További nevezetes eloszlások (9. előadás)

Az alábbi eloszlások többek között **statisztikai alkalmazásokban** fordulnak elő:

- Pareto-eloszlás: végtelen momentumokkal rendelkező eloszlások modellezésére (például jövedelmek, kárnagyságok)
- t-eloszlás: például két eloszlás **várható értékének** összehasonlítására
- F-eloszlás: például két eloszlás **szórásának** összehasonlítására
- χ^2-eloszlás: például annak eldöntésére, hogy két tulajdonság között van-e **összefüggés**
- gamma-eloszás: nemnegatív valószínűségi változók modellezésére
- beta-eloszlás: $[0,1]$-értékű valószínűségi változók modellezésére

Pareto-eloszlás

Az X valószínűségi változó Pareto-eloszlású, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\beta}{\alpha^{\beta+1}} \cdot x^{-\beta}; & \text{ha } x \leq \alpha \\ 0, & \text{ha } x < \alpha. \end{cases}$$

Itt $\alpha > 0, \beta > 1$ rögzített számok. Ekkor X eloszlásfüggvénye:

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

Pareto-eloszlás

Az X valószínűségi változó Pareto-eloszlású, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\beta}{\alpha^{\beta+1}} \cdot x^{-\beta}; & \text{ha } x \leq \alpha \\ 0, & \text{ha } x < \alpha. \end{cases}$$

Itt $\alpha > 0, \beta > 1$ rögzített számok. Ekkor X eloszlásfüggvénye:

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

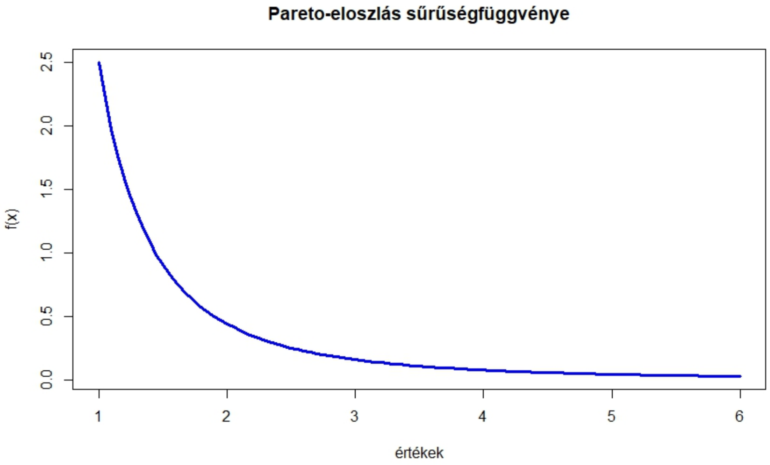
Az X valószínűségi változó k. momentuma:

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k \cdot f(x)\, dx = \frac{\beta}{\alpha^{\beta+1}} x^{k-\beta}\, dx < \infty \Leftrightarrow k - \beta < -1.$$

Tehát a Pareto-eloszlásnak csak $\beta - 1$-nél kisebb k-ra véges a k. momentuma.

Például ha $\beta = 2,5$, akkor a várható érték létezik és véges, de szórása nem létezik.

Pareto-eloszlás



Az $\alpha = 1, \beta = 2,5$ paraméterű Pareto-eloszlás sűrűségfüggvénye: $f(x) = 2,5 \cdot x^{-2,5}$, ha $x \geq 1$, és 0 különben. A **várható értéke véges**, a **szórása végtelen**.

t-eloszlás

Legyenek $X_1, X_2, \ldots, X_f$ és Y független standard normális eloszlású valószínűségi változók. Ekkor a

$$Z = \frac{Y}{\sqrt{(X_1^2 + X_2^2 + \ldots + X_f^2)/f}}$$

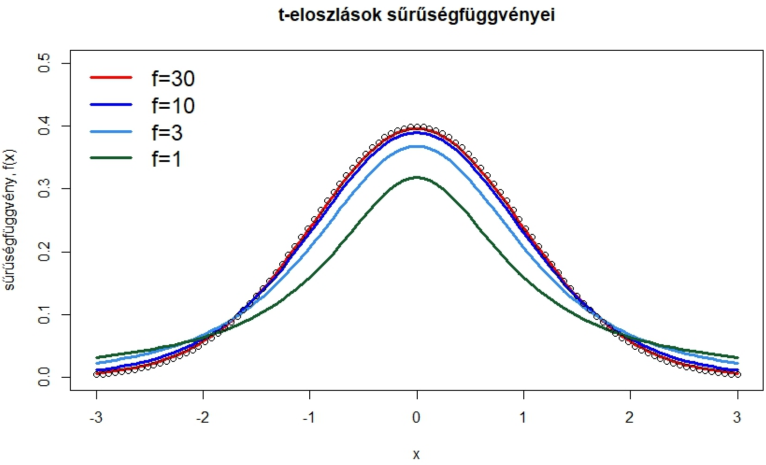
valószínűségi változó eloszlását f szabadsági fokú **t-eloszlásnak** (vagy Student-eloszlásnak) nevezzük.

