Várható érték, szórás, momentumok

$P(X \in A)$ tetsz. A esetén ki tudjuk számolni.

Az eloszlás teljes leírást ad a valószínűségi változóról.

A gyakorlatban az eloszlást nehéz becsülni → ezért így csak részleges leírásai a valószínűségi változónak

Ha X diszkrét, akkor várható érték

$$EX = \sum_i x_i P(X = x_i)$$

az értékeket súlyozzuk a valószínűséggel.

$x_1, x_2, \dots$, az X értékei

Ha X folytonos, akkor egy sűrűségfgv van

↓ a várható érték akkor

$$EX = \int_{-\infty}^{\infty} x f(x)\, dx$$

A nagy számok (erős) törvénye:

Ha $X_1, X_2, \dots$ független, azonos eloszlásúak, akkor az átlag: $\frac{X_1 + \dots + X_n}{n} \to EX_1$, ha $n \to \infty$ 1 valószínűséggel.

/vagyis ha n nagy, akkor tetszőlegesen becsülhető lesz [illegible]

Így pl. ha feldobunk egy dobókockát sokszor, akkor:

X = Dobókocka eredménye

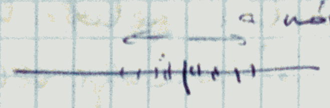
$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

Szórás, szórásnégyzet

Egy X valószínűségi változó szórásnégyzete

$$D^2(X) = E\Big(\underbrace{(X - EX)^2}_{}\Big)$$

az átlagtól való eltérés négyzete, s ennek vesszük az átlagát majd a szóródás mértékét mutatja meg



Ex

A szórás pedig:

$$D(X) = \sqrt{D^2(X)}$$

Kiszámoláshoz:

$$D^2(X) = E\left(X^2 - 2X\,E(X) + E^2(X)\right) =$$

$$= E(X^2) - \left(E(X)\right)^2$$

Momentum: Egy X valószínűségi változó n. momentuma

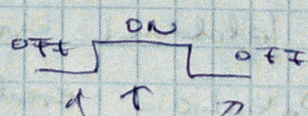
$$E(X^n) = \sum x_i^n P(X = x_i)$$

diszkrét esetben

ill.

folytonos esetben: $\int x^n f(x)\,dx$

Példa: A felhasználók internetezési szokásait on/off folyamattal jellemezzük: ha nem generál forgalmat, akkor 0, ha bekapcsol, akkor 1.



ezeknek a hossza véletlen hosszú.

Egy on periódus hossza vastagfarkú → Pareto-eloszlású

↓

legyen ez X

X sűrűségfüggvénye (eltekintünk most attól, hogy ez valójában egy diszkrét valószínűségi változó):

$$f(x) = \frac{1}{x^{2,5}} \quad , \text{ha } x \geq 5$$

Várhatóan (vagy átlagosan) mekkora egy on periódus hossza? Határozzuk meg a szórását is.

$$EX = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{5}^{\infty} x \cdot \frac{1}{x^{2,5}}\, dx =$$

$$= \left[\frac{x^{-0,5}}{-0,5} \right]_5^{\infty} = 0 + \frac{2}{\sqrt{5}}$$

szórásnégyzet: $D^2(X) = E(X^2) - \underbrace{(E(X))^2}_{\left(\frac{2}{\sqrt{5}}\right)^2}$

$$E(X^2) = \int_5^{\infty} x^2 \cdot \frac{1}{x^{2,5}}\, dx = \int x^2 \frac{1}{x^{2,5}}\, dx = \left[\frac{\sqrt{x}}{\frac{1}{2}} \right]_5^{\infty} \to \infty$$

ugyanis $E(X^2) \to \infty$, így

$D^2(X) \to \infty$ nem létezik a szórásnégyzete.

Pareto-eloszlás (vastagfarkú eloszlás)

↓ hatványrendben cseng le a farka.

így $P(X > x) \approx \dfrac{c}{x^{1.5}}$

A nagy értékek is valószínűek

(nem úgy mint az exp. eloszlásnál:

$Y = \exp(\lambda)$

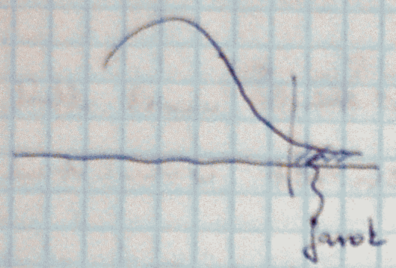
↓

ill $P(Y > x) = e^{-\lambda x}$

↑

exp. csökken le a valószínűség

Exp egy vékonyfarkú.



pl. ha veszünk sok homokszemet, akkor egy homokszem mérete Pareto-eloszlást mutat.

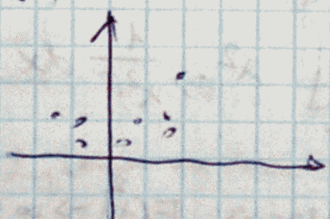
Többdimenziós valószínűségi változók

2 dimenzióssal foglalkozunk.

Legyen (X, Y) egy kétdimenziós valószínűségi változó.

A legfontosabb leírás itt is az eloszlás.

A diszkrét értékű valószínűségi változó értékei felsorolhatók: (x_1, y_1), (x_2, y_2)



A folytonos értékű valószínűségi változó:

bármilyen értéket felvehet ebben.

Kell:

$$P((X,Y)\in A)$$, ahol A a sík egy tetszőleges részhalmaza

Diszkrét esetben:

eloszlás: $(x_1, y_1), (x_2, y_2) \ldots$

$P(X=x_1, Y=y_1)$, $P(X=x_2, Y=y_2) \ldots$

felsorolással.

$\forall A$ $$P((X,Y)\in A) = \sum_{(x_i, y_i)\in A} P((X,Y)=(x_i, y_i))$$

- halmazba esés valószínűsége

Folytonos esetben

Van egy sűrűségfgv. együttes sűrűségfgv.

(X,Y) együttes sűrűségfgv-e $f(x,y)\geq 0$

$$P((X,Y)\in A) = \iint_A f(x,y)\,dx\,dy$$

az A halmazon való integrálás

X és Y egydimenziósak.

Ha csak az egyikre kérdezünk az eloszlást →

→ marginális eloszlásuk.

<u>Marginális eloszlások</u>

Ha (X,Y) egy kétdimenziós valószínűségi változó, akkor X eloszlása az első változó szerinti marginális eloszlása. Y-ra hasonlóan kapjuk (a második változó szerinti marginális eloszlás).

