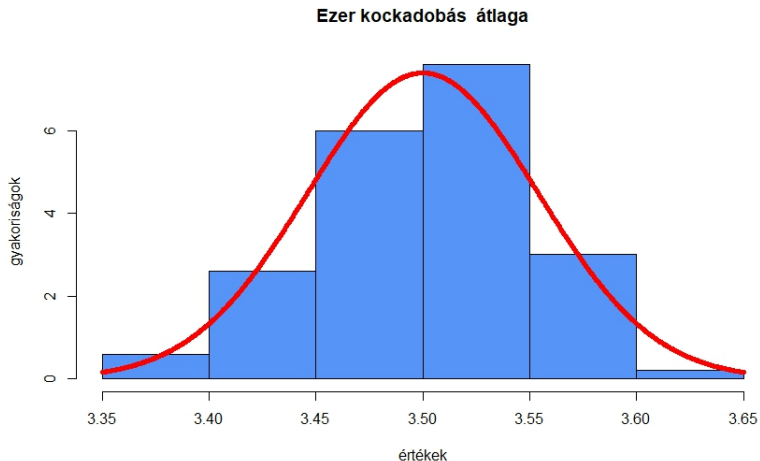
Az átlag eloszlása százelemű mintából, ha X_j eloszlása **exponenciális** $\lambda = 1/3$ -dal, és $n = 1000$, és az $N(3, 3/\sqrt{1000})$ **normális** eloszlás sűrűségfüggvénye ($\bar{x} = 2,98, s_n^* = 0,098$)

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 ($\bar{x} = 3,501, s_n^* = 0,098, \sigma/\sqrt{n} = 0,051$)

Centrális határeloszlástétel

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

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

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

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

teljesül $n \rightarrow \infty$ esetén eloszlásban \rightarrow a normalizált összeg eloszlása „közel van” a standard normálishoz

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**.