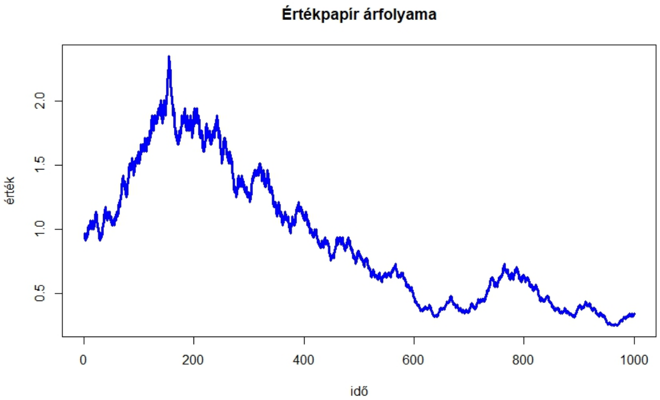
Eloszlásfüggvény: példa (7. előadás)



Egy elképzelt értékpapír árfolyama 1000 napon keresztül, 1000 forintban

Eloszlásfüggvény: bevezetés

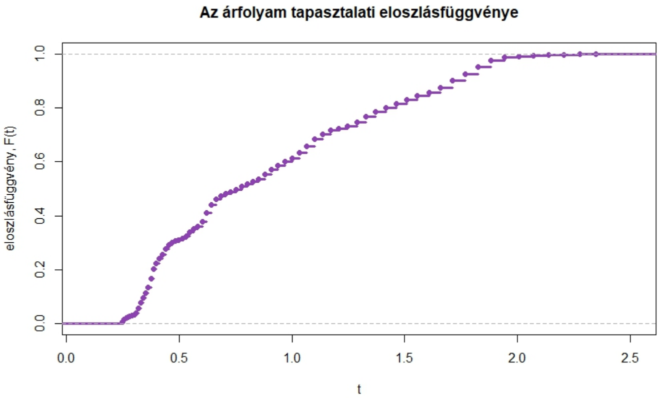
- X valószínűségi változó: egy véletlen kísérlet eredménye
- eddig: X **diszkrét**, és a $\mathbb{P}(X = x)$ valószínűségekkel lehet leírni az eloszlását
- ha a lehetséges értékek halmaza „túl nagy”, vagy a valószínűségek „túl kicsik”, ez nem informatív
- például: X az értékpapír árfolyama holnap, $\mathbb{P}(X = 784) = 0,0038$, $\mathbb{P}(X = 785) = 0,004$, stb. egy előrejelzés szerint $\to$ 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 $X : \Omega \to \mathbb{R}$ valószínűségi változó. Ekkor X **eloszlásfüggvénye** az alábbi $F : \mathbb{R} \to [0, 1]$ 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 $\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}$ egy esemény, tehát van valószínűsége.

Eloszlásfüggvény: példa

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

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

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

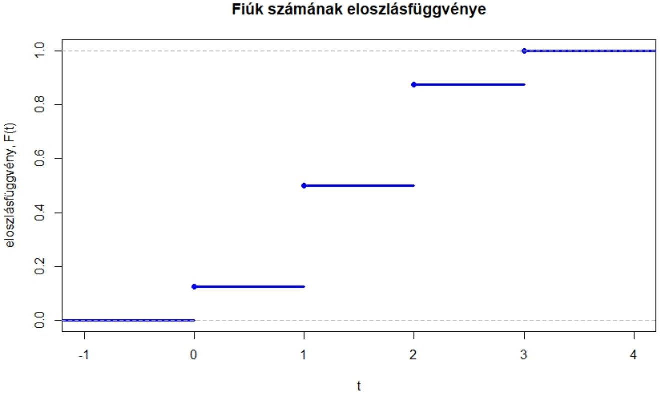
$$\mathbb{P}(X=0)=\frac{1}{8}, \quad \mathbb{P}(X=1)=\frac{3}{8}, \quad \mathbb{P}(X=2)=\frac{3}{8}, \quad \mathbb{P}(X=3)=\frac{1}{8}.$$

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

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

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

Eloszlásfüggvény: példa



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

Eloszlásfüggvény

Definíció (Eloszlásfüggvény)

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

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

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:

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

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

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

Eloszlásfüggvény

Definíció (Eloszlásfüggvény)

