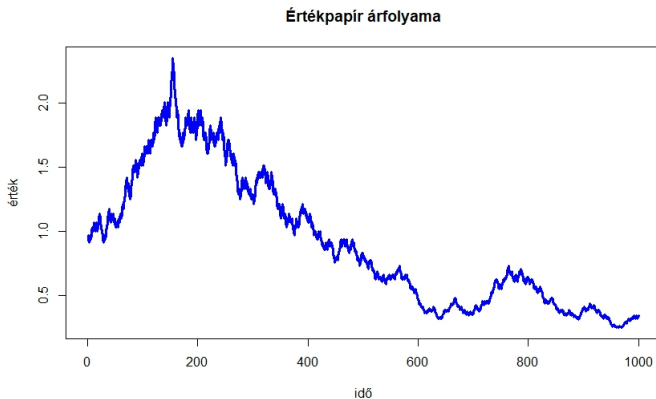


Eloszlásfüggvény: példa (8. előadás)



Egy elképzelt értékpapír árfolyama 1000 napon keresztül, 1000 forintban

Eloszlásfüggvény: bevezetés

- X valószínűségi változó: egy véletlen kísérlet eredménye
- eddig: X **diszkrét**, és a $\mathbb{P}(X = x)$ valószínűségekkel lehet leírni az eloszlását
- ha a lehetséges értékek halmaza „túl nagy”, vagy a valószínűségek „túl kicsik”, ez nem informatív
- például: X az értékpapír árfolyama holnap, $\mathbb{P}(X = 784) = 0,0038$, $\mathbb{P}(X = 785) = 0,004$, stb. egy előrejelzés szerint \rightarrow ennél hasznosabb információ lehet, hogy

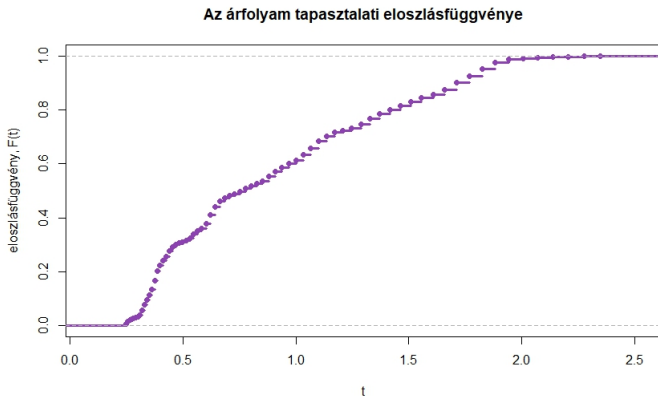
$$\mathbb{P}(X \leq 785) = 0,5,$$

azaz az értékpapír 50% valószínűséggel nem haladja meg a 785 szintet

- **eloszlásfüggvény**: $F(t)$ annak valószínűsége, hogy **a valószínűségi változó értéke legfeljebb t** , azaz

$$F(t) = \mathbb{P}(X \leq t).$$

Eloszlásfüggvény: példa



t függvényében a t -nél nem nagyobb árfolyamú napok aránya az előző példában

Eloszlásfüggvény: definíció

Legyen $X : \Omega \rightarrow \mathbb{R}$ valószínűségi változó. Ekkor X **eloszlásfüggvénye** az alábbi $F : \mathbb{R} \rightarrow [0, 1]$ függvény:

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\})$$

tetszőleges $t \in \mathbb{R}$ valós számra.

Ez **minden valószínűségi változóra** és minden $t \in \mathbb{R}$ valós számra értelmes: éppen úgy definiáltuk a valószínűségi változót, hogy $\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}$ esemény, tehát van valószínűsége.

Eloszlásfüggvény: példa

Valakinek három gyereke születik, a gyerekek mindegyike egymástól függetlenül $1/2$ valószínűséggel fiú. Nyolc egyformán valószínű eset van:

$$\{\text{LLL, FLL, LFL, LLF, FFL, FLF, LFF, FFF}\}$$

Legyen X a fiúk száma. X diszkrét valószínűségi változó, lehetséges értékei: 0, 1, 2, 3, és

$$\mathbb{P}(X = 0) = \frac{1}{8}, \quad \mathbb{P}(X = 1) = \frac{3}{8}, \quad \mathbb{P}(X = 2) = \frac{3}{8}, \quad \mathbb{P}(X = 3) = \frac{1}{8}.$$

Az X **eloszlásfüggvényének**, F -nek az értéke néhány helyen:

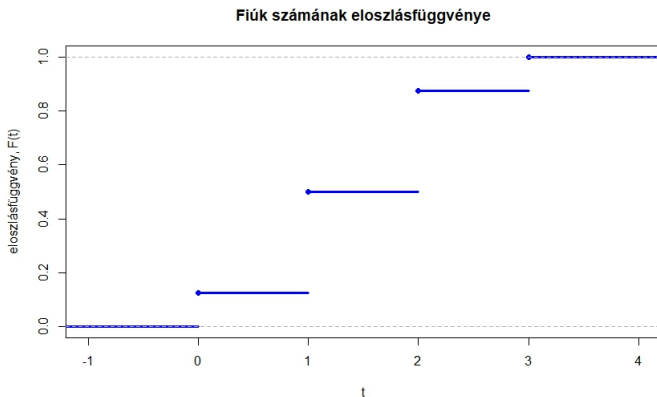
$$F(0) = \mathbb{P}(X \leq 0) = \frac{1}{8};$$

$$F(1) = \mathbb{P}(X \leq 1) = \frac{1}{2};$$

$$F(2, 4) = \mathbb{P}(X \leq 2, 4) = \frac{7}{8};$$

$$F(4) = \mathbb{P}(X \leq 4) = 1.$$

Eloszlásfüggvény: példa



Három gyerek közül a fiúk számának eloszlásfüggvénye vízszintes: t , függőleges: $F(t) = \mathbb{P}(X \leq t)$.

