Valószínűségszámítás

2020. október 21.

Mészáros Szabolcs

Tárgyhonlap:
cs.bme.hu/valszam

Múltkori dia: exponenciális eloszlás

Definíció: Egy Z val. változó *exponenciális eloszlású* λ paraméterrel (ahol $\lambda > 0$ valós), ha

Miért nem ugyanaz a betű?
Mert Z val. változó, míg x szám.

$$f_Z(x) = \begin{cases} \lambda e^{-\lambda x} & \text{ha } x > 0, \\ 0 & \text{egyébként,} \end{cases}$$

$$F_Z(x) = \begin{cases} 1 - e^{-\lambda x} & \text{ha } x > 0, \\ 0 & \text{egyébként.} \end{cases}$$

Jelölés:

$$Z \sim \mathrm{Exp}(\lambda)$$

Példák:

- Idei első áramszünet időpontja,
- Mikor hív már fel XY?
- Sűrű, egymás utáni “kísérletek” közül az első siker időpontja.

Val. változók függetlensége

Definíció: Legyenek X és Y val. változók ugyanazon az $(\Omega, \mathcal{F}, \mathbb{P})$ eseménytéren. Ezek *függetlenek*, ha minden $x, y \in \mathbb{R}$ esetén

$$\{X < x\} \text{ és } \{Y < y\}$$

független események.

Megjegyzések:

- Eml.: $\mathbb{P}\big(\{X < x\} \cap \{Y < y\}\big) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$
- Értelmezés: a változók által meghatározott események függetlenek.
- Miért nem $\{X = x\}$ és $\{Y = y\}$? Hogy folytonosra is működjön.

Val. változók függetlensége

Példák:

1) X egy kockadobás eredménye, Y a ma leeső csapadék mennyisége.

2) $\Omega = \{1, 2, 3, \ldots, 11, 12\}$
X a szám hármas maradék, Y a négyes maradéka

3) Tegyük fel, hogy egy telefonközpontba egy nap N hívás érkezik, ahol $N \sim \mathrm{Pois}(\lambda)$. Az egyes hívások egymástól függetlenül p valószínűséggel gép által indított hívások. Jelölje X a gép által indított hívások számát, Y pedig az többi hívást.

Val. változók függetlensége

Állítás: Az X és Y diszkrét valószínűségi változók pontosan akkor függetlenek, ha minden $x \in \mathrm{Ran}(X)$ és $y \in \mathrm{Ran}(Y)$ esetén

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

Használata:

- Cáfolni vele a függetlenséget: könnyű. Elég egy x-y párt találni.
- Ellenőrizni a függetlenséget: számolós. Minden párra le kell ellenőrizni.

Biz ötlete: $\{X < x\}$ felbontható $\{X = x'\}$ alakú események diszjunkt uniójára.

Szorzat várható értéke, állítás

Állítás: Ha X és Y független valószínűségi változók, és léteznek az $\mathbb{E}(X)$, $\mathbb{E}(Y)$ és $\mathbb{E}(XY)$ várható értékek, akkor

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Megjegyzés:

- A val. változók nem feltétlenül diszkrétek.
- Függetlenség nélkül ez nem igaz, pl. X értéke azonos eséllyel 1 vagy -1.
- A fenti egyenlet nem elég a függetlenséghez.
  (De ha $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$
  minden f, g nemnegatív valós függvényre, abból következik.)

Szorzat várható értéke, indikátor

Lemma: Legyenek A és B független események. Ekkor

$$\mathbb{E}(\mathbf{1}_A \mathbf{1}_B) = \mathbb{E}(\mathbf{1}_A)\mathbb{E}(\mathbf{1}_B)$$

Biz:

$$\mathbb{E}(\mathbf{1}_A \mathbf{1}_B) = \mathbb{E}(\mathbf{1}_{A\cap B}) = \mathbb{P}(A \cap B) =$$

$$= \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}(\mathbf{1}_A)\mathbb{E}(\mathbf{1}_B)$$

Kérdés: miért segít ez az általánosabb esetben?