Az $f = 1$ szabadsági fokú t-eloszlás, vagyis Y/X eloszlása a **Cauchy-eloszlás**. Ennek sűrűségfüggvénye:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

A Cauchy-eloszlásnak sem várható értéke, sem szórása nem létezik: $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$ nem értelmezhető, mert $\int_0^{\infty} \frac{1}{x}\, dx$ integrál nem véges.

A t-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú t-eloszlások sűrűségfüggvényei. A pöttyözött vonal a standard normális eloszlás sűrűségfüggvényét jelöli, ez közel van a t-eloszlás sűrűségfüggvényéhez, ha f nagy.

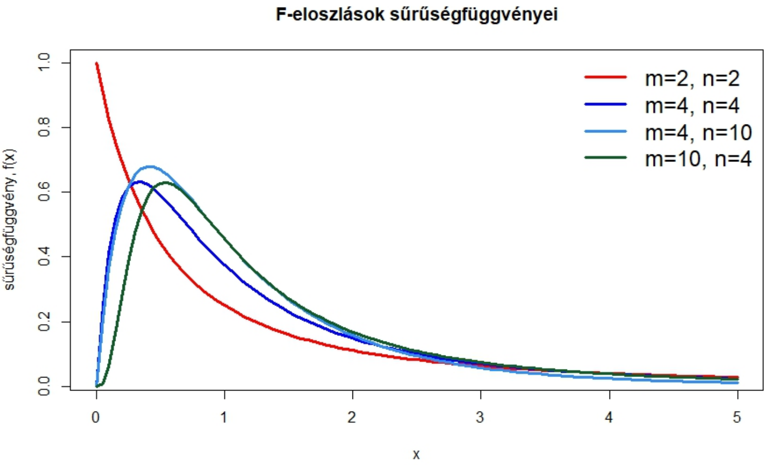
F-eloszlás

Legyenek m, n pozitív egészek, $X_1, \ldots, X_m$, $Y_1, Y_2, \ldots, Y_n$ pedig független standard normális eloszlású valószínűségi változók. Ekkor az

$$F = \frac{n(X_1^2 + X_2^2 + \ldots + X_m^2)}{m(Y_1^2 + Y_2^2 + \ldots + Y_n^2)}$$

valószínűségi változó eloszlását m, n szabadsági fokú **F-eloszlásnak** nevezzük.

Az F-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú F-eloszlások sűrűségfüggvényei

χ^2-eloszlás

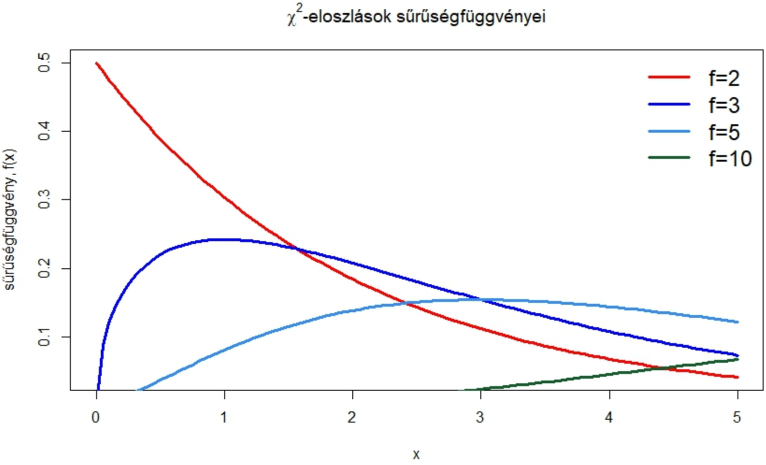
Legyenek $X_1, X_2, \ldots, X_q$ független standard normális eloszlású valószínűségi változók. Az

$$Y = X_1^2 + X_2^2 + \ldots + X_q^2$$

valószínűségi változó eloszlását q szabadsági fokú **χ^2-eloszlásnak** nevezzük. Ennek sűrűségfüggvénye:

$$f(t) = \begin{cases} \frac{t^{q/2-1}}{2^{q/2}\Gamma(q/2)} e^{-t/2}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