Eloszlásfüggvény

Definíció (Eloszlásfüggvény)

Legyen $X : \Omega \rightarrow \mathbb{R}$ valószínűségi változó. Ekkor X eloszlásfüggvénye az alábbi $F : \mathbb{R} \rightarrow [0, 1]$ függvény:

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \quad \text{minden } t \in \mathbb{R} \text{ valós számra.}$$

Az eloszlásfüggvény minden t valós számhoz hozzárendeli, hogy mennyi annak valószínűsége, hogy a valószínűségi változó értéke legfeljebb t . Például ha X a fiúk száma három gyerek közül:

$$F(1) = \mathbb{P}(X \leq 1) = \mathbb{P}(\text{legfeljebb egy fiú van}) = 1/2;$$

$$F(2) = \mathbb{P}(X \leq 2) = \mathbb{P}(\text{legfeljebb két fiú van}) = 7/8;$$

$$F(2,3) = \mathbb{P}(X \leq 2, 3) = \mathbb{P}(\text{legfeljebb 2,3 fiú van}) = 7/8;$$

Eloszlásfüggvény

Definíció (Eloszlásfüggvény)

Legyen $X : \Omega \rightarrow \mathbb{R}$ valószínűségi változó. Ekkor X eloszlásfüggvénye az alábbi $F : \mathbb{R} \rightarrow [0, 1]$ függvény:

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \quad \text{minden } t \in \mathbb{R} \text{ valós számra.}$$

Az eloszlásfüggvény minden t valós számhoz hozzárendeli, hogy mennyi annak valószínűsége, hogy a valószínűségi változó értéke legfeljebb t . Például ha X a fiúk száma három gyerek közül:

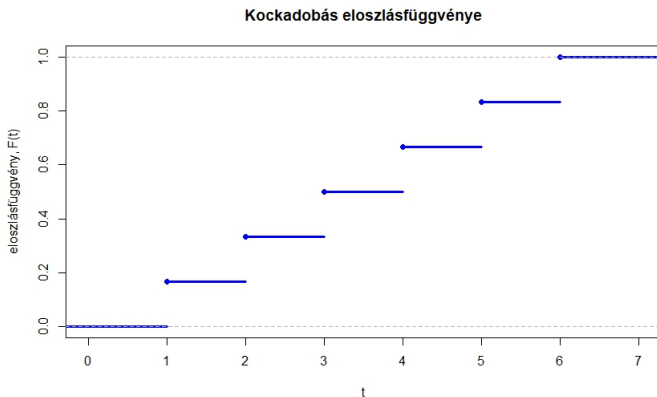
$$F(1) = \mathbb{P}(X \leq 1) = \mathbb{P}(\text{legfeljebb egy fiú van}) = 1/2;$$

$$F(2) = \mathbb{P}(X \leq 2) = \mathbb{P}(\text{legfeljebb két fiú van}) = 7/8;$$

$$F(2,3) = \mathbb{P}(X \leq 2,3) = \mathbb{P}(\text{legfeljebb 2,3 fiú van}) = 7/8;$$

Véges értékészletű valószínűségi változók esetén az eloszlásfüggvény lépcsős (véges sok értéket vesz fel), és az ugrások nagyságát az egyes lehetséges értékek valószínűségei adják meg.

Eloszlásfüggvény: példa



Szabályos dobókockával dobott szám eloszlásfüggvénye
vízszintes: t , függőleges: $F(t) = \mathbb{P}(X \leq t)$.

Az eloszlásfüggvény tulajdonságai

Ha $a, b \in \mathbb{R}$ valós számok, és F az X eloszlásfüggvénye, akkor

$$\mathbb{P}(a < X \leq b) = F(b) - F(a),$$

hiszen annak valószínűségét, hogy X az a és b közé esik, megkaphatjuk úgy, hogy $\mathbb{P}(X \leq b)$ -ből levonjuk $\mathbb{P}(X \leq a)$ -t.

Legyen F egy tetszőleges valószínűségi változó eloszlásfüggvénye. Ekkor

- ❶ F monoton növekvő: $a < b$ esetén $F(a) \leq F(b)$.
- ❷ $\lim_{t \rightarrow -\infty} F(t) = 0$; $\lim_{t \rightarrow \infty} F(t) = 1$.
- ❸ F jobbról folytonos, azaz minden $t \in \mathbb{R}$ valós számra $\lim_{s \rightarrow t+} F(s) = F(t)$.

Fordítva: ha F -re érvényesek ezek a tulajdonságok, akkor van olyan X , aminek F az eloszlásfüggvénye.

Eloszlásfüggvény: példa

Legyen X négy rendű $1/2$ paraméterű binomiális eloszlású valószínűségi változó. Mennyi X eloszlásfüggvényének az értéke az 1,5 helyen?

Eloszlásfüggvény: példa

Legyen X négy rendű $1/2$ paraméterű binomiális eloszlású valószínűségi változó. Mennyi X eloszlásfüggvényének az értéke az $1,5$ helyen?

X -re a következőképpen gondolhatunk: $n = 4$ független kísérlet, mindegyik $p = 0,5$ valószínűséggel sikerül, X a **sikeres kísérletek száma**.

Definíció szerint, ha F az X eloszlásfüggvénye, akkor

$$F(1,5) = \mathbb{P}(X \leq 1,5) = \mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1),$$

hiszen X értéke nemnegatív egész. Így

$$F(1,5) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \left(\frac{1}{2}\right)^4 + 4 \cdot \left(\frac{1}{2}\right)^4 = \frac{5}{16}.$$

Egyenletes eloszlás

Tekintsük a következő hétköznapi példát.

- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- Mennyi a valószínűsége, hogy a futár 11 óráig megérkezik?
- **Feltéve, hogy a futár 10 óráig még nem érkezett meg, mennyi a valószínűsége, hogy 11 óra előtt megérkezik?**

Legyen X a futár érkezésének időpontja. Így fogunk tudni számolni:

$$\mathbb{P}(X \leq 11)$$

Egyenletes eloszlás

Tekintsük a következő hétköznapi példát.

- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- Mennyi a valószínűsége, hogy a futár 11 óráig megérkezik?
- **Feltéve, hogy a futár 10 óráig még nem érkezett meg, mennyi a valószínűsége, hogy 11 óra előtt megérkezik?**

Legyen X a futár érkezésének időpontja. Így fogunk tudni számolni:

$$\mathbb{P}(X \leq 11) = \frac{11 - 8}{12 - 8} = \frac{3}{4} = 75\%.$$

$$\mathbb{P}(X \leq 11 | X > 10)$$

Egyenletes eloszlás

Tekintsük a következő hétköznapi példát.

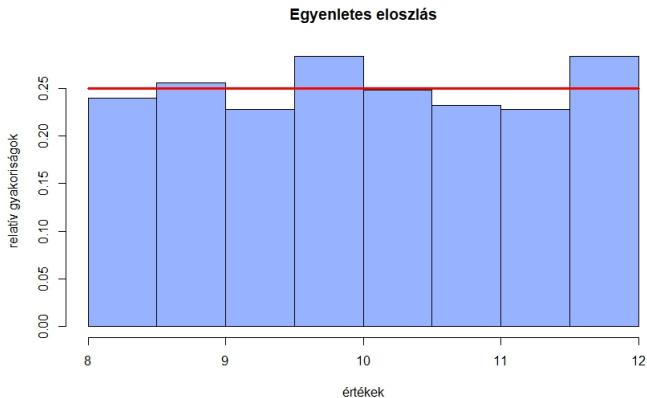
- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- Mennyi a valószínűsége, hogy a futár 11 óráig megérkezik?
- **Feltéve, hogy a futár 10 óráig még nem érkezett meg, mennyi a valószínűsége, hogy 11 óra előtt megérkezik?**

Legyen X a futár érkezésének időpontja. Így fogunk tudni számolni:

$$\mathbb{P}(X \leq 11) = \frac{11 - 8}{12 - 8} = \frac{3}{4} = 75\%.$$

$$\mathbb{P}(X \leq 11 | X > 10) = \frac{\mathbb{P}(\{X \leq 11\} \cap \{X > 10\})}{\mathbb{P}(X > 10)} = \frac{1/4}{2/4} = \frac{1}{2}.$$

Egyenletes eloszlás



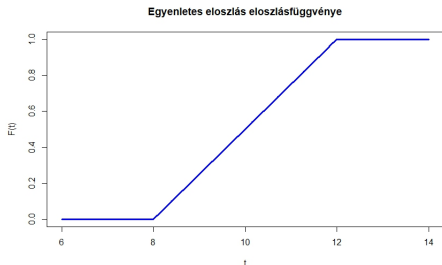
A $[8, 12]$ intervallumon egyenletes eloszlásból vett 500 elemű minta hisztogramja

Egyenletes eloszlás

Definíció (Egyenletes eloszlás (uniform distribution))

Az X valószínűségi változó **egyenletes eloszlású** az $[a, b]$ intervallumon, ha eloszlásfüggvénye:

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0, & \text{ha } t \leq a; \\ \frac{t-a}{b-a}, & \text{ha } a < t < b; \\ 1, & \text{ha } t \geq b. \end{cases}$$



Egyenletes eloszlás

Csomagot várunk, a futár 10 és 12 óra között érkezik. Feltesszük, hogy érkezésének időpontja egyenletes eloszlású a $[10, 12]$ intervallumon. Ekkor az előző állítás alapján az alábbiak igazak ($a = 10, b = 12$).

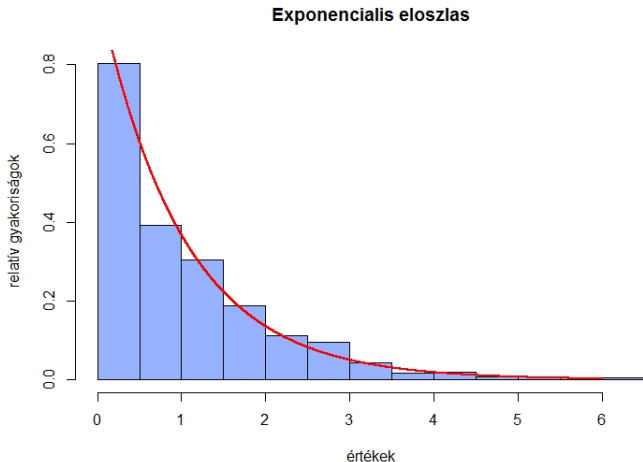
- Annak valószínűsége, hogy 10 és 11 óra között érkezik: $(11 - 10)/(12 - 10) = 1/2$.
- Annak valószínűsége, hogy 10:15 és 10:30 között érkezik: $1/8 = 0,125$.
- Annak valószínűsége, hogy 10 : 30 után érkezik: $3/4 = 0,75$.

Exponenciális eloszlás

Az **exponenciális eloszlás** sokszor használható véletlen időtartamok modellezésére, például

- egy művelet elvégzésének ideje: egy ember kiszolgálása egy boltban, vagy egy számítás elvégzése egy számítógépen
- egy ember reakcióideje
- két esemény bekövetkezése között eltelt idő, például egy üzletben két ügyfél érkezése közötti idő
- járványterjedés modellezésénél: a fertőzés átadásának vagy a gyógyulásnak az ideje
- radioaktív részecske bomlási ideje