Mert az egyszerű val. változók felbonthatók indikátorok összegére.

$$X = \sum_{k \in \mathrm{Ran}(X)} k \cdot \mathbf{1}_{\{X=k\}} \qquad Y = \sum_{l \in \mathrm{Ran}(Y)} l \cdot \mathbf{1}_{\{Y=l\}}$$

Szorzat várható értéke

Biz.:

$$\mathbb{E}(XY) = \mathbb{E}\left(\sum_{k\in\mathrm{Ran}(X)} k\cdot\mathbf{1}_{\{X=k\}} \sum_{l\in\mathrm{Ran}(Y)} l\cdot\mathbf{1}_{\{Y=l\}}\right) =$$

$$= \sum_{k\in\mathrm{Ran}(X)} \sum_{l\in\mathrm{Ran}(Y)} k\cdot l\cdot\mathbb{E}\left(\mathbf{1}_{\{X=k\}}\mathbf{1}_{\{Y=l\}}\right)$$

$$= \sum_{k\in\mathrm{Ran}(X)} \sum_{l\in\mathrm{Ran}(Y)} k\cdot l\cdot\mathbb{E}\left(\mathbf{1}_{\{X=k\}}\right)\mathbb{E}\left(\mathbf{1}_{\{Y=l\}}\right)$$

$$= \mathbb{E}(X)\mathbb{E}(Y)$$

Diszkrét együttes eloszlás, példa

Kérdés: A fent használt $\mathbb{P}(\{X = k\} \cap \{Y = l\})$ valószínűségeket, nem lehetne valahogy koncepciózusan, “együtt kezelni”?

Példa: $\mathrm{Ran}(X) = \{2, 3, 5\}$ $\mathrm{Ran}(Y) = \{0, 1, 2\}$

$Y \backslash X$	2	3	5	
0	0,05	0,15	0,1	0,3
1	0,1	0,2	0,1	0,4
2	0,05	0,2	0,05	0,3
	0,2	0,55	0,25	

$$\mathbb{P}(X = 5) \cdot \mathbb{P}(Y = 0) = 0{,}075$$

$$\mathbb{P}(X = 5,\ Y = 0) = 0{,}1$$

Kérdések: Független-e X és Y? Mennyi $\mathbb{E}(XY)$?

Diszkrét együttes eloszlás, def.

Definíció: Legyenek X és Y egyszerű valószínűségi változók.

- *Együttes eloszlásuk:*

$$\mathbb{P}(X = k, Y = l) \quad (k \in \mathrm{Ran}(X), l \in \mathrm{Ran}(Y))$$

- *Marginális eloszlások:* az X és Y eloszlása külön-külön.

Marginális kiszámolása:

$$\mathbb{P}(X = k) = \sum_{l \in \mathrm{Ran}(Y)} \mathbb{P}(X = k,\ Y = l)$$

Diszkrét együttes eloszlás, példa

Példa (folytatás): Ha X és Y függetlenek lettek volna, akkor $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Mit csináljunk amikor ez nem teljesül?

Ötlet: XY is csak egy szokásos val. változó, használhatjuk a definíciót.

$$\mathrm{Ran}(XY) = \{k \cdot l \mid k \in \mathrm{Ran}(X),\ l \in \mathrm{Ran}(Y)\}$$

$$\mathbb{E}(XY) = \sum_{m \in \mathrm{Ran}(XY)} m \cdot \mathbb{P}\Big(\bigcup_{\substack{k \in \mathrm{Ran}(X) \\ l \in \mathrm{Ran}(Y) \\ k \cdot l = m}} \{X = k,\ Y = l\} \Big) =$$

$$= \sum_{k \in \mathrm{Ran}(X)} \sum_{l \in \mathrm{Ran}(Y)} k \cdot l \cdot \mathbb{P}(X = k,\ Y = l) =$$

Diszkrét együttes eloszlás, példa

$$
\begin{aligned}
&= 0 \cdot 0{,}05 + 0 \cdot 0{,}15 + 0 \cdot 0{,}1 \\
&\quad +2 \cdot 0{,}1 + 3 \cdot 0{,}2 + 5 \cdot 0{,}1 \\
&\quad +4 \cdot 0{,}05 + 6 \cdot 0{,}2 + 10 \cdot 0{,}05 \\
&= 3{,}2
\end{aligned}
$$

$Y \backslash X$	2	3	5	
0	0,05	0,15	0,1	0,3
1	0,1	0,2	0,1	0,4
2	0,05	0,2	0,05	0,3
	0,2	0,55	0,25	

Megjegyzés: A val. változók nem-függetlensége azzal is megmutatható, hogy

$$\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$$

Kovariancia, motiváció

Kérdés: hogyan mérjük a függetlenséget / nem függetlenséget?

	Nagysága	**Viszony mérése**
Esemény	valószínűség	feltételes valószínűség
Valószínűségi változó	várható érték	???