A χ^2-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú χ^2-eloszlások sűrűségfüggvényei

Gamma-eloszlás

Gamma-függvény. Ha $a > 0$ pozitív szám, legyen

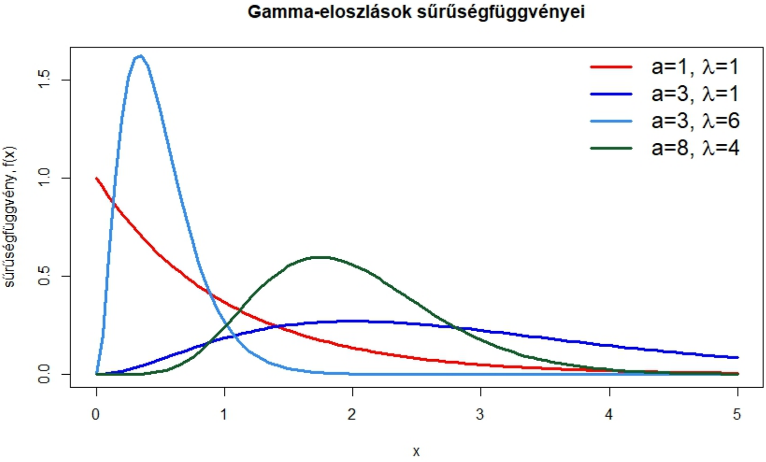
$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Parciális integrálással belátható, hogy $\Gamma(a) = (a-1)\Gamma(a-1)$ minden $a > 1$-re, és így $\Gamma(n) = (n-1)!$, ha n pozitív egész.

Legyenek a és λ pozitív számok. Az X valószínűségi változó **gamma-eloszlású** a renddel és λ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

A gamma-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú gamma-eloszlások sűrűségfüggvényei

A gamma-eloszlás tulajdonságai

Az X valószínűségi változó **gamma-eloszlású** a renddel és λ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

- **Kapcsolat az exponenciális eloszlással**: ha $a = 1$, akkor a sűrűségfüggvény $\lambda e^{-\lambda x}$, ha $x \geq 0$, és az exponenciális eloszlást kapjuk vissza.
- **Exponenciális eloszlások összege:** ha $X_1, X_2, \ldots, X_n$ független λ paraméterű exponenciális eloszlású valószínűségi változók, akkor $X_1 + X_2 + \ldots + X_n$ gamma-eloszlású $a = n$ renddel és λ paraméterrel.
- **Kapcsolat a χ^2-eloszlással**: ha $a = q/2$ és $\lambda = 1/2$, akkor a q szabadsági fokú χ^2-eloszlást kapjuk vissza.
- **Várható érték** és **szórás:**

$$\mathbb{E}(X) = \frac{a}{\lambda}; \qquad D(X) = \frac{\sqrt{a}}{\lambda}.$$

Beta-eloszlás

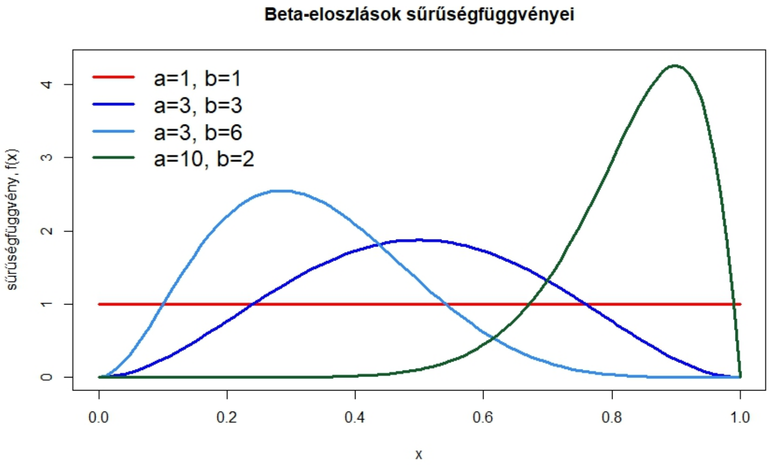
Legyenek $a, b > 1$ számok. Az X valószínűségi változó **beta-eloszlású** a és b paraméterekkel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, & t \in [0,1]; \\ 0, & x < 0 \text{ vagy } x > 1. \end{cases}$$

Ha $X_1, X_2, \ldots, X_n$ független, a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változók, és X_k^* jelöli ebben a mintában a k. legnagyobb számot, akkor X_k^* eloszlása beta-eloszlás $a = k$ és $b = n - k + 1$ paraméterekkel.