Exponenciális eloszlás



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

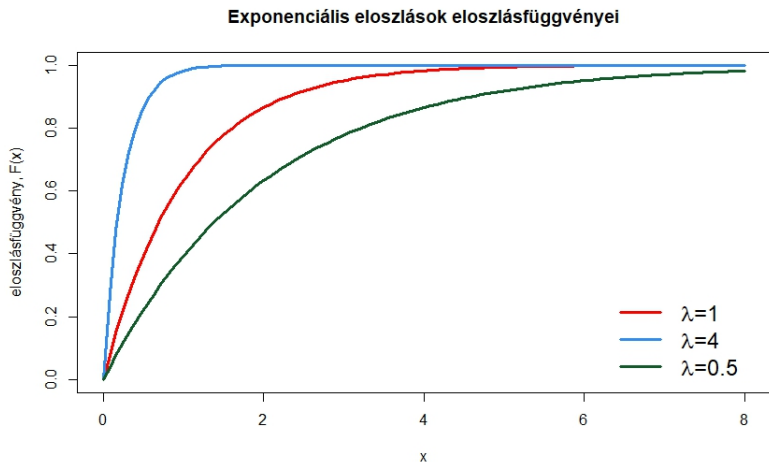
Exponenciális eloszlás

Definíció

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

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

Exponenciális eloszlás



Különböző paraméterű ($\lambda = \frac{1}{2}, 1$, illetve 4) exponenciális eloszlások eloszlásfüggvényei

Exponenciális eloszlás

Példa. Tegyük fel, hogy egy boltban egy vevő kiszolgálásának ideje (percben számolva) $1/3$ **paraméterű 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?

Az eloszlásfüggvény definíciója alapján

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

Exponenciális eloszlás

Példa. Tegyük fel, hogy egy boltban egy vevő kiszolgálásának ideje (percben számolva) $1/3$ **paraméterű 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?

Az eloszlásfüggvény definíciója alapján

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

Hasonlóképpen

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

Exponenciális eloszlás

Állítás (Az exponenciális eloszlás örökifjú tulajdonsága)

Legyen X exponenciális eloszlású valószínűségi változó, s, t pozitív számok.
Ekkor

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

Bizonyítás. A feltételes valószínűség definícióját és az exponenciális eloszlás eloszlásfüggvényének alakját felhasználva

$$\begin{aligned}\mathbb{P}(X \geq s + t | X \geq s) &= \frac{\mathbb{P}(\{X \geq s + t\} \cap \{X \geq s\})}{\mathbb{P}(X \geq s)} = \frac{1 - \mathbb{P}(X < s + t)}{1 - \mathbb{P}(X < s)} = \\ &= \frac{1 - F(s + t)}{1 - F(s)} = \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \\ &= e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) = 1 - F(t) = \mathbb{P}(X \geq t).\end{aligned}$$

□

Pareto-eloszlás

Biztosításmatematikában, vagy például a jövedelmek eloszlásának modellezésére használják gyakran az alábbi eloszlást.

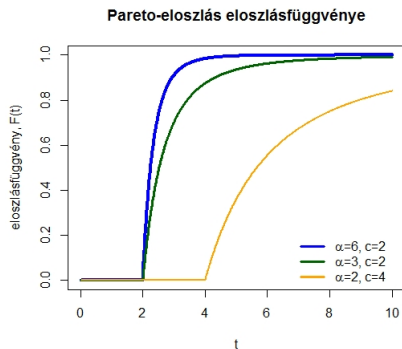
Definíció (Pareto-eloszlás)

Az X valószínűségi változó Pareto-eloszlású $\alpha > 0$ és $c > 0$ paraméterekkel, ha eloszlásfüggvénye

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

A definícióból látszik, hogy a Pareto-eloszlás a c paraméternél kisebb értékeket nem vehet fel. Gyakran használják jövedelmek eloszlásának modellezésére. Ahogy később látni fogjuk, van olyan α , melyre az eloszlás várható értéke nem létezik, és olyan is, amire a várható érték létezik, de a szórás nem.

Pareto-eloszlás



Különböző paraméterű Pareto-eloszlások eloszlásfüggvényei

Kvantilisek

Definíció (Kvantilis)

Legyen $0 \leq z \leq 1$, és X egy valószínűségi változó. Ekkor az X valószínűségi változó z -kvantilise:

$$q_z = \inf\{t \in \mathbb{R} : F(t) \geq z\}.$$

Ha az X eloszlásfüggvénye, azaz az $F(t) = \mathbb{P}(X \leq t)$ függvény folytonos, akkor az igaz, hogy

$$F(q_z) = \mathbb{P}(X \leq q_z) = z,$$

azaz z az a szám, aminél X éppen z valószínűséggel kisebb.

Kvantilisek

Definíció (Kvantilis)

Legyen $0 \leq z \leq 1$, és X egy valószínűségi változó. Ekkor az X valószínűségi változó z -kvantilise:

$$q_z = \inf\{t \in \mathbb{R} : F(t) \geq z\}.$$

Ha az X eloszlásfüggvénye, azaz az $F(t) = \mathbb{P}(X \leq t)$ függvény folytonos, akkor az igaz, hogy

$$F(q_z) = \mathbb{P}(X \leq q_z) = z,$$

azaz z az a szám, aminél X éppen z valószínűséggel kisebb.

A $z = 1/2$ -kvantilis a medián: $\mathbb{P}(X \leq m) = 1/2$.

Kvantilisek

Definíció (Kvantilis)

Legyen $0 \leq z \leq 1$, és X egy valószínűségi változó. Ekkor az X valószínűségi változó z -kvantilise:

$$q_z = \inf\{t \in \mathbb{R} : F(t) \geq z\}.$$

Ha az X eloszlásfüggvénye, azaz az $F(t) = \mathbb{P}(X \leq t)$ függvény folytonos, akkor az igaz, hogy

$$F(q_z) = \mathbb{P}(X \leq q_z) = z,$$

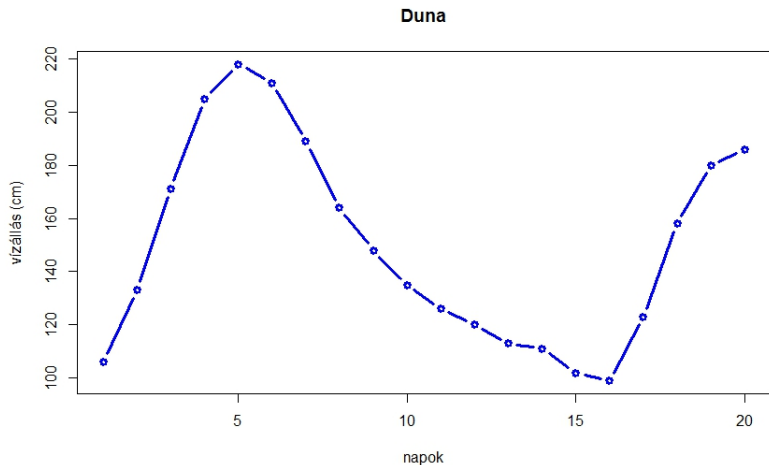
azaz z az a szám, aminél X éppen z valószínűséggel kisebb.