- Feltételes várható érték?

  Igen, de csak később.

- Kovariancia.

Kovariancia, definíció

Definíció: Az X és Y valószínűségi változók *kovarianciá*ja:

$$\mathrm{cov}(X,Y) = \mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big)$$

A továbbiakban mindig feltesszük, hogy $\mathrm{cov}(X,Y)$ létezik.

Állítás:

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Bizonyítás: $\mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big) =$

$$= \mathbb{E}(XY) - \mathbb{E}\Big(\mathbb{E}(X)Y\Big) - \mathbb{E}\Big(X\mathbb{E}(Y)\Big) + \mathbb{E}\Big(\mathbb{E}(X)\mathbb{E}(Y)\Big) =$$

Kovariancia, tulajdonságok

Biz. (folytatás): $\operatorname{cov}(X, Y) = \ldots$

$$= \mathbb{E}(XY) + (-1 - 1 + 1)\mathbb{E}(X)\mathbb{E}(Y)$$

$$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Következmény:

1. Ha valamelyikük konstans, akkor a kovariancia nulla.
2. Ha függetlenek, akkor a kovariancia nulla.
3. Attól, hogy a kovariancia nulla, még nem feltétlenül lesznek függetlenek.

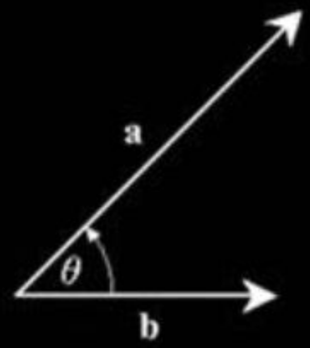
Kovariancia, példa

Példa (folytatás):

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) =$$
$$= 3{,}2 - 3{,}8 \cdot 1 = -0{,}6$$

Geometriai kép:

A kovariancia hasonlóan működik a skaláris szorzathoz. A merőlegesség szerepét, a *"korrelálatlanság"* veszi át, azaz amikor a kovariancia 0 (ami a függetlenségnél gyengébb feltétel).



Szórásnégyzet, definíció

Definíció: Egy X valószínűségi változó szórásnégyzete

$$\operatorname{cov}(X, X) = \mathbb{E}\left((X - \mathbb{E}X)^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Jelölés: $\mathbb{D}^2(X)$

Értelmezés:

0) Önmagával való együtt változás? Annak így nincs sok értelme.
1) “átlagos négyzetes eltérés” (a várható értéktől)
2) Ha a kovariancia a skalár szorzat analógiája, akkor a szórásnégyzet a hossznégyzeté.

Szórásnégyzet, tulajdonságok

Definíció: X *szórása* $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$

Kérdés: Miért nem a következővel mérjük a várható értéktől való eltérést:

$$\mathbb{E}\Big(\big|X - \mathbb{E}(X)\big|\Big)$$

Állítás: Ha X és Y függetlenek, akkor

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$$

Szórásnégyzet, tulajdonságok

Bizonyítás:

$$
\begin{aligned}
\mathbb{D}^2(X+Y) = \mathbb{E}\big((X+Y)^2\big) - \big(\mathbb{E}(X+Y)\big)^2 &= \\
= \mathbb{E}(X^2)+\mathbb{E}(Y^2)+2\mathbb{E}(XY)-\mathbb{E}(X)^2& \\
-\mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y)& \\
= \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\mathrm{cov}(X,Y)&
\end{aligned}
$$

Mivel a változók függetlenek, ezért a kovariancia 0 és az állítás következik.

Állítás:

$$\mathbb{D}(X+c) = \mathbb{D}(X) \qquad \mathbb{D}(cX) = |c|\mathbb{D}(X) \qquad (\forall c \in \mathbb{R})$$

Szórásnégyzet, példák

1) kockadobás szórásnégyzete: $\mathbb{E}(K^2) = \dfrac{91}{6} \quad \mathbb{E}(K) = \dfrac{7}{2}$

$$\Rightarrow \mathbb{D}^2(K) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 \approx 2{,}9167$$