Az $a = 1$ és $b = 1$ választással az **egyenletes eloszlást** kapjuk vissza.

A beta-eloszlás sűrűségfüggvénye



Különböző szabadsági fokú beta-eloszlások sűrűségfüggvényei

Házi feladat november 20., 10:15-ig

Tegyük fel, hogy egy mosógép élettartama

(i) exponenciális eloszlású 10 év várható értékkel;

(ii) normális eloszlású 10 év várható értékkel és 10 év szórással.

Mennyi garanciaidőt adhatunk (évben számolva), ha azt szeretnénk, hogy annak valószínűsége, hogy egy mosógép a garanciaidőn belül elromlik, 5% legyen?

Házi feladat november 20., 10:15-ig

Tegyük fel, hogy egy mosógép élettartama, T exponenciális eloszlású 10 év várható értékkel.

Mennyi garanciaidőt adhatunk (évben számolva), ha azt szeretnénk, hogy annak valószínűsége, hogy egy mosógép a garanciaidőn belül elromlik, 5% legyen?

Mivel exponenciális esetén $\mathbb{E}(T) = 1/\lambda$, az eloszlás paramétere $\lambda = 0,1$. Legyen t a garanciaidő. Ekkor az exponenciális eloszlás eloszlásfüggvénye alapján

$$0,05 = \mathbb{P}(T \leq t) = 1 - \exp(-\lambda \cdot t) = 1 - \exp(-0,1 \cdot t)$$

$$0,95 = \exp(-0,1 \cdot t) \qquad \Rightarrow \qquad -0,1 \cdot t = \log 0,95$$

$$t = -\frac{\log 0,95}{0,1} = 0,513.$$

Házi feladat november 20., 10:15-ig

Tegyük fel, hogy egy mosógép élettartama, T normális eloszlású 10 év várható értékkel és 10 szórással.

Mennyi garanciaidőt adhatunk (évben számolva), ha azt szeretnénk, hogy annak valószínűsége, hogy egy mosógép a garanciaidőn belül elromlik, 5% legyen?

$$0,05 = \mathbb{P}(T \leq t) = \Phi\left(\frac{t-10}{10}\right)$$

$$\Phi^{-1}(0,05) = -1,645 = \frac{t-10}{10} \qquad \Rightarrow \qquad t = -1,645 \cdot 10 + 10 < 0,$$

vagyis ebben az esetben nincs ilyen t.

Másképpen:

$$\mathbb{P}(T \leq 0) = \Phi\left(\frac{0-10}{10}\right) = \Phi(-1) = 15,9\%$$

A $\Phi^{-1}(0,05)$ kiszámítása R-ben: `qnorm(0.05)`.

A $\Phi(-1)$ kiszámítása R-ben: `pnorm(-1)`.

Házi feladat november 27., 10:15-ig

Válasszunk egy tetszőleges f pozitív egészt (legalább 10 legyen).

Generáljunk 100 darab olyan $f + 1$ hosszú vektort, melyek koordinátái független standard normális eloszlásúak.

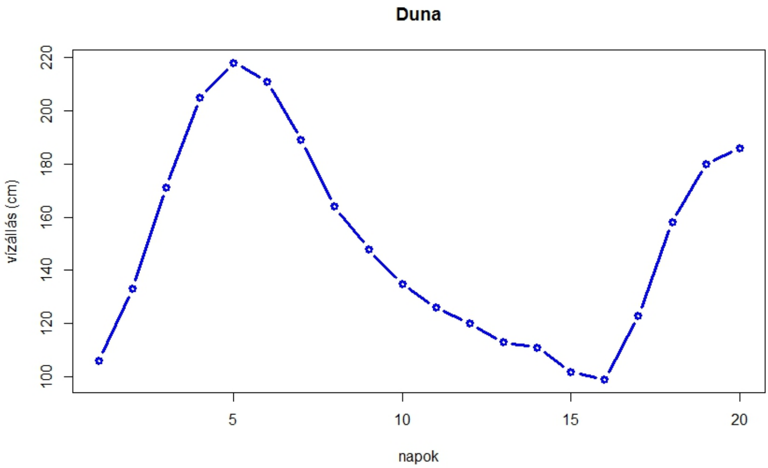
Minden ilyen vektorra számítsuk ki a

$$z = \frac{y}{\sqrt{(x_1^2 + x_2^2 + \ldots + x_f^2)/f}}$$

mennyiséget, ahol Y a vektor első koordinátája, $x_1, x_2, \ldots, x_f$ pedig a további koordináták.