A $z = 1/2$ -kvantilis a medián: $\mathbb{P}(X \leq m) = 1/2$.

A medián fontos tulajdonsága: a $\mathbb{E}(|X - u|)$ érték az $u = m$ esetén a legkisebb, ahol m a medián.

Összehasonlításképpen: az $\mathbb{E}((X - u)^2)$ érték az $u = \mathbb{E}(X)$ esetén a legkisebb.

Valószínűségi vektorváltozó: példa



A Duna vízállása 20 napon keresztül (az adatok forrása: *Országos Vízelző Szolgálat*): $X_1 = 106, X_2 = 133, \dots, X_{20} = 186$

Valószínűségi vektorváltozó

Sok esetben nem egyetlen valószínűségi változó viselkedését vizsgáljuk, hanem több valószínűségi változó **együttes viselkedését**. Például:

- egy véletlen folyamat (tőzsdeindex, egy folyó vízállása, egy ország népessége) különböző időpontokban;
- egy ember (vagy ország, cég stb.) több különböző jellemzője (például egy ember életkora, jövedelme és kiadásai);
- egy mérésorozatban a különböző mérések során megfigyelt értékek (például egy mérést tízszer megismételve tíz különböző valószínűségi változót kapunk).

Valószínűségi változók együttesét **valószínűségi vektorváltozónak** nevezük. Ez állhat **összefüggő** (mint az első két esetben) vagy **független** (mint tipikusan a harmadik esetben) valószínűségi változókból is.

Valószínűségi vektorváltozó

Az

$$\underline{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$$

függvény **valószínűségi vektorváltozó**, ha X_1, X_2, \dots, X_n valószínűségi változók.

- 1000 embert megkérdezzük a havi jövedelméről. Legyen X_i az i . megkérdezett jövedelme. Ekkor $(X_1, X_2, \dots, X_{1000})$ valószínűségi vektorváltozó.
- $(X_1, X_2, \dots, X_{20})$ is valószínűségi vektorváltozó, ahol X_j a Duna vízálása a j . napon ($j = 1, 2, \dots, 20$).

Ha \underline{X} valószínűségi vektorváltozó, akkor az X_j valószínűségi változó eloszlását az \underline{X} **i . peremeloszlásának** nevezzük.

Az \underline{X} valószínűségi vektorváltozó **diszkrét**, ha értékészlete véges vagy megszámlálhatóan végtelen.

Együttes eloszlásfüggvény

Az $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó **együttes eloszlásfüggvénye** az $F : \mathbb{R}^n \rightarrow [0, 1]$ függvény, melyre

$$F(\underline{t}) = F(t_1, \dots, t_n) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n),$$

ha $(t_1, \dots, t_n) \in \mathbb{R}^n$ valós számok. Például:

- egy véletlenszerűen választott embert megkérdezzük a havi jövedelméről (X_1), a havi kiadásairól (X_2), és az életkoráról (X_3);
- ekkor (X_1, X_2, X_3) valószínűségi vektorváltozó, és
- ha eloszlásfüggvénye F , akkor például

$$F(200000, 150000, 40) = \mathbb{P}(X_1 \leq 200000, X_2 \leq 150000, X_3 \leq 40)$$

annak valószínűsége, hogy egy véletlenszerűen választott ember havi jövedelme legfeljebb 200000 (forint), havi kiadása legfeljebb 150000 (forint), életkora pedig legfeljebb 40 (év).

Valószínűségi változók függetlensége

- **két valószínűségi változóra:** az $X, Y : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\mathbb{P}(X \leq t_1, Y \leq t_2) = \mathbb{P}(X \leq t_1) \cdot \mathbb{P}(Y \leq t_2)$$

teljesül tetszőleges $t_1, t_2 \in \mathbb{R}$ valós számokra.

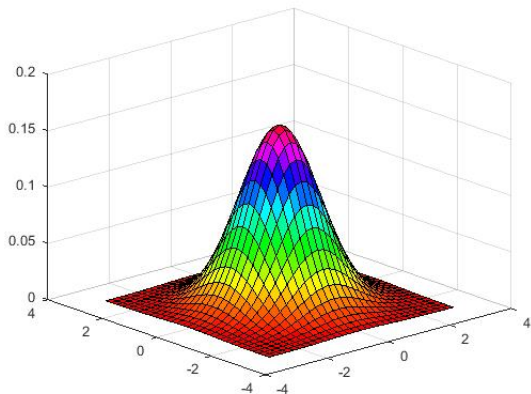
- **véges sok valószínűségi változóra:** $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) &= \\ &= \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \dots \mathbb{P}(X_n \leq t_n) \end{aligned}$$

teljesül tetszőleges t_1, t_2, \dots, t_n valós számokra.

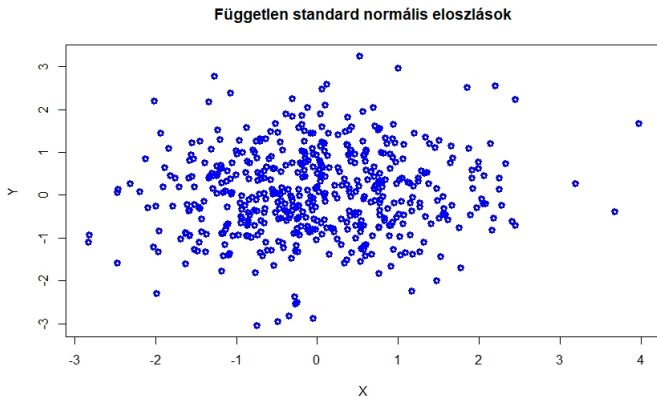
- **megszámlálható sok valószínűségi változóra:** az $X_1, X_2, X_3 \dots$ valószínűségi változók **függetlenek**, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.