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

$$F(t) = \mathbb{P}(X \le t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \le t\}) \qquad \textit{minden } t \in \mathbb{R} \textit{ valós számra.}$$

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:

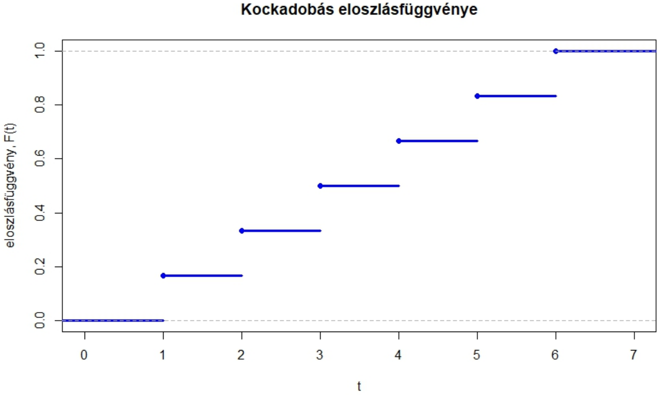
$$F(1) = \mathbb{P}(X \le 1) = \mathbb{P}(\text{legfeljebb egy fiú van}) = 1/2;$$

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

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

Véges értékkészletű 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

1. (i) F monoton növő: $a < b$ esetén $F(a) \leq F(b)$.
2. (ii) $\lim_{t\to-\infty} F(t) = 0$; $\quad \lim_{t\to\infty} F(t) = 1$.
3. (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.

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

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

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

Először a kovarianciát határozzuk meg, felhasználva, hogy a kovariancia mind a két változójában lineáris, valamint független valószínűségi változók kovarianciája 0.

$$\begin{aligned}\operatorname{cov}(X - Y, X + 2Y) &= \operatorname{cov}(X, X) + 2\operatorname{cov}(X, Y) - \operatorname{cov}(Y, X) - 2\operatorname{cov}(Y, Y)\\ &= D^2(X) - 2D^2(Y) = 30 - 2 \cdot 40 = -50,\end{aligned}$$

hiszen Poisson-eloszlású valószínűségi változóknál a várható érték és a szórásnégyzet megegyezik.

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

A szórásnégyzet tulajdonságai alapján, felhasználva, hogy X és Y függetlenek:

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

A szórásnégyzet tulajdonságai alapján, felhasználva, hogy X és Y függetlenek:

$$D^2(X - Y) = D^2(X) + (-1)^2 D^2(Y) = 30 + 40 = 70;$$

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

A szórásnégyzet tulajdonságai alapján, felhasználva, hogy X és Y függetlenek:

$$D^2(X - Y) = D^2(X) + (-1)^2 D^2(Y) = 30 + 40 = 70;$$

$$D^2(X + 2Y) = D^2(X) + 2^2 D^2(Y) = 30 + 4 \cdot 40 = 190;$$

Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

A szórásnégyzet tulajdonságai alapján, felhasználva, hogy X és Y függetlenek:

$$D^2(X - Y) = D^2(X) + (-1)^2 D^2(Y) = 30 + 40 = 70;$$

$$D^2(X + 2Y) = D^2(X) + 2^2 D^2(Y) = 30 + 4 \cdot 40 = 190;$$

$$R(X - Y, X + 2Y) = \frac{\operatorname{cov}(X - Y, X + 2Y)}{D(X - Y)D(X + 2Y)} = \frac{-50}{\sqrt{70} \cdot \sqrt{190}} = -0,43.$$

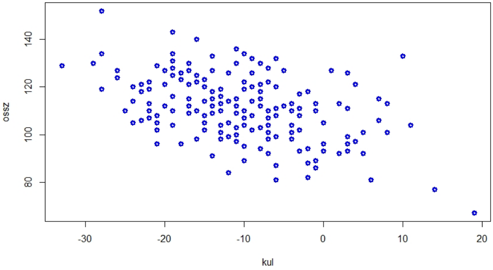
Házi feladat november 6., hétfő, 10:15-ig

Egy boltban a vásárlók száma délelőtt Poisson-eloszlású 30 várható értékkel, délután szintén Poisson-eloszlású 40 várható értékkel, és a délelőtti és délutáni vásárlók száma egymástól független.

