Valószínűségszámítás előadás, 9. hét, 2020. november 11.

Sűrűségfüggvény, egyenletes eloszlás

1. A sűrűségfüggvény

Azt már láttuk (például az egyenletes eloszlás, exponenciális eloszlás, Pareto-eloszlás esetében), hogy véletlen mennyiségek modellezésénél nem csak diszkrét (véges vagy megszámlálhatóan sok értéket felvevő) valószínűségi változókat érdemes használni, hanem olyanokat is, melyek értékkészlete egy intervallum vagy például a pozitív számok halmaza. Emlékeztetőül, az X valószínűségi változó eloszlásfüggvényét így definiáltuk:

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

ha $t \in \mathbb{R}$ tetszőleges valós szám. Ebből következik, hogy

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

is teljesül tetszőleges $a < b$ valós számokra.

Az eloszlásfüggvény azonban nem mindig a „legkényelmesebb” leírása a valószínűségi változó viselkedésének. Ezért gyakran használjuk inkább a sűrűségfüggvényt (olyankor, amikor van egyáltalán sűrűségfüggvény), melynek fő jellemzője, hogy annak valószínűségét, hogy X egy adott $[a, b]$ intervallumba esik, a sűrűségfüggvény alatti területnek az a és b közé eső része, azaz a sűrűségfüggvény a-tól b-ig vett integrálja adja meg.

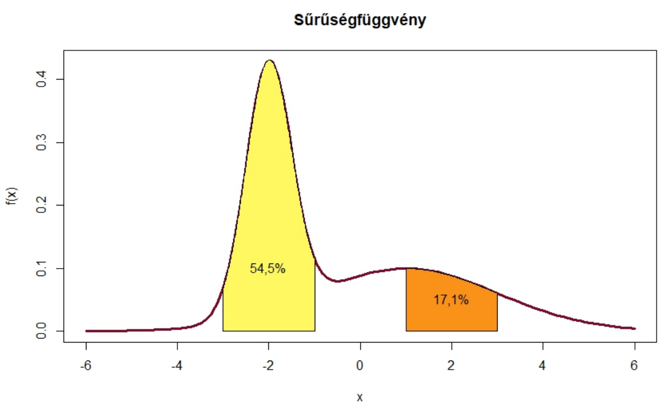
Ezt láthatjuk az 1. ábrán. Ha X sűrűségfüggvénye f (ami most az ábrán látható függvény), akkor például annak valószínűsége, hogy X értéke -3 és -1 közé esik:

$$\mathbb{P}(-3 \leq X \leq -1) = \int_{-3}^{-1} f(x)dx = \mathbf{54,5}\%.$$

Hasonlóképpen, annak valószínűsége, hogy X értéke 1 és 3 közé esik:

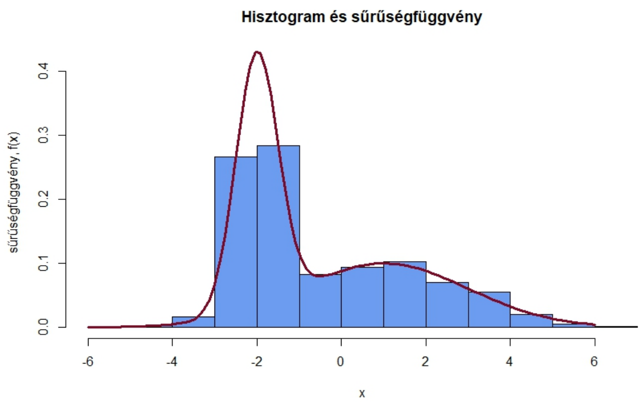
$$\mathbb{P}(1 \leq X \leq 3) = \int_{1}^{3} f(x)dx = \mathbf{17,1\%}.$$

Láthatjuk, hogy bár a két intervallum egyformán hosszú, az első esetben a sűrűségfüggvény nagyobb értékei miatt a sűrűségfüggvény alatti terület és így a valószínűség is nagyobb lesz. Vagyis minél nagyobbak egy intervallumban a sűrűségfüggvény értékei, annál nagyobb valószínűséggel esik oda a megfigyelés.