Készítsünk hisztogramot az így kapott 100 értékből.

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

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

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

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

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 megszámlálható végtelen.

Együttes eloszlásfüggvény

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

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

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

- egy véletlenszerűen választott embert 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).

Együttes sűrűségfüggvény

Az $\underline{X} = (X_1, \ldots, X_n)$ valószínűségi vektorváltozó **abszolút folytonos**, ha van olyan $f : \mathbb{R}^n \to \mathbb{R}$ függvény, melyre

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_n} \ldots \int_{-\infty}^{t_1} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ esetén. Ilyenkor az f függvényt az $(X_1, X_2, \ldots, X_n)$ **együttes sűrűségfüggvényének** nevezzük.

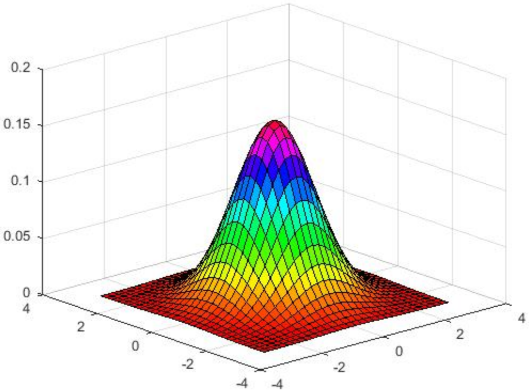
Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Ekkor minden (megfelelő) $A \subseteq \mathbb{R}^n$ halmazra

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in A) = \int_A f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n.$$

Következmény:

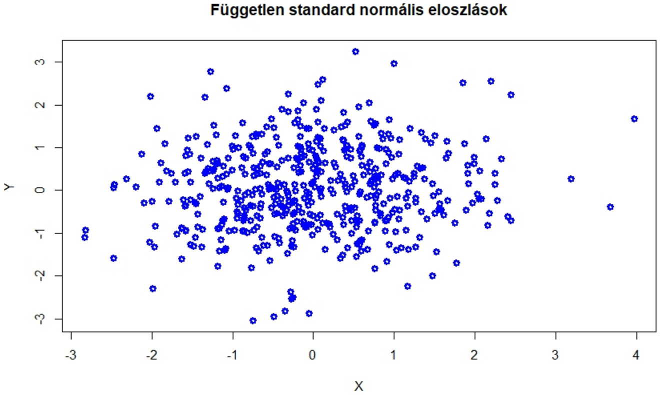
$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n = 1.$$

Kétdimenziós normális eloszlás



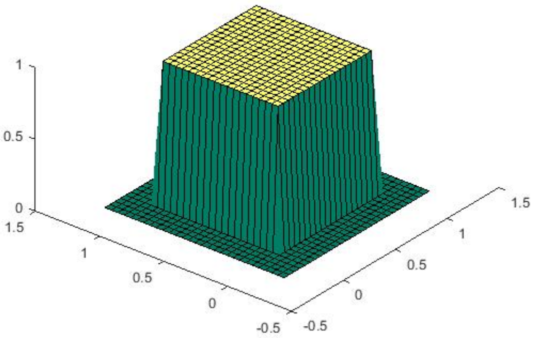
Két független standard normális eloszlás együttes sűrűségfüggvénye Azaz: (X, Y) együttes sűrűségfüggvénye, ahol X, Y függetlenek, $N(0, 1)$ eloszlásúak

Kétdimenziós normális eloszlás



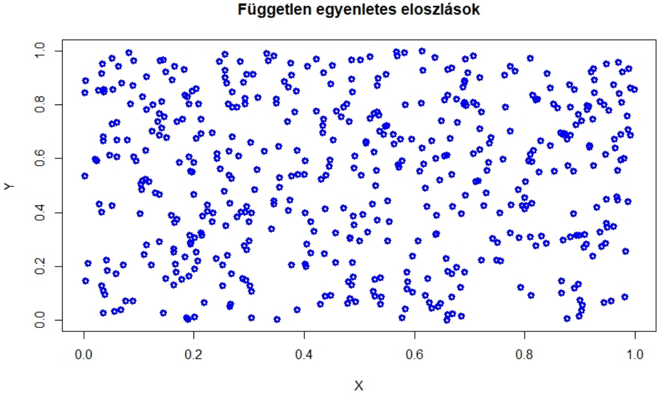
500 darab véletlen pont a síkon, melyek koordinátái független standard nomális eloszlásúak. Ahol nagyobb az együttes sűrűségfüggvény (előző ábra), oda több pont esik.