Határozzuk meg $X - Y$ és $X + 2Y$ korrelációs együtthatóját.

```
x=rpois(200, lambda=30)

y=rpois(200, lambda=40)

kul =x-y

ossz = x+2*y

plot(ossz kul, lwd="3", col="blue")

cov(kul, ossz)

[1] -54.23739

cor(kul, ossz)

[1] -0.45
```

Házi feladat november 6., hétfő, 10:15-ig



Az együttes eloszlás 200 kísérletből

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.

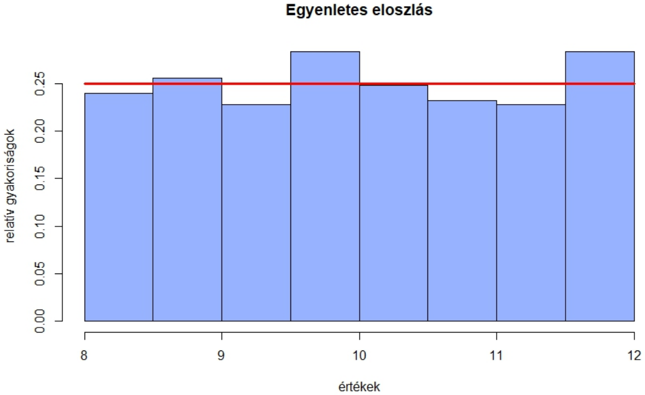
- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- Mennyi a valószínűsége, hogy a futár 11 óráig megérkezik?
- **Feltéve, hogy a futár 10 óráig még nem érkezett meg**, **mennyi a valószínűsége, hogy 11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



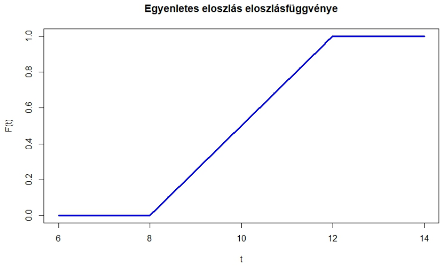
A $[8, 12]$ intervallumon egyenletes eloszlásból vett 500 elemű minta hisztogramja

Egyenletes eloszlás

Definíció (Egyenletes eloszlás (uniform distribution))

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

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



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 ($a = 10$, $b = 12$).

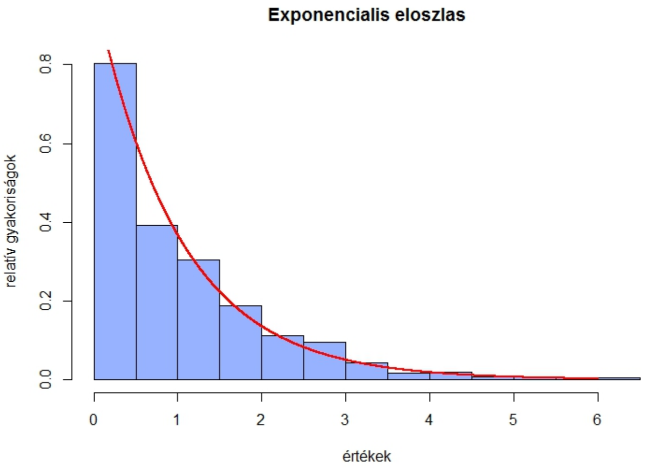
- Annak valószínűsége, hogy 10 és 11 óra között érkezik: $(11 - 10)/(12 - 10) = 1/2$.
- Annak valószínűsége, hogy 10:15 és 10:30 között érkezik: $1/8 = 0,125$.
- Annak valószínűsége, hogy $10 : 30$ után érkezik: $3/4 = 0,75$.

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

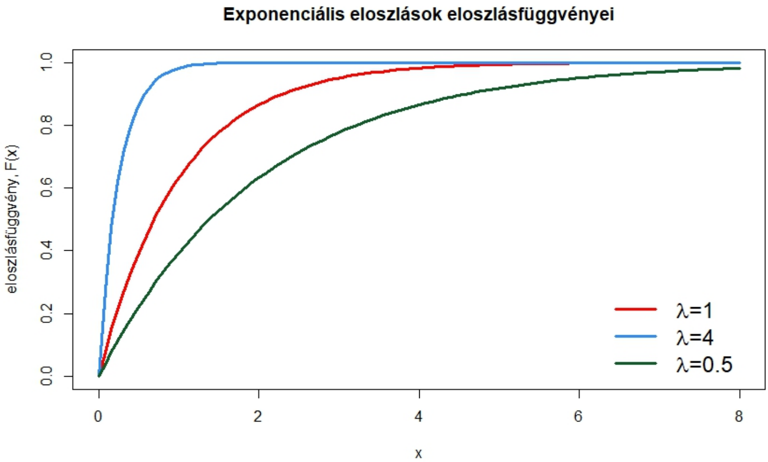
Exponenciális eloszlás

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 $\left(\lambda = \frac{1}{2}, 1,\ \text{illetve } 4\right)$ 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

