Valószínűségszámítás

2020. október 28.

Mészáros Szabolcs

Tárgyhonlap:
cs.bme.hu/valszam

A prezentáció anyagát és az abból készült videofelvételt a tárgy hallgatói jogosultak használni, kizárólag saját célra. A felvétel másolása, videómegosztókra való feltöltése részben vagy egészben tilos, illetve csak a tantárgyfelelős előzetes engedélyével történhet.

Copyright © 2020, BME VIK

valószínűségi vektorváltozó

Definíció: Legyenek $X_1, \ldots, X_n$ val. változók. Ekkor az

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

függvényt *valószínűségi vektorváltozó*nak hívjuk.

Definíció: Az $\underline{X}$ val. vektorváltozó *(együttes) eloszlásfüggvény*e az

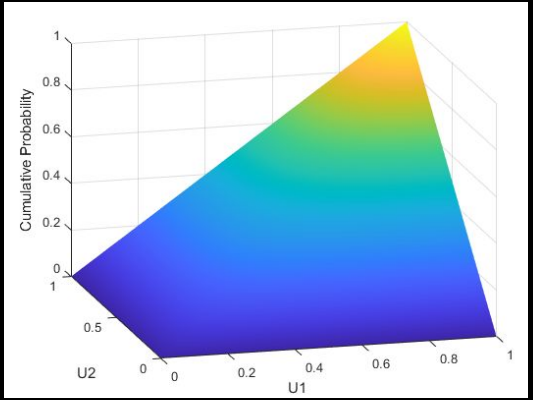
$$F_{\underline{X}} : \mathbb{R}^n \to [0, 1]$$

$$F_{\underline{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 < x_1, \ldots, X_n < x_n)$$

vektorváltozós, skalárértékű függvény.

Val. vektorváltozó, példa

Példa: Válasszunk egyenletesen véletlenszerűen egy pontot az egységnégyzetben. Jelölje (X,Y) a pont koordinátáit. Mi (X,Y) együttes eloszlásfüggvénye?



Forrás: www.mathworks.com/help/stats/copulas-generate-correlated-samples.html

$$F_{(X,Y)}(x,y) = \mathbb{P}(X < x, Y < y) =$$

$$= \begin{cases} 1 & \text{ha } x > 1, y > 1 \\ x & \text{ha } 0 < x \leq 1, y > 1 \\ y & \text{ha } 0 < y \leq 1, x > 1 \\ xy & \text{ha } 0 < x, y \leq 1 \\ 0 & \text{egyébkent} \end{cases} = \max\left(0, \min(1,x)\right) \cdot \max\left(0, \min(1,y)\right)$$

Val. vektorváltozó, példa

Példa: Aladárral és Bélával csetelünk.

Ha nincs meccs: egymástól független, Exp(6) eloszlás szerint válaszolnak.

Ha meccs van: Aladár dupla annyi idő alatt reagál, Béla feleannyi idő alatt.

Annak az esélye, hogy ma meccs van: $\frac{1}{5}$

Kérdés: mi a válaszidők együttes eloszlásfüggvénye?

X : Aladár válaszideje

Y : Béla válaszideje

$M = \{\text{meccs van}\}$

$U, V \sim \mathrm{Exp}(6)$

$$Z = \begin{cases} 2 & \text{ha meccs van,} \\ 1 & \text{egyébként.} \end{cases}$$

$$X = U \cdot Z \qquad Y = V/Z$$

Val. vektorváltozó, példa

$$F_{(X,Y)}(x,y) = \mathbb{P}(X < x, Y < y) = \mathbb{P}(U \cdot Z < x, V/Z < y) =$$

$$= \mathbb{P}(U \cdot Z < x, V/Z < y \mid M)\mathbb{P}(M)$$

$$+\mathbb{P}(U \cdot Z < x, V/Z < y \mid \overline{M})\mathbb{P}(\overline{M})$$

$$= \mathbb{P}(U \cdot 2 < x, V/2 < y)\frac{1}{5} + \mathbb{P}(U < x, V < y)\frac{4}{5}$$

$$(x, y > 0)$$

$$= \mathbb{P}(U < \frac{x}{2})\mathbb{P}(V < 2y)\frac{1}{5} + \mathbb{P}(U < x)\mathbb{P}(V < y)\frac{4}{5}$$

$$= (1 - e^{-6\frac{x}{2}})(1 - e^{-6 \cdot 2y})\frac{1}{5} + (1 - e^{-6x})(1 - e^{-6y})\frac{4}{5}$$

Együttes sűrűségfüggvény, def.