1. ábra. Példa sűrűségfüggvényre

A 2. ábrán ugyanezt a sűrűségfüggvényt használva sorsoltunk 1000 elemű mintát, vagyis tekintettünk ezer, egymástól független valószínűségi változót úgy, hogy mindegyiknek ugyanaz az ábrán szereplő függvény a sűrűségfüggvénye. Azt látjuk, hogy a hisztogram nagyrészt követi a sűrűségfüggvény alakját: ahol nagyobb a sűrűségfüggvény értéke, oda nagyobb valószínűséggel esnek a megfigyelések, így a hisztogramon magasabb oszlopokat látunk.



2. ábra. A sűrűségfüggvény és hozzá tartozó ezer elemű minta hisztogramja

1.1. Definíció. *Az $X : \Omega \to \mathbb{R}$ valószínűségi változó* **sűrűségfüggvénye** *az $f : \mathbb{R} \to \mathbb{R}$ függvény, ha*

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} f(x)dx$$

teljesül minden $t \in \mathbb{R}$ számra.

Nem minden valószínűségi változónak van sűrűségfüggvénye, például a diszkréteknek nincs.

1.2. Definíció. *Ha az X valószínűségi változónak* **van sűrűségfüggvénye**, *akkor* **abszolút folytonos** *valószínűségi változónak nevezzük.*

1.1. A sűrűségfüggvény tulajdonságai

1.1. Állítás. *Ha az X valószínűségi változó sűrűségfüggvénye f, akkor tetszőleges $a < b$ számokra*

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f(x)dx.$$

1.2. Állítás. *Legyen X abszolút folytonos valószínűségi változó, melynek F az eloszlásfüggvénye.*
(a) Ha f az X sűrűségfüggvénye, akkor minden $t \in \mathbb{R}$ számra

$$F(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f(x)\, dx.$$

(b) Az $f(t) = F'(t)$ függvény (azokra a t-kre, ahol F differenciálható) az X sűrűségfüggvénye.
Ha az $f : \mathbb{R} \to \mathbb{R}$ függvény sűrűségfüggvény, akkor

(i) $f(x) \geq 0$ teljesül „majdnem minden” $x \in \mathbb{R}$-re (például véges vagy megszámlálható sok kivétel lehetséges).

(ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

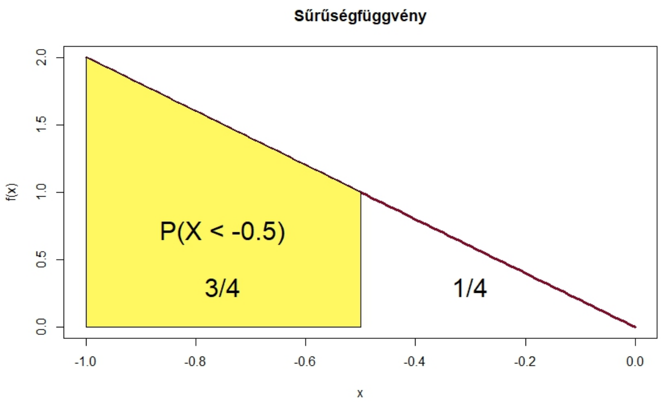
Fordítva: ha f teljesíti ezt a két tulajdonságot, akkor van olyan valószínűségi változó, aminek f a sűrűségfüggvénye.

1.2. Példa

Legyen az X valószínűségi változó sűrűségfüggvénye $f(x) = 2|x|$, ha $-1 < x < 0$, és 0 különben. Mennyi X eloszlásfüggvényének értéke a $-1/2$ helyen?

Felhasználva az **eloszlásfüggvény** és a **sűrűségfüggvény** definícióját (az 1.2. állítás (a) részét), illetve hogy $x \leq -1$ esetén $f(x) = 0$, azt kapjuk, hogy

$$F(-1/2) = \mathbb{P}(X \leq -1/2) = \int_{-\infty}^{-1/2} f(x)dx = \int_{-1}^{-1/2} f(x)dx = \int_{-1}^{-1/2} 2|x|\,dx =$$
$$= -\int_{-1}^{-1/2} 2x\,dx = -\left[x^2\right]_{x=-1}^{x=-1/2} = -\left(\frac{1}{4} - 1\right) = \mathbf{\frac{3}{4}}.$$



2. Várható érték és szórás abszolút folytonos esetben

Mennyi lehet a 2. ábrán látható eloszlás várható értéke és szórása?