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).

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó **együttes sűrűségfüggvénye** f. Hogyan kapható meg például az **első peremeloszlás, azaz X_1 sűrűségfüggvénye**?

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó **együttes sűrűségfüggvénye** f. Hogyan kapható meg például az **első peremeloszlás, azaz X_1 sűrűségfüggvénye**?

Az X_j valószínűségi változó sűrűségfüggvénye (melyet f_j-vel jelölünk), azaz a j. peremsűrűségfüggvény így kapható meg f-ből:

$$f_j(t) = \int_{\mathbb{R}^{n-1}} f(s_1, \ldots, s_{j-1}, t, s_{j+1}, \ldots, s_n) ds_1 \ldots ds_{j-1} ds_{j+1} \ldots ds_n.$$

Speciálisan $n = 2$-re:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Hogyan kapható meg például az első peremeloszlás, azaz X_1 sűrűségfüggvénye?

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Hogyan kapható meg például az első peremeloszlás, azaz X_1 sűrűségfüggvénye?

Állítás

Tegyük fel, hogy az $\underline{X} = (X_1, \ldots, X_n)$ valószínűségi változók együttes sűrűségfüggvénye f. Ekkor az X_j valószínűségi változó sűrűségfüggvénye (melyet f_j-vel jelölünk), azaz a j. peremsűrűségfüggvény így kapható meg f-ből:

$$f_j(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(s_1, \ldots, s_{j-1}, t, s_{j+1}, \ldots, s_n) ds_1 \ldots ds_{j-1} ds_{j+1} \ldots ds_n.$$

Speciálisan $n = 2$-re:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Függetlenség

Definíció (Véges eset)

Azt mondjuk, hogy az $X_1, \dots, X_n : \Omega \to \mathbb{R}$ valószínűségi változók **függetlenek**, *ha*

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

teljesül tetszőleges $t_1, t_2, \dots, t_n$ valós számokra, azaz

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

ahol F_j az X_j valószínűségi változó eloszlásfüggvénye.

Definíció (Végtelen eset)

Az $X_1, X_2, X_3 \dots$ valószínűségi változók függetlenek, ha közülük bármely véges sokat kiválasztva független valószínűségi változókat kapunk.

Függetlenség és együttes sűrűségfüggvény

Állítás (Függetlenség és sűrűségfüggvény)

Legyen az $\underline{X} = (X_1, \dots, X_n)$ *valószínűségi vektorváltozó együttes sűrűségfüggvénye* f*, továbbá az* X_j *valószínűségi változó sűrűségfüggvénye* f_j *minden* $j = 1, 2, \dots, n$ *esetén. Ezekkel a jelölésekkel:* $X_1, X_2, \dots, X_n$ *pontosan akkor függetlenek, ha*

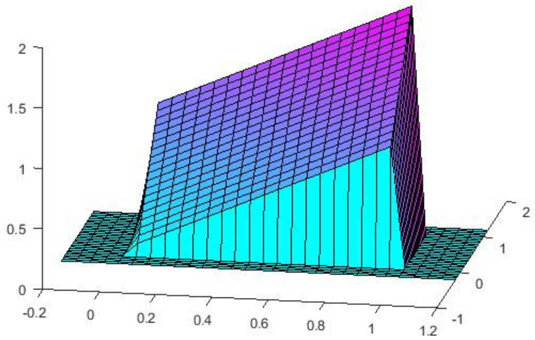
$$f(t_1, \dots, t_n) = f_1(t_1) \cdot f_2(t_2) \dots f_n(t_n)$$

teljesül bármely $t_1, t_2, \dots, t_n$ *valós számokra.*

Például: ha X_1, X_2 független standard normális eloszlású valószínűségi változók, akkor együttes sűrűségfüggvényük

$$f(t_1, t_2) = f_1(t_1) f_2(t_2) = \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_1^2}{2} \right) \right) \cdot \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_2^2}{2} \right) \right)$$
$$= \frac{1}{2\pi} \exp\left(-\frac{t_1^2 + t_2^2}{2} \right).$$

Együttes sűrűségfüggvény: példa



A $[0,1]\times[0,1]$ négyzeten $x+y$ alakú együttes sűrűségfüggvény