Definíció: Legyen $\underline{X} = (X_1, \ldots, X_n)$ val. vektorváltozó. Egy

$$f_{\underline{X}} : \mathbb{R}^n \to [0, \infty)$$

függvény az $\underline{X}$ *(együttes) sűrűségfüggvénye*, ha

$$\int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\underline{X}}(z_1, \ldots, z_n) \mathrm{d}z_1 \ldots \mathrm{d}z_n = F_{\underline{X}}(x_1, \ldots, x_n)$$

tetszőleges $x_1, \ldots, x_n \in \mathbb{R}$ esetén.

Definíció: Az $\underline{X}$ val. vektorváltozót *folytonos*nak hívjuk, ha létezik együttes sűrűségfüggvénye.

Együttes sfv., valószínűségek

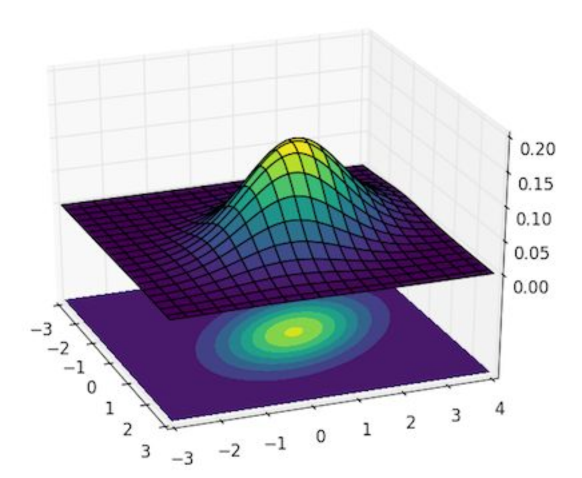
Állítás: Legyen $\underline{X}$ val. vektorváltozó, és $H \subseteq \mathbb{R}^n$ Jordan-mérhető.

Ekkor

$$\mathbb{P}(\underline{X} \in H) = \int_H f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x}$$

Megjegyzések:

- Jordan-mérhető $\sim$ van térfogata.
- Vesd össze: egy dimenziós állítás a $\mathbb{P}(a < X < b)$ mennyiségről.



Forrás: scipython.com/blog/visualizing-the-bivariate-gaussian-distribution/

Együttes sfv., meghatározás

Állítás: Legyen $\underline{X} = (X_1, \dots, X_n)$ val. vektorváltozó. Ha $\underline{X}$ folytonos, akkor az alábbi függvény a sűrűségfüggvénye lesz:

$$f_{\underline{X}}(x_1, \dots, x_n) = \begin{cases} \partial_{x_1} \dots \partial_{x_n} F_{\underline{X}}(x_1, \dots, x_n) & \text{ha ez létezik,} \\ 0 & \text{egyébként.} \end{cases}$$

Megjegyzések:

- A deriválások sorrendje tetszőleges.
- Feltesszük, hogy létezik a sűrűségfüggvény (nem pedig következtetjük). Pl. ha $\underline{X} = (X_1, X_2)$ és $X_1 = X_2$ akkor $\underline{X}$ nem folytonos.

Együttes sfv., meghatározás

Példa: Mi az esélye, hogy Aladár előbb ír vissza, mint Béla?

$$F_{(X,Y)}(x,y) = (1-e^{-6\frac{x}{2}})(1-e^{-6\cdot 2y})\frac{1}{5} + (1-e^{-6x})(1-e^{-6y})\frac{4}{5}$$

$$(x, y > 0)$$

$$\partial_y F_{(X,Y)}(x,y) = (1-e^{-6\frac{x}{2}})12e^{-12y}\frac{1}{5} + (1-e^{-6x})6e^{-6y}\frac{4}{5}$$

$$f_{(X,Y)}(x,y) = \partial_x \partial_y F_{(X,Y)}(x,y) =$$

$$= 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5}$$

Együttes sfv., meghatározás

Példa: Mi az esélye, hogy Aladár előbb ír vissza, mint Béla?

$$f_{(X,Y)}(x,y) = 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5} \qquad (x,y>0)$$

$$\mathbb{P}(X<Y) = \int_{\{x<y\}} f_{(X,Y)}(x,y)\mathrm{d}x\mathrm{d}y =$$

$$= \int_0^\infty \int_0^y \left(3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5}\right)\mathrm{d}x\mathrm{d}y$$

$$= \int_0^\infty \left(\left[3e^{-3x}\right]_0^y \cdot 12e^{-12y}\frac{1}{5} + \left[6e^{-6x}\right]_0^y \cdot 6e^{-6y}\frac{4}{5}\right)\mathrm{d}x\mathrm{d}y = 0{,}44$$

Együttes sfv., karakterizáció

Állítás: Legyen $f : \mathbb{R}^n \to [0, \infty)$. Ekkor

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n = 1$$

pontosan akkor teljesül, ha létezik $\underline{X}$ val. vektorváltozó, aminek sűrűségfüggvénye éppen f.