A **várható értéket** olyan formában nem tudjuk definiálni abszolút folytonos eloszlásokra, mint diszkrét esetben, hiszen $\mathbb{P}(X = x) = 0$ minden x-re. Helyette:

2.1. Definíció. *Legyen X abszolút folytonos valószínűségi változó, melynek sűrűségfüggvénye f. Ekkor X* **várható értéke***:*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx,$$

ha ez az integrál létezik és véges.

2.2. Definíció. *Tegyük fel, hogy az X valószínűségi változó abszolút folytonos, sűrűségfüggvénye f, és $\mathbb{E}(X^2)$ létezik, azaz az $\int_{-\infty}^{\infty} x^2 f(x)\,dx$ integrál véges. Ekkor X* **szórásnégyzete***:*

$$D^2(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

diszkrét	**abszolút folytonos**
X lehetséges értékei: $x_1, x_2, \ldots$	X sűrűségfüggvénye: f.
$\boxed{\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j \cdot \mathbb{P}(X = x_j)}$	$\boxed{\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx}$
$\boxed{\mathbb{E}(X^2) = \sum_{j=1}^{\infty} x_j^2 \cdot \mathbb{P}(X = x_j)}$	$\boxed{\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx}$

$$\boxed{D(X) = \sqrt{\mathbb{E}\big((X - \mathbb{E}(X))^2\big)} = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}}$$

$\mathbb{E}(X^k) = \sum_{j=1}^{\infty} x_j^k \cdot \mathbb{P}(X = x_j)$	$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx$
$\boxed{\mathbb{E}(g(X)) = \sum_{j=1}^{\infty} g(x_j) \cdot \mathbb{P}(X = x_j)}$	$\boxed{\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx}$

szórása *pedig*

$$D(X) = \sqrt{\mathbb{E}\big((X - \mathbb{E}(X))^2\big)} = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}.$$

Láthatjuk, hogy a szórás definíciója megegyezik a diszkrét esetben használttal.

2.1. Momentumok

Az X valószínűségi változók **k. momentuma** a k. hatványának várható értéke, ha ez létezik:

$$\mathbb{E}(X^k).$$

Ennek a többek között a statisztikában van jelentősége, az eloszlások ismeretlen paramétereinek becslésekor.

2.1. Állítás. *Ha X abszolút folytonos valószínűségi változó, f a sűrűségfüggvénye, és $\mathbb{E}(g(X))$ létezik, akkor*

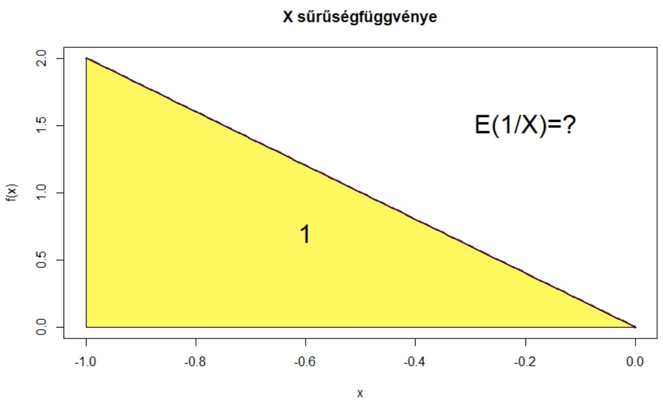
$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx.$$

Ezért a k. momentum kiszámítása:

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx.$$

2.2. Következmény. *A* **szórásnégyzetet** *a következőképpen számíthatjuk ki abszolút folytonos X valószínűségi változó esetén:*

$$D^2(X) = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} x \cdot f(x)dx\right]^2.$$



Példa. Legyen az X valószínűségi változó sűrűségfüggvénye $f(x) = 2|x|$, ha $-1 < x < 0$, és 0 különben. Mennyi az $1/X$ valószínűségi változó várható értéke?

Mivel X sűrűségfüggvénye azonosan 0, ha $x > 0$, ezért $X < 0$ és $1/X < 0$ biztosan teljesül. Így $\mathbb{E}(\mathbf{1}/\mathbf{X}) < \mathbf{0}$ teljesülni fog.