2) indikátor val. változó szórásnégyzete:

$$\mathbb{D}^2(\mathbf{1}_A) = \mathbb{E}(\mathbf{1}_A^2) - \mathbb{E}(\mathbf{1}_A)^2 = \mathbb{E}(\mathbf{1}_A) - \mathbb{E}(\mathbf{1}_A)^2 =$$

$$= \mathbb{P}(A) - \mathbb{P}(A)^2 = p(1-p)$$

Szórásnégyzet, példák

3) binomiális eloszlás szórásnégyzete: $X \sim B(n;p)$

$$X = \mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}$$

$$\mathbb{D}^2(X) = \mathbb{D}^2(\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}) =$$

$$= \mathbb{D}^2(\mathbf{1}_{A_1}) + \cdots + \mathbb{D}^2(\mathbf{1}_{A_n}) = np(1-p)$$

4) geometriai eloszlás szórásnégyzete: $T \sim \mathrm{Geo}(p)$

$$\mathbb{D}^2(T) = \mathbb{E}(T^2) - \mathbb{E}(T)^2 =$$

$$= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

Szórásnégyzet, példák

5) Poisson-eloszlás szórásnégyzete: $Y \sim \mathrm{Pois}(\lambda)$

$$\mathbb{D}^2(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 =$$

$$= \mathbb{E}(Y^2 - Y) + \mathbb{E}(Y) - \mathbb{E}(Y)^2$$

$$= \sum_{k=0}^{\infty} (k^2 - k) \frac{\lambda^k}{k!} e^{-\lambda} + \lambda - \lambda^2$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda - \lambda^2 = \lambda$$

Korreláció, definíció

Kérdés: a példában kapott $\operatorname{cov}(X, Y) = -0{,}6$ mit jelent a val. változókra nézve? Ők most mennyire nem függetlenek?

Definíció: Legyenek X és Y olyan valószínűségi változók, amikre $\operatorname{cov}(X, Y)$, $\mathbb{D}(X)$ és $\mathbb{D}(Y)$ létezik és véges. Ekkor a korrelációjuk:

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}$$

Értelmezés:

A geometriai analógiában ez a bezárt szög koszinusza

Korreláció, tulajdonságok

Állítás:

- $-1 \leq \operatorname{corr}(X, Y) \leq 1$
- Ha a korreláció +1, akkor $Y = aX + b$, ahol $a > 0$.
- Ha a korreláció -1, akkor $Y = aX + b$, ahol $a < 0$.

Példa (korábbi):

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)} = \frac{-0{,}6}{\sqrt{9{,}2} \cdot \sqrt{0{,}6}} \approx -0{,}2554$$

