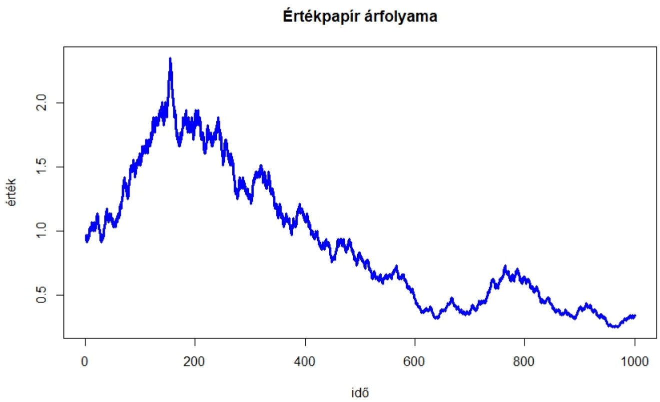
Valószínűségszámítás előadás, 7. hét, 2021. október 22.

Eloszlásfüggvény, együttes eloszlás

1. Eloszlásfüggvény



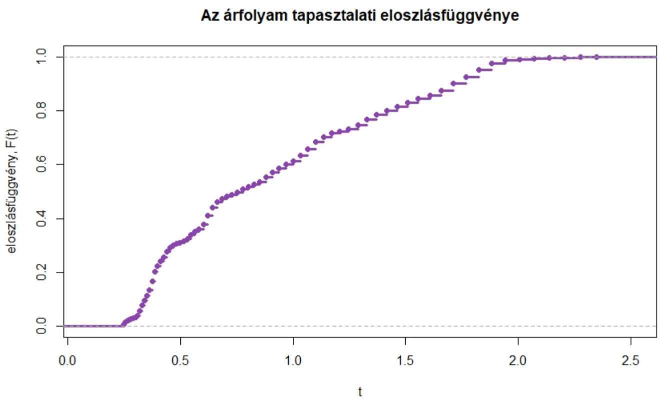
1. ábra. Egy elképzelt értékpapír árfolyama 1000 napon keresztül, 1000 forintban

- X valószínűségi változó: egy véletlen kísérlet eredménye
- eddig: X **diszkrét**, és a $\mathbb{P}(\mathbf{X} = \mathbf{x})$ valószínűségekkel lehet leírni az eloszlását
- ha a lehetséges értékek halmaza „túl nagy” (1. ábra), 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 $\to$ ennél hasznosabb információ lehet, hogy

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

  azaz az értékpapír 50% valószínűséggel nem haladja meg a 785 szintet
- **eloszlásfüggvény**: $F(t)$ annak valószínűsége, hogy **a valószínűségi változó értéke legfeljebb** t, azaz

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



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

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

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

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 $\{X \leq t\} = \{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}$ egy esemény, tehát van valószínűsége.

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

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

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

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

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

$$F(0) = \mathbb{P}(X \leq 0) = \frac{\mathbf{1}}{\mathbf{8}}; \qquad F(1) = \mathbb{P}(X \leq 1) = \frac{\mathbf{1}}{\mathbf{2}};$$

$$F(2,4) = \mathbb{P}(X \leq 2,4) = \frac{\mathbf{7}}{\mathbf{8}}; \qquad F(4) = \mathbb{P}(X \leq 4) = \mathbf{1}.$$

3. ábra. Három gyerek közül a fiúk számának eloszlásfüggvénye; a függőleges tengelyen: $F(t) = \mathbb{P}(X \leq t)$.

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

1.1. Az eloszlásfüggvény tulajdonságai

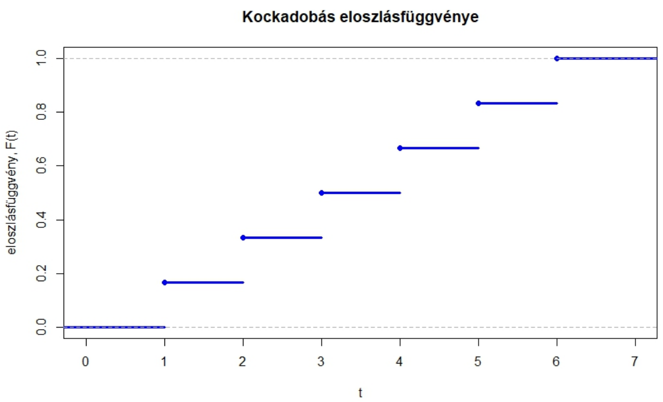
1.1. Állítás. *Ha* $a, b \in \mathbb{R}$ *valós számok, és* F *az* X *eloszlásfüggvénye, akkor*

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

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

(i) F monoton növő: $a < b$ esetén $F(a) \leq F(b)$.