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

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

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

Hasonlóképpen

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

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

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

$\square$

Pareto-eloszlás

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

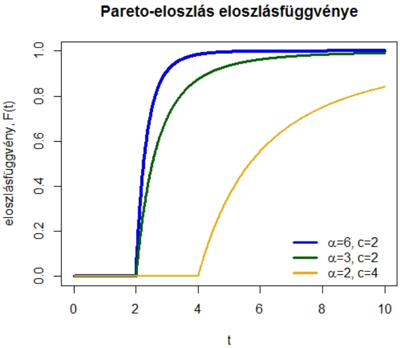
Definíció (Pareto-eloszlás)

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

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

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.

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

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

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

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

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

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

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

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

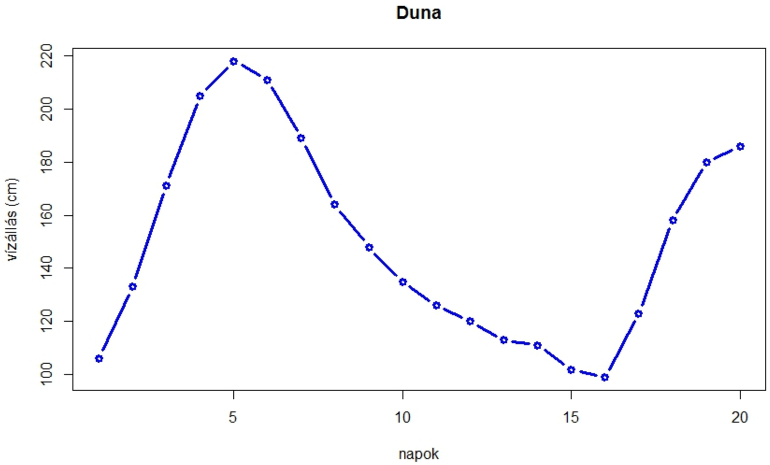
azaz z az a szám, aminél X éppen z valószínűséggel kisebb.

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.

Valószínűségi vektorváltozó: példa



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

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 méréssorozatban a különböző mérések során megfigyelt értékek (például egy mérést tízszer megismételve tíz különböző valószínűségi változót kapunk).

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

Amint látni fogjuk, az eloszlásfüggvény, sűrűségfüggvény az együttes esetben is definiálható.

Valószínűségi vektorváltozó

Az

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

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

- 1000 embert megkérdezünk 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 vízállása a j. napon $(j = 1, 2, \dots, 20)$.

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

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

Együttes eloszlásfüggvény

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

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

ha $(t_1, \ldots, t_n) \in \mathbb{R}^n$ 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

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

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 $X, Y : \Omega \to \mathbb{R}$ 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.

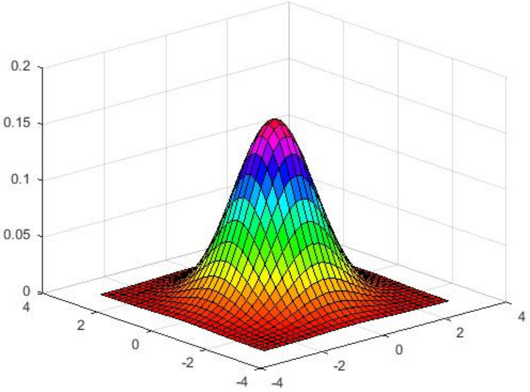
- **véges sok valószínűségi változóra**: $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ valószínűségi változók **függetlenek**, ha

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

  teljesül tetszőleges $t_1, t_2, \ldots, t_n$ valós számokra.