Kitérő: statisztikai korreláció

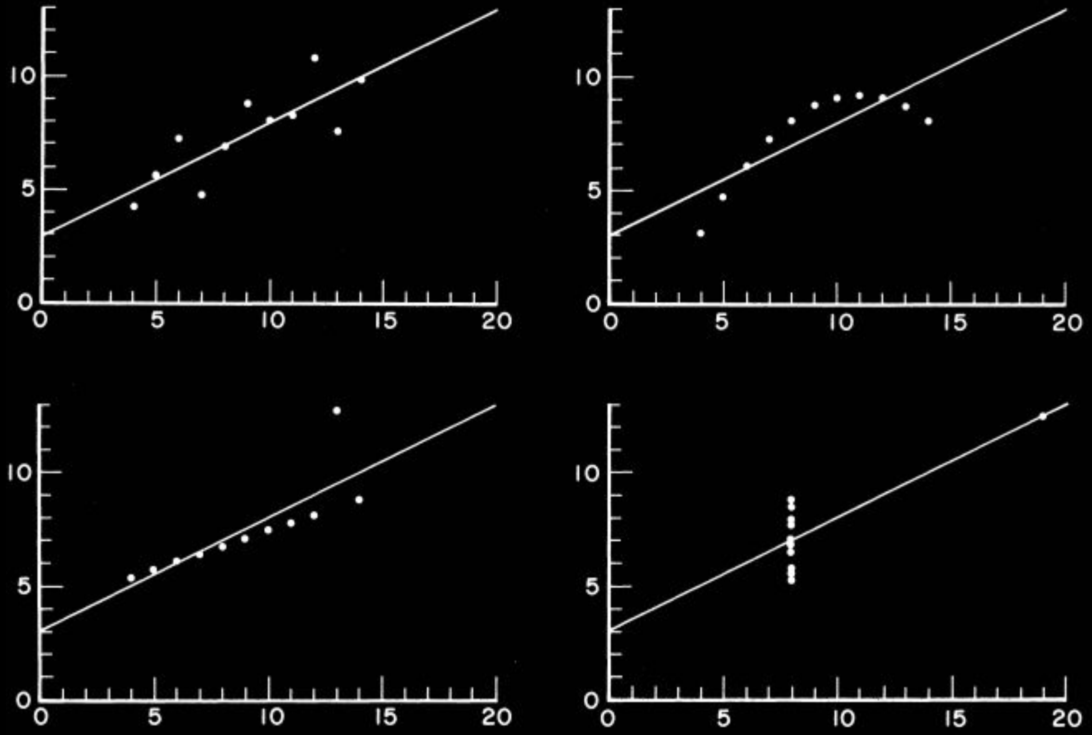
Vizsgált mennyiségek: X, Y

Minta: $X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n$

Minta realizációja: adott $\omega \in \Omega$ kimenetelen a minta értékei.

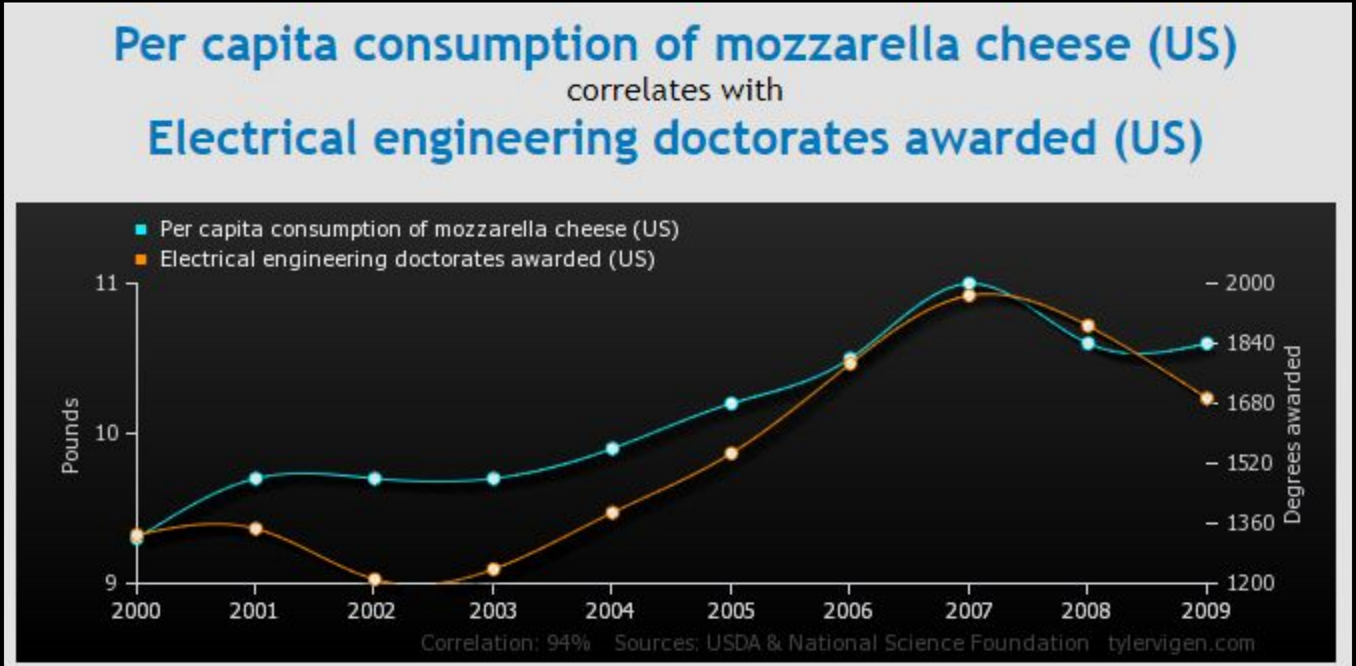
Pearson-féle korreláció: Vegyük a minta realizációját (ez n db számpár). Tegyünk úgy, mintha ez az n számpár n különböző esetben a vizsgált mennyiségek értékeit jelölné, amely esetek egyforma valószínűséggel következhetnek be. Számoljunk eszerint korrelációt. Ezzel egy közelítést kapunk a vizsgált mennyiségek (elméleti) korrelációjára.

Kitérő: Anscombe négyese



Forrás: www.lithoguru.com/scientist/statistics/Anscombe_Graphs%20in%20Statistical%20Analysis_1973.pdf

Hamis korrelációk



Forrás: www.tylervigen.com/spurious-correlations

Továbbá, “a korreláció nem implikált kauzalitást.”

Köszönöm a figyelmet!