4. ábra. Szabályos dobókockával dobott szám eloszlásfüggvénye; a függőleges tengelyen: $F(t) = \mathbb{P}(X \leq t)$.

(ii) $\lim_{t\to-\infty} F(t) = 0$; $\quad \lim_{t\to\infty} F(t) = 1$.

(iii) F jobbról folytonos, azaz minden $t \in \mathbb{R}$ valós számra $\lim_{s\to t+} F(s) = F(t)$.

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

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

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

2. Nevezetes eloszlások és eloszlásfüggvényeik

Nem csak a diszkrét valószínűségi változók között, hanem általában is találhatunk olyan eloszlásokat, melyek gyakran előfordulnak hétköznapi példákban vagy matematikai modellekben. Ezeket most az eloszlásfüggvényeik segítségével mutatjuk be.

2.1. Egyenletes eloszlás

Tekintsük a következő hétköznapi példát.

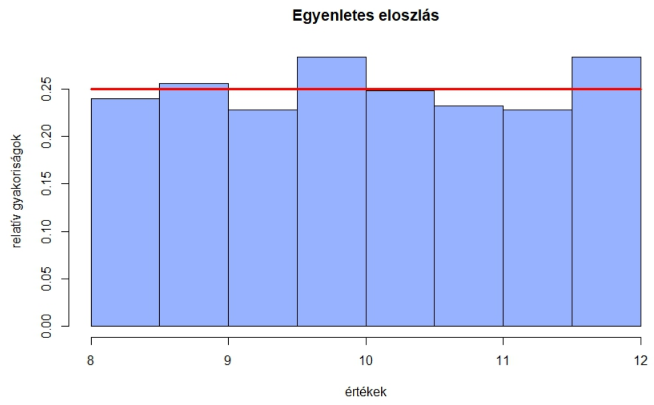
- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- Mennyi a valószínűsége, hogy a futár 11 óráig megérkezik?
- **Feltéve, hogy a futár 10 óráig még nem érkezett meg**, **mennyi a valószínűsége, hogy 11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

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

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

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

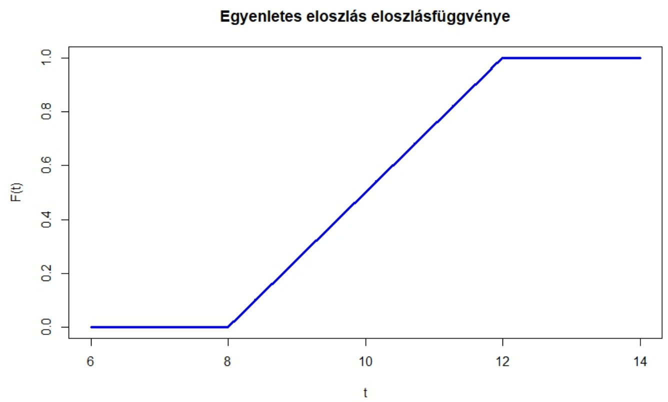
Általában azt vehetjük észre, hogy annak valószínűsége, hogy X egy adott intervallumba esik, a $(8, 12)$ intervallumon belül, az intervallum hosszával arányos.



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

2.1. Definíció (Egyenletes eloszlás (uniform distribution)). *Az X valószínűségi változó* **egyenletes eloszlású** *az $[a, b]$ intervallumon, ha eloszlásfüggvénye:*