Pontosabban, mivel

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx,$$

a $g(x) = 1/x$ függvénnyel:

$$\mathbb{E}(1/X) = \int_{-\infty}^{\infty} \frac{1}{x} \cdot f(x)\,dx = \int_{-1}^{0} \frac{2|x|}{x}\,dx = \int_{-1}^{0} (-2)\,dx = \mathbf{-2}.$$

Példa. Tegyük fel, hogy a holnap hulló csapadék mennyiségének sűrűségfüggvénye az alábbi:

$$f(x) = \begin{cases} 0,2, & \text{ha } 0 \le x \le 1; \\ 0,4, & \text{ha } 1 < x \le 3; \\ 0 & \text{ha } x < 0 \text{ vagy } x > 2. \end{cases}$$

Jelölje a csapadékmennyiséget X. A csapadékmennyiség **várható értéke**:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot 0,2dx + \int_1^3 x \cdot 0,4dx =$$
$$= 0,2 \cdot \left(\frac{1^2}{2} - \frac{0^2}{2}\right) + 0,4 \cdot \left(\frac{3^2}{2} - \frac{1^2}{2}\right) = \mathbf{1,7}.$$

A csapadékmennyiség négyzetének várható értéke:

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, ds = \int_0^1 x^2 \cdot 0,2\, dx + \int_1^3 x^2 \cdot 0,4\, dx =$$
$$= 0,2 \cdot \left(\frac{1^3}{3} - \frac{0^3}{3}\right) + 0,4 \cdot \left(\frac{3^3}{3} - \frac{1^3}{3}\right) = \mathbf{3,53}.$$

Ez alapján a csapadékmennyiség szórása:

$$D(X) = \sqrt{\mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2} = \sqrt{3,53 - 1,7^2} = \mathbf{0,8}.$$

3. Egyenletes eloszlás

A korábban látott három eloszlás, az egyenletes, exponenciális– és Pareto-eloszlás valójában abszolút folytonos eloszlások. Eddig az eloszlásfüggvény segítségével definiáltuk őket, de a sűrűségfüggvényüket is meghatározhatjuk, abból pedig a várható értéküket, szórásukat is. Egyelőre az egyenletes eloszlást nézzük meg részletesebben.

3.1. Egyenletes eloszlás

- Csomagot várunk, amit a futár véletlen Y időpontban hoz ki.
- Feltételezzük, hogy Y egyenletes eloszlású a $[8, 12]$ intervallumon (órában mérve).
- **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:

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

Ahogy az 1.2. állításban láttuk, a sűrűségfüggvény, ha létezik, az eloszlásfüggvény deriváltjaként áll elő. Így kaphatjuk meg a már ismert eloszlásfüggvényből a sűrűségfüggvényt.

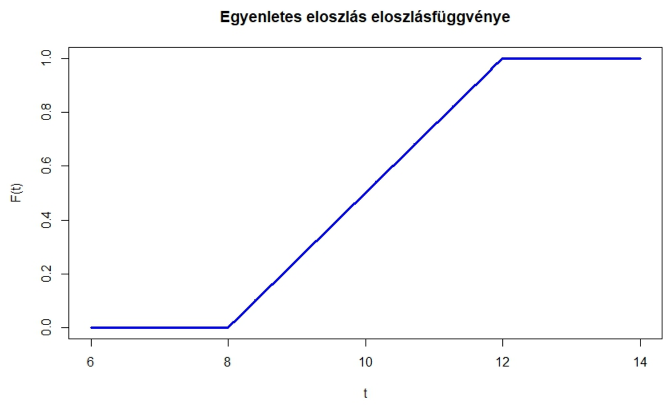
3.1. Definíció. *Az X valószínűségi változó* **egyenletes eloszlású** *az $[a, b]$ intervallumon, ha eloszlásfüggvénye:*