Kétdimenziós normális eloszlás



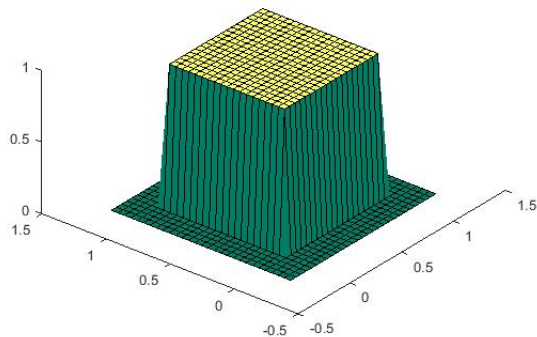
Két független standard normális eloszlás együttes sűrűségfüggvénye. Azaz: (X, Y) együttes sűrűségfüggvénye, ahol X, Y függetlenek, $N(0, 1)$ eloszlásúak.

Kétdimenziós normális eloszlás



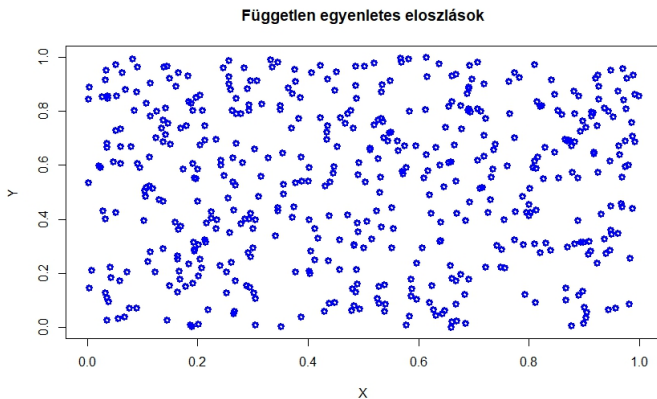
500 darab véletlen pont a síkon, melyek koordinátái független standard normális eloszlásúak. Ahol nagyobb az együttes sűrűségfüggvény (előző ábra), oda több pont esik.

Kétdimenziós egyenletes eloszlás



(X, Y) együttes sűrűségfüggvénye, ahol X és Y függetlenek és a $[0, 1]$ intervallumon egyenletes eloszlásúak

Kétdimenziós egyenletes eloszlás



500 darab véletlen pont a síkon, melyek koordinátái függetlenek és a $[0, 1]$ intervallumon egyenletes eloszlásúak (az együttes sűrűségfüggvény az előző ábrán látható).

Együttes eloszlás: példa

Kétszer dobunk szabályos kockával. Legyen X az első dobás, Y pedig a dobott számok közül a nagyobb. Ekkor az (X, Y) valószínűségi vektorváltozó együttes eloszlása: a lehetséges (k, l) párok, és a $\mathbb{P}(X = k, Y = l)$ valószínűségek, melyeket az alábbi táblázatban láthatunk:

X/Y	1	2	3	4	5	6	összesen
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	0	$1/18$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	0	0	$1/12$	$1/36$	$1/36$	$1/36$	$1/6$
4	0	0	0	$1/9$	$1/36$	$1/36$	$1/6$
5	0	0	0	0	$5/36$	$1/36$	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
összesen	$1/36$	$1/12$	$5/36$	$7/36$	$1/4$	$11/36$	1

Konvolúció

Állítás

Legyenek X és Y független, nemnegatív egész értékű valószínűségi változók. Ekkor az $X + Y$ valószínűségi változó eloszlását az alábbi módon határozhatjuk meg:

$$\mathbb{P}(X + Y = k) = \sum_{l=0}^k \mathbb{P}(X = l)\mathbb{P}(Y = k - l) \quad (k \geq 0).$$

Továbbá

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y); \quad D(X + Y) = \sqrt{D^2(X) + D^2(Y)}.$$

Bizonyítás. Diszjunkt eseményekre való szétbontással, illetve a függetlenség definíciójának felhasználásával:

$$\mathbb{P}(X + Y = k) = \sum_{l=0}^k \mathbb{P}(X = l, Y = k - l) = \sum_{l=0}^k \mathbb{P}(X = l)\mathbb{P}(Y = k - l).$$


Konvolúció: példa

Állítás

Legyenek X és Y független Poisson-eloszlású valószínűségi változók, X paramétere λ , az Y paramétere μ . Ekkor az $X + Y$ valószínűségi változó is Poisson-eloszlású, paramétere $\lambda + \mu$, várható értéke és szórásnégyzete is $\lambda + \mu$.

Bizonyítás. Legyen $k \geq 0$ tetszőleges. Ekkor a Poisson-eloszlás definíciója alapján

$$\begin{aligned}\mathbb{P}(X + Y = k) &= \sum_{l=0}^k \mathbb{P}(X = l)\mathbb{P}(Y = k - l) = \sum_{l=0}^k \frac{\lambda^l}{l!} e^{-\lambda} \cdot \frac{\mu^{k-l}}{(k-l)!} e^{-\mu} = \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^k \frac{k!}{l!(k-l)!} \lambda^l \mu^{k-l} = \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} \lambda^l \mu^{k-l} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!},\end{aligned}$$

ahol az utolsó lépésben a binomiális tételt használtuk. 