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



6. ábra. A $[8, 12]$ intervallumon egyenletes eloszlás eloszlásfüggvénye

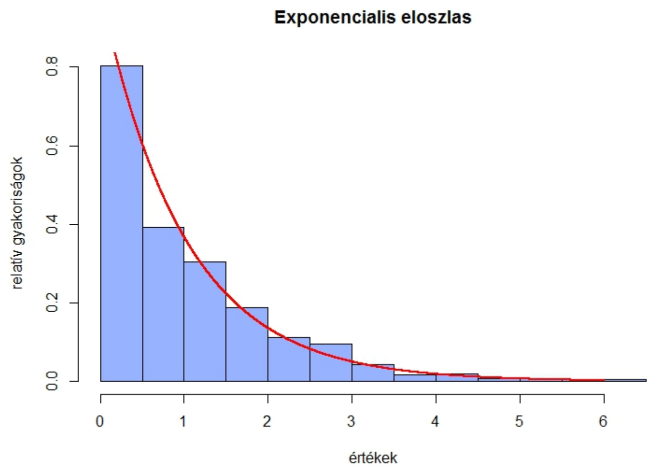
Példa. Csomagot várunk, a futár 10 és 12 óra között érkezik. Feltesszük, hogy érkezésének időpontja egyenletes eloszlású a $[10, 12]$ intervallumon. Ekkor az előző állítás alapján az alábbiak igazak ($a = 10, b = 12$).

- 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: $3/4 = 0,75$.

2.2. 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



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

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

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

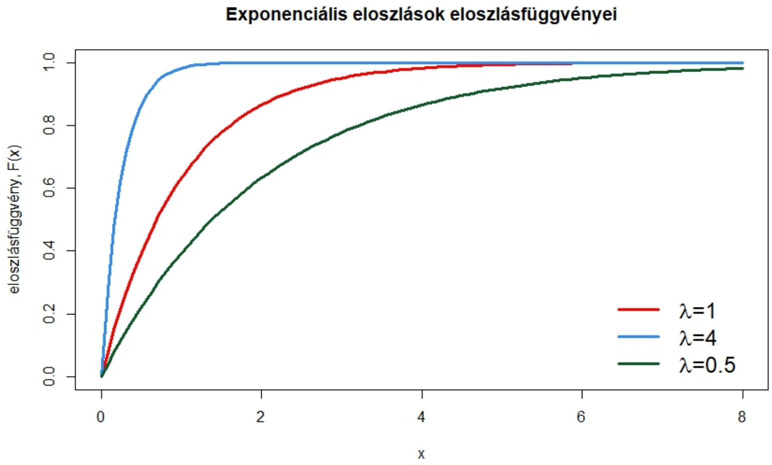
2.1. Állítás (Az exponenciális eloszlás örökifjú tulajdonsága). *Legyen X exponenciális eloszlású valószínűségi változó, s, t pozitív számok. Ekkor*

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

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

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

□



7. ábra. Különböző paraméterű $\left(\lambda = \frac{1}{2}, 1, \text{ illetve } 4\right)$ exponenciális eloszlások eloszlásfüggvényei

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

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

Hasonlóképpen

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

2.3. Pareto-eloszlás

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

2.3. Definíció (Pareto-eloszlás). *Az X valószínűségi változó Pareto-eloszlású $\alpha > 0$ és $c > 0$ paraméterekkel, ha eloszlásfüggvénye*

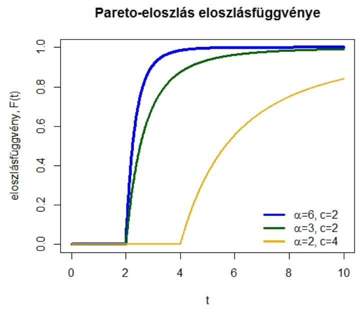
$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0; & ha\ t \leq c; \\ 1 - \left(\frac{c}{t}\right)^{\alpha}; & ha\ t > c. \end{cases}$$

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.

3. Kvantilisek

3.1. Definíció (Kvantilis). *Legyen $0 \leq z \leq 1$, és X egy valószínűségi változó. Ekkor az X valószínűségi változó z-kvantilise:*

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



8. ábra. Különböző paraméterű Pareto-eloszlások eloszlásfüggvényei

Ha az X eloszlásfüggvénye folytonos, akkor az igaz, hogy

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

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.

4. Valószínűségi vektorváltozó

Gyakran több mennyiséget egyszerre vizsgálunk, vagy ugyanazt többször megmérjük. Például: véletlenszerűen választott ember havi jövedelme, kiadása rezsire, illetve élelmiszerre; vagy egy helyen öt különböző műszerrel megmérjük a hőmérsékletet.