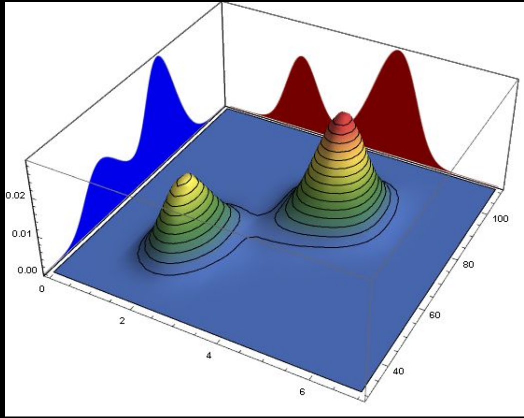
(Vesd össze: egy-változós sűrűségfüggvény karakterizáció.)

Marginális eloszlás, definíció

Definíció: Ha $\underline{X} = (X_1, \ldots, X_n)$ val. vektorváltozó, akkor az X_i eloszlását az $\underline{X}$ i-edik *marginális eloszlás*ának (avagy peremeloszlásának) hívjuk.

Állítás: Ha $\underline{X}$ folytonos val. vektorváltozó, akkor X_i is folytonos, és a sűrűségfüggvénye:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_n$$



Forrás: wolfram.com,
Visualize Marginal Distributions

Marginális eloszlás, példa

Példa: Béla válaszidejének eloszlása,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\mathrm{d}x =$$

$$= \int_0^{\infty} \left(3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5}\right)\mathrm{d}x$$

$$= \frac{12}{5}e^{-12y}\left[-e^{-3x}\right]_0^{\infty} + \frac{24}{5}e^{-6y}\left[-e^{-6x}\right]_0^{\infty}$$

$$= \frac{12}{5}e^{-12y} + \frac{24}{5}e^{-6y}$$

Név: kevert exponenciális eloszlás

Val. változók együttes függetlensége

Definíció: Az $X_1, \ldots, X_n$ val. változók *(együttesen) függetlene*k, ha az

$$\{X_1 < x_1\}, \ldots, \{X_n < x_n\}$$

események függetlenek minden $x_1, \ldots, x_n \in \mathbb{R}$ esetén.

Megjegyzések:

- Példákat lásd még: két-változós, diszkrét esetnél.
- Tipikus példa: független kísérletek (számszerű) eredményei.
- Intuitívan “összefüggő” val. változókról is kiderülhet, hogy függetlenek.
- Ahogy események esetén is, függetlenek részhalmaza független.

Függetlenség, karakterizáció

Állítás: Az $X_1, \ldots, X_n$ val. változók pontosan akkor függetlenek, ha

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

tetszőleges $x_1, \ldots, x_n \in \mathbb{R}$ esetén.

Állítás: Az $X_1, \ldots, X_n$ folytonos val. változók pontosan akkor függetlenek, ha $(X_1, \ldots, X_n)$ folytonos val. vektorváltozó, és

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

tetszőleges $x_1, \ldots, x_n \in \mathbb{R}$ esetén.

Függetlenség, karakterizáció

Példa: Igaz-e, hogy Aladár válaszideje független Béla válaszidejétől?

$$f_{(X,Y)}(x, y) = 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5}$$

$$f_X(x) = \frac{3}{5}e^{-3x} + \frac{24}{5}e^{-6x} \qquad f_Y(y) = \frac{12}{5}e^{-12y} + \frac{24}{5}e^{-6y}$$

$$f_X(x) \cdot f_Y(y) \neq 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5}$$

Konvolúció

Emlékeztető:

Legyenek X és Y független val. változók, amik eloszlása egyenletes az $\{1, 2, 3, 4, 5, 6\}$ halmazon. Mi $X+Y$ eloszlása?