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

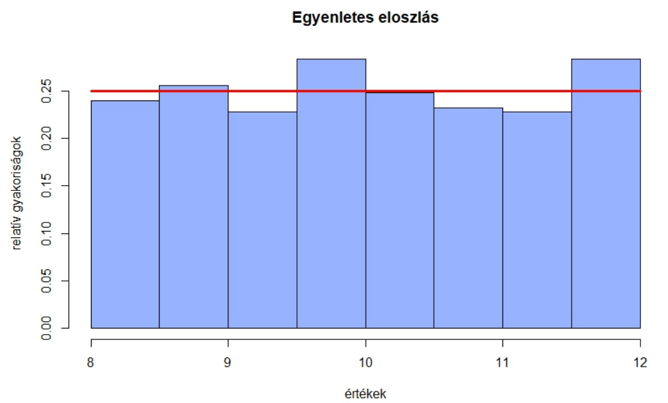
3.1. Állítás. *Legyen X egyenletes eloszlású az (a, b) intervallumon. Ekkor X sűrűségfüggvénye:*

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \le x \le b; \\ 0, & \text{különben.} \end{cases}$$

Továbbá



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



4. ábra. A $[8, 12]$ intervallumon egyenletes eloszlás sűrűségfüggvénye és 500 darab független, $[8, 12]$ intervallumon egyenletes eloszlású valószínűségi változóból álló minta hisztogramja

(i) Ha $a \le c \le d \le b$, akkor

$$\mathbb{P}(c \le X \le d) = \int_c^d f(x)\,dx = \int_c^d \frac{1}{b-a}\,dx = \frac{d-c}{b-a}.$$

(ii) Az X valószínűségi változó **várható értéke** *és* **szórása***:*

$$\mathbb{E}(X) = \frac{a+b}{2}; \qquad D(X) = \frac{b-a}{\sqrt{12}}.$$

Bizonyítás. Könnyen látható, hogy a megadott $f(x)$ függvényt $-\infty$-től t-ig integrálva éppen $F(t) = \mathbb{P}(X \le t)$-t kapjuk, ez lesz tehát a sűrűségfüggvény.

(i) Itt az állítás a bizonyítást is tartalmazza, a sűrűségfüggvény egyik tulajdonságát (az 1.2. állítás) használjuk, majd a konstans integrálját (amihez valójában csak a téglalap alatti területre van szükség).

(ii) A várható értéket a definíció alapján számíthatjuk ki:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} s \cdot f(s)ds = \int_a^b s \cdot \frac{1}{b-a}ds = \frac{1}{b-a}\left[\frac{s^2}{2}\right]_{s=a}^{b} = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2},$$

hiszen az x függvény primitív függvénye $\frac{x^2}{2}$, és $b^2 - a^2 = (b-a)(b+a)$.

(iii) A szórást szintén definíció alapján számoljuk, először a négyzet várható értékét kiszámítva:

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} s^2 \cdot f(s) ds = \int_a^b s^2 \cdot \frac{1}{b-a} ds = \frac{1}{b-a}\left[\frac{s^3}{3}\right]_{s=a}^{b} \\ &= \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} = \frac{a^2+ab+b^2}{3},\end{aligned}$$

hiszen az x^2 függvény primitív függvénye $\frac{x^3}{3}$, és $b^3 - a^3 = (b-a)(a^2+ab+b^2)$.

Ebből a szórásnégyzet:

$$\begin{aligned}D^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12} \quad \Rightarrow \quad D(X) = \frac{b-a}{\sqrt{12}}.\end{aligned}$$

Vagyis a várható érték az intervallum közepe, a szórás pedig egyenesen arányos az intervallum hosszával

Példa. Csomagot várunk, a futár 8 és 12 óra között érkezik. Feltesszük, hogy érkezésének időpontja egyenletes eloszlású a $[8, 12]$ intervallumon. Ekkor az előző állítás alapján az alábbiak igazak ($a = 8, b = 12$).

- Annak valószínűsége, hogy 10 és 11 óra között érkezik: $(11-10)/(12-8) = 1/4$.
- Annak valószínűsége, hogy 10:15 és 10:30 között érkezik: $1/16 = 0,0625$.
- Érkezési időpontjának várható értéke: $(8+12)/2 = 10$ óra.
- Érkezési időpontjának szórása: $(12-8)/\sqrt{12} = 2/\sqrt{3} = 1,154$.

Házi feladat november 18., szerda, 8:00-ig Tegyük fel, hogy az X valószínűségi változó sűrűségfüggvénye $x+c$, ha $0 \leq x \leq 1$, és 0 különben. Határozzuk meg c értékét, számítsuk ki X várható értékét és szórását.