Nevezetes eloszlások összege

- X, Y független Poisson-eloszlásúak λ_1 és λ_2 paraméterrel $\Rightarrow X + Y$ Poisson-eloszlású $\lambda_1 + \lambda_2$ paraméterrel;
- X, Y független binomiális eloszlásúak, n_1 , illetve n_2 renddel, és azonos p paraméterrel

Nevezetes eloszlások összege

- X, Y független Poisson-eloszlásúak λ_1 és λ_2 paraméterrel $\Rightarrow X + Y$ Poisson-eloszlású $\lambda_1 + \lambda_2$ paraméterrel;
- X, Y független binomiális eloszlásúak, n_1 , illetve n_2 renddel, és azonos p paraméterrel $\Rightarrow X + Y$ binomiális eloszlású $n_1 + n_2$ renddel és p paraméterrel;
- X_1, X_2, \dots, X_n független normális eloszlásúak \Rightarrow az összegük és az átlaguk is normális eloszlású;

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, vagy: X_j és X_1 eloszlásfüggvénye megegyezik tetszőleges j -re.

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

Valószínűségi változók függetlensége

- **két valószínűségi változóra:** az $X, Y : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\mathbb{P}(X \leq t_1, Y \leq t_2) = \mathbb{P}(X \leq t_1) \cdot \mathbb{P}(Y \leq t_2)$$

teljesül tetszőleges $t_1, t_2 \in \mathbb{R}$ valós számokra.

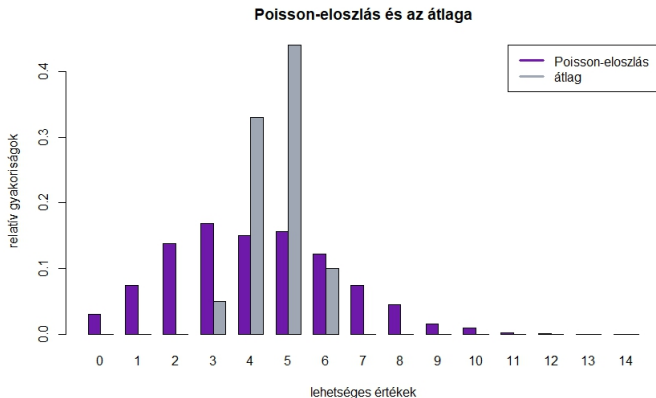
- **véges sok valószínűségi változóra:** $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ valószínűségi változók **függetlenek**, ha

$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) &= \\ &= \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \dots \mathbb{P}(X_n \leq t_n) \end{aligned}$$

teljesül tetszőleges t_1, t_2, \dots, t_n valós számokra.

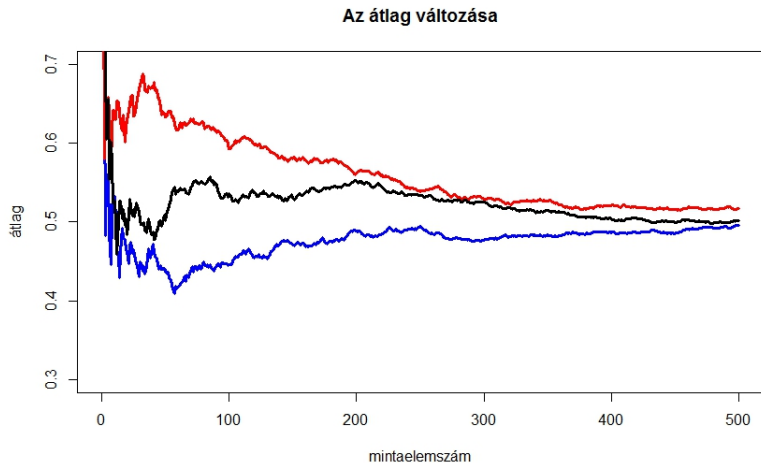
- **megszámlálható sok valószínűségi változóra:** az $X_1, X_2, X_3 \dots$ valószínűségi változók **függetlenek**, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.

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



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

Az átlag várható értéke

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $m = \mathbb{E}(X_1) < \infty$. Ekkor

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

Az átlag várható értéke

Állítás

Legyenek X_1, \dots, X_n független azonos eloszlású valószínűségi változók, melyekre $m = \mathbb{E}(X_1) < \infty$. Ekkor

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

Bizonyítás.

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

Felhasználtuk a várható érték linearitását, és hogy csak 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 (azaz eloszlásfüggvényük) megegyezik, akkor $\mathbb{E}(Y) = \mathbb{E}(Z)$

Az átlag szórása

Állítás

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

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

Az átlag szórása

Állítás

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

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

Bizonyítás.

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

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 $D(Y) = D(Z)$

Egyenlőtlenségek

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra – azonban p -t **nem ismerjük**.

Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya **legfeljebb 1%-kal** tér el p -től, **tetszőleges p esetén legalább 95% legyen?**

Egyenlőtlenségek

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra – azonban p -t **nem ismerjük**.

Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya **legfeljebb 1%-kal** tér el p -tól, **tetszőleges p esetén legalább 95% legyen?**

Ennek megértéséhez ezekre van szükség:

- az arány átlagolás \rightarrow milyen gyorsan csökken a szórás?
- ha a **szórás kicsi**, abból hogyan következik, hogy nagy valószínűséggel csak keveset tévedünk \rightarrow ebben segít a **Markov– és a Csebisev-egyenlőtlenség**.

A Markov-egyenlőtlenség bizonyítása

Markov-egyenlőtlenség. Legyen $t > 0$, és X **nemnegatív, véges várható értékű** valószínűségi változó, vagyis melyre $X \geq 0$ biztosan teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Bizonyítás. Tekintsük az alábbi Y valószínűségi változót:

$$Y = \begin{cases} t, & \text{ha } X \geq t; \\ 0, & \text{ha } X < t. \end{cases}$$

A Markov-egyenlőtlenség bizonyítása

Markov-egyenlőtlenség. Legyen $t > 0$, és X **nemnegatív, véges várható értékű** valószínűségi változó, vagyis melyre $X \geq 0$ biztosan teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Bizonyítás. Tekintsük az alábbi Y valószínűségi változót:

$$Y = \begin{cases} t, & \text{ha } X \geq t; \\ 0, & \text{ha } X < t. \end{cases}$$

Mindkét esetben Y értéke legfeljebb X értéke (hiszen $X \geq 0$):

$$Y \leq X \quad \Rightarrow \quad \mathbb{E}(Y) \leq \mathbb{E}(X).$$

Másrészt, mivel Y értéke összesen kétféle lehet, a várható értékét kiszámítva:

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + t \cdot \mathbb{P}(Y = t) = t \cdot \mathbb{P}(X \geq t) \leq \mathbb{E}(X).$$

A Markov-egyenlőtlenség bizonyítása

Markov-egyenlőtlenség. Legyen $t > 0$, és X **nemnegatív, véges várható értékű** valószínűségi változó, vagyis melyre $X \geq 0$ biztosan teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Bizonyítás. Tekintsük az alábbi Y valószínűségi változót:

$$Y = \begin{cases} t, & \text{ha } X \geq t; \\ 0, & \text{ha } X < t. \end{cases}$$

Mindkét esetben Y értéke legfeljebb X értéke (hiszen $X \geq 0$):

$$Y \leq X \quad \Rightarrow \quad \mathbb{E}(Y) \leq \mathbb{E}(X).$$

Másrészt, mivel Y értéke összesen kétféle lehet, a várható értékét kiszámítva:

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + t \cdot \mathbb{P}(Y = t) = t \cdot \mathbb{P}(X \geq t) \leq \mathbb{E}(X).$$

Az utolsó egyenlőtlenség mindkét oldalát a t pozitív számmal osztva a Markov-egyenlőtlenséget kapjuk.

A Csebisev-egyenlőtlenség bizonyítása

Markov-egyenlőtlenség. Legyen $t > 0$, és X **nemnegatív, véges várható értékű** valószínűségi változó, vagyis melyre $X \geq 0$ biztosan teljesül. Ekkor

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Csebisev-egyenlőtlenség. Legyen X **véges szórású** valószínűségi változó, $t > 0$ pozitív szám. Ekkor

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{D^2(X)}{t^2}.$$

Bizonyítás. Legyen $Z = (X - \mathbb{E}(X))^2$. Ez a valószínűségi változó nemnegatív, ezért alkalmazható a Markov-egyenlőtlenség, a t^2 pozitív számmal:

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq t) &= \mathbb{P}((X - \mathbb{E}(X))^2 \geq t^2) = \mathbb{P}(Z \geq t^2) \leq \\ &\stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}(Z)}{t^2} = \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{t^2} = \frac{D^2(X)}{t^2} \end{aligned}$$

a szórásnégyzet definíciója alapján.

A Csebisev-egyenlőtlenség alkalmazása

Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya **legfeljebb 1%-kal** 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 – hiszen p -t nem ismerjük.

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

Mivel X binomiális eloszlású:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \quad D\left(\frac{X}{n}\right) = \frac{1}{n} \sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

Mivel X binomiális eloszlású:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \quad D\left(\frac{X}{n}\right) = \frac{1}{n} \sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

Csebisev-egyenlőtlenség alkalmazva az X/n valószínűségi változóra:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{D^2\left(\frac{X}{n}\right)}{0,01^2} = \frac{p(1-p)}{0,01^2 \cdot n} \leq \frac{1}{4 \cdot 0,01^2 \cdot n},$$

mivel $p(1-p) \leq 1/4$ a számtani-mértani közepek közötti egyenlőtlenség szerint.

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

A Csebisev-egyenlőtlenség szerint

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{1}{4 \cdot 0,01^2 \cdot n}$$

Kell:

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

azaz

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0,01\right) \leq 0,05.$$

Tehát ennyi embert **biztosan elég megkérdezni**:

$$\frac{1}{4 \cdot 0,01^2 \cdot n} \leq 0,05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000.$$

A Csebisev-egyenlőtlenség alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X : a pártot támogatók száma a megkérdezettek között

A Csebisev-egyenlőtlenség szerint

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0,01\right) \leq \frac{1}{4 \cdot 0,01^2 \cdot n}$$

Kell:

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

azaz

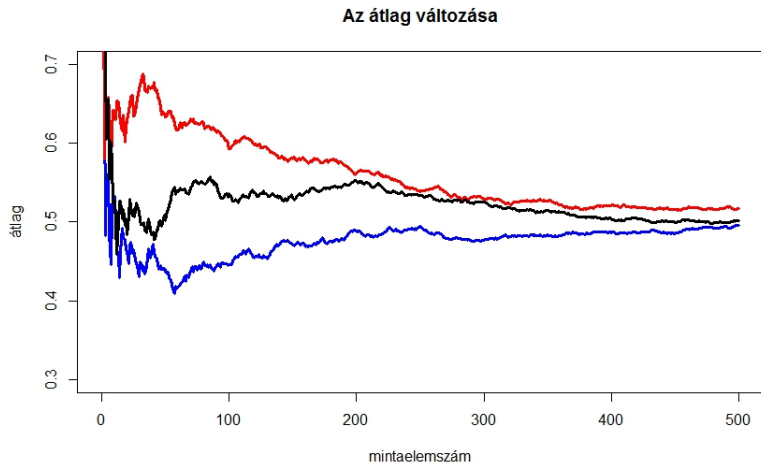
$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0,01\right) \leq 0,05.$$

Tehát ennyi embert **biztosan elég megkérdezni**:

$$\frac{1}{4 \cdot 0,01^2 \cdot n} \leq 0,05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000.$$

Ha 0,01 helyett 0,005-öt írnánk (a felét), $n \geq 200000$ (négyyszer annyi) adódna.

Az átlag konvergenciája



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

Házi feladat november 12., kedd, 8:15-ig

Legyenek X és Y egymástól független Poisson-eloszlású valószínűségi változók, melyek várható értéke 5.

a) Határozzuk meg $(X, X + Y)$ együttes eloszlásfüggvényének értékét az $s = 1, t = 2$ helyen, azaz $F(1, 2)$ -t, ahol F az együttes eloszlásfüggvény.

b) Sorsoljunk két független, 1000 elemű mintát 5 várható értékű Poisson-eloszlásból, és számítsuk ki, hogy az 1000 sorsolás hányadrészében lesz az első mintaelem legfeljebb 1, az összeg pedig legfeljebb 2.