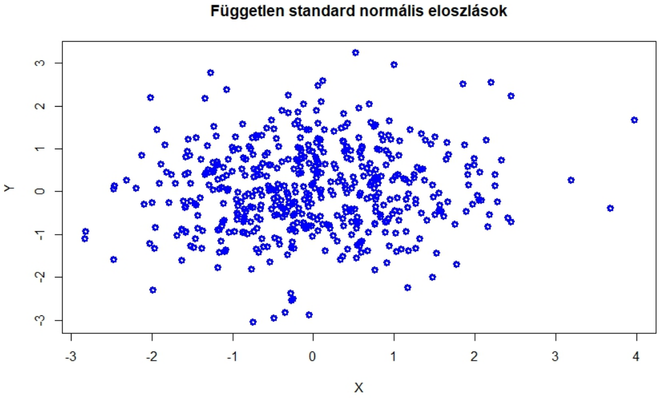
- **megszámlálható sok valószínűségi változóra**: az $X_1, X_2, X_3 \ldots$ valószínűségi változók **függetlenek**, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.

Kétdimenziós normális eloszlás



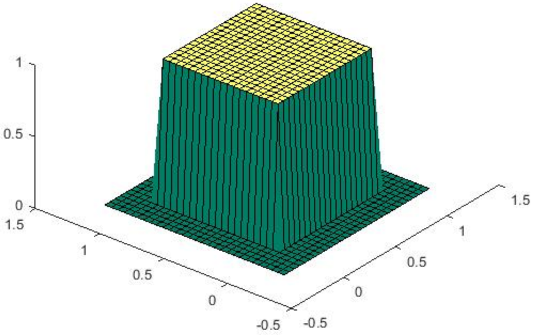
Két független standard normális eloszlás együttes sűrűségfüggvénye Azaz: (X, Y) együttes sűrűségfüggvénye, ahol X, Y függetlenek, $N(0, 1)$ eloszlásúak

Kétdimenziós normális eloszlás



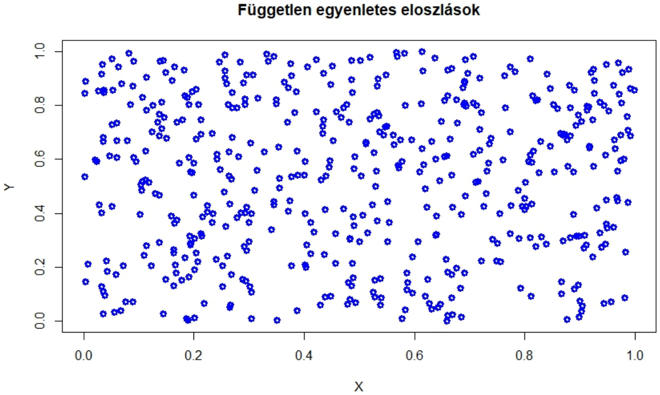
500 darab véletlen pont a síkon, melyek koordinátái független standard nomális eloszlásúak. Ahol nagyobb az együttes sűrűségfüggvény (előző ábra), oda több pont esik.

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:

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

Továbbá

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

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

$$\mathbb{P}(X+Y=k) = \sum_{l=0}^{k} \mathbb{P}(X=l)\mathbb{P}(Y=k-l) = \sum_{l=0}^{k} \frac{\lambda^l}{l!}e^{-\lambda} \cdot \frac{\mu^{k-l}}{(k-l)!}e^{-\mu} =$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^{k} \frac{k!}{l!(k-l)!} \lambda^l \mu^{k-l} =$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} \lambda^l \mu^{k-l} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

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;
- $X_1, X_2, \ldots, X_n$ függgetlen normális eloszlásúak $\Rightarrow$ az összegük és az átlaguk is normális eloszlású;

Házi feladat november 13., hétfő, 10:15-ig

Sorsoljunk az R programban $X_1, X_2, \ldots, X_{1000}$, Y_1, $Y_2, \ldots$, $Y_{1000}, Z_1, Z_2, \ldots, Z_1$ egymástól független, 1 paraméterű exponenciális eloszlású valószínűségi változókat az `rexp` parancs segítségével.

a) Ábrázoljuk az $(X_1, X_2, \ldots, X_{1000})$ minta hisztogramját és tapasztalati eloszlásfüggvényét.

b) Ábrázoljuk az $(X_1 + Y_1 + Z_1, X_2 + Y_2 + Z_2, \ldots, X_{1000} + Y_{1000} + Z_{1000})$ ezer elemű minta hisztogramját és tapasztalati eloszlásfüggvényét.

c) A fentiek közül melyik mintának nagyobb a 95%-os kvantilise? Indokoljunk is (vagy az ábra alapján, vagy a kvantilisek kiszámításával).