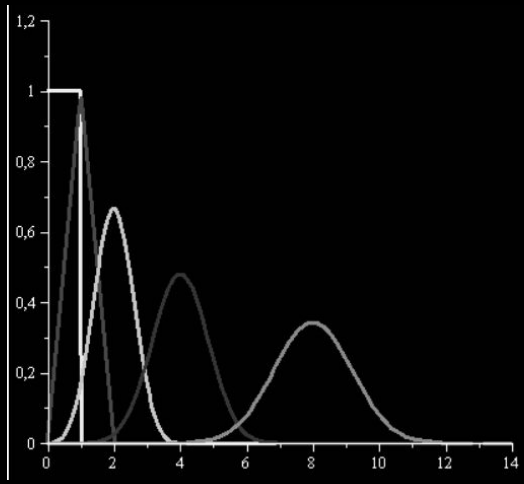
$$\{X+Y=7\} = \{X=6, Y=1\} \cup \{X=5, Y=2\} \cup$$
$$\cup \{X=4, Y=3\} \cup \{X=3, Y=4\}$$
$$\cup \{X=2, Y=5\} \cup \{X=1, Y=6\}$$

Konvolúció, példák

Definíció: Legyenek X és Y független val. változók.
Ekkor $X+Y$ eloszlását az X és Y *konvolúció*jának hívjuk.

Példák:

- A binomiális eloszlás független indikátorok összege.
- Ha a két változó Geo(p) eloszlású, akkor az összeg "negatív binomiális".
- Ha a két változó U(0;1) eloszlású, akkor az összeg "háromszög eloszlású".
- Ha több független U(0;1) eloszlású változót adunk össze: Irwin-Hall eloszlás.



Diszkrét konvolúció

Állítás: Legyenek X és Y független, diszkrét val. változók, amik értékei nemnegatív egészek. Ekkor

$$\mathbb{P}(X+Y=k)=\sum_{i=0}^{k}\mathbb{P}(X=i)\mathbb{P}(Y=k-i) \quad (\forall k\in\mathbb{N})$$

Megjegyzések:

- A jobb oldalon az eredeti változóknak csak az eloszlása van.
- Általánosítható egész számos esetre.
- Ha a jobb oldalra $\mathbb{P}(X=i, Y=k-i)$ kerül, akkor függetlenség nélkül is igaz.

Diszkrét konvolúció

Bizonyítás:

$$\mathbb{P}(X+Y=k)=\mathbb{P}\Bigg(\bigcup_{i+j=k}\{X=i\}\cap\{Y=j\}\Bigg)=$$

$$=\sum_{i=0}^{k}\mathbb{P}\Big(\{X=i\}\cap\{Y=k-i\}\Big)$$

$$=\sum_{i=0}^{k}\mathbb{P}(X=i)\mathbb{P}(Y=k-i)$$

Diszkrét konvolúció, példa

Példa: Legyenek X és Y független val. változók, $\mathrm{Pois}(\lambda)$ és $\mathrm{Pois}(\mu)$ eloszlásúak. Mi $X+Y$ eloszlása?

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

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

$$=e^{-(\lambda+\mu)}\frac{1}{k!}(\lambda+\mu)^k \Rightarrow X+Y\sim \mathrm{Pois}(\lambda+\mu)$$

Folytonos konvolúció

Állítás: Legyenek X és Y független, folytonos val. változók. Ekkor a

$$z \mapsto \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x$$

függvény az $X + Y$ sűrűségfüggvénye.

Megjegyzés:

- Ha a jobb oldalra $f_{(X,Y)}(x, z-x)$ kerül, akkor függetlenség nélkül is igaz.

Folytonos konvolúció, példa

Példa: Legyenek X és Y független, $\mathrm{Exp}(\lambda)$ eloszlású val. változók. Határozzuk meg $X+Y$ eloszlását.

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\mathrm{d}x = \qquad (z>0)$$

$$= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\mathrm{d}x = \lambda^2 e^{-\lambda z} \int_0^z 1\mathrm{d}x$$

$$= \lambda^2 e^{-\lambda z} z \Rightarrow \text{nem exponenciális}$$

Név: gamma-eloszlás

Érdekesség: végtelenül osztható

Megfigyelés:

- Poisson-eloszlású val. változók felírhatók tetszőleges, véges sok független, Poisson-eloszlású konvolúciójaként.
- Nem-nyilvánvaló, de egy geometriai eloszlású val. változó is felírható tetszőleges, véges sok független, azonos eloszlású konvolúciójaként.

Igaz-e ez mindig? Nem, az egyenletes eloszlás ellenpélda.

Név: végtelenül osztható.

Ez egy nevezetes eloszlás-osztály: csak ezek állnak elő független val. változók összegeinek határeloszlásaiként.

Köszönöm a figyelmet!