4.1. Definíció (Valószínűségi vektorváltozó). *Egy $\underline{X} : \Omega \to \mathbb{R}^n$ függvény valószínűségi vektorváltozó, ha minden koordinátája valószínűségi változó, azaz $\underline{X} = (X_1, X_2, \ldots, X_n)$, ahol $X_1, \ldots, X_n$ valószínűségi változók.*

Belátható, hogy ilyenkor tetszőleges $t_1, t_2, \ldots, t_n$ számokra

$$\{\omega \in \Omega : X_1(\omega) \leq t_1, X_2(\omega) \leq t_2, \ldots, X_n(\omega) \leq t_n\} \in \mathcal{A},$$

azaz a $\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n)$ valószínűség értelmes.

Itt $\underline{X}$ értékei n dimenziós vektorok, vagyis n hosszú számsorozatok, és X_k ennek a k. koordinátáját (elemét) jelöli.

Valószínűségi vektorváltozó együttes eloszlása: $\underline{x}_k$ lehetséges értékek és $p_k = \mathbb{P}(X = \underline{x}_k)$ valószínűségek $(k = 1, 2, \ldots)$.

Az $\underline{X} = (X_1, X_2, \ldots, X_n)$ függvény pontosan akkor valószínűségi vektorváltozó, ha minden koordinátája valószínűségi változó. Ez alapján:

4.2. Definíció. *Legyen $\underline{X} = (X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó. Ekkor az X_k valószínűségi változó eloszlását az $\underline{X}$ valószínűségi vektorváltozó k. peremeloszlásának nevezzük.*

Az együttes eloszlás meghatározza a peremeloszlásokat. Fordítva nem, kivéve, ha a koordináták függetlenek.

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:

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

Például, mivel X az első dobás, Y pedig a két dobás közül a nagyobb:

$$\mathbb{P}(X=3, Y=3) = \mathbb{P}(\text{az első dobás 3, a második 1, 2 vagy 3}) = \frac{1}{6}\cdot\frac{1}{2} = \frac{1}{12},$$

hiszen a két dobás egymástól független.

Vagy:

$$\mathbb{P}(X=2, Y=5) = \mathbb{P}(\text{az első dobás 3, a második 5}) = \frac{1}{36}.$$

A peremeloszlások:

X: $(1,1/6),(2,1/6),(3,1/6),\dots,(6,1/6)$

Y: $(1,1/36),(2,1/18),\dots,(6,1/6)$

4.1. Együttes eloszlásfüggvény

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

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

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

Például:

- egy véletlenszerűen választott embert megkérdezünk 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

  $$F(200000, 150000, 40) = \mathbb{P}(X_1 \le 200000, X_2 \le 1500000, X_3 \le 40)$$

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

Az $X_1,\dots,X_n$ függetlenségét úgy is megfogalmazhatjuk, hogy minden $t_1,t_2,\dots,t_n$-re

$$F(t_1,\dots,t_n) = F_1(t_1)F_2(t_2)\dots F_n(t_n),$$

ahol F az együttes sűrűségfüggvény, F_j pedig az X_j eloszlásfüggvénye ($j = 1,2,\dots,n$). Ugyanis ez pont azt jelenti, hogy minden $t_1,t_2,\dots,t_n$-re

$$\mathbb{P}(X_1 \le t_1, X_2 \le t_2, \dots, X_n \le t_n) = \mathbb{P}(X_1 \le t_1)\cdot\ldots\cdot\mathbb{P}(X_n \le t_n),$$

és éppen ez volt a függetlenség definíciója. Másképpen: a függetlenség azzal ekvivalens, hogy az együttes eloszlásfüggvény a peremeloszlásfüggvények szorzata.

Házi feladat november 5., péntek, 12:15-ig Egy országban a diplomások havi jövedelme $\alpha = 3$ és $c = 15000$ paraméterű Pareto-eloszlással jellemezhető, míg a diplomával nem rendelkezők jövedelme $\alpha = 4$ és $c = 10000$ paraméterű Pareto-eloszlással.

Megkérdezünk egy embert, aki elárulja, hogy a jövedelme 18000 és 20000 között van. Ez alapján mennyi a feltételes valószínűsége, hogy az illetőnek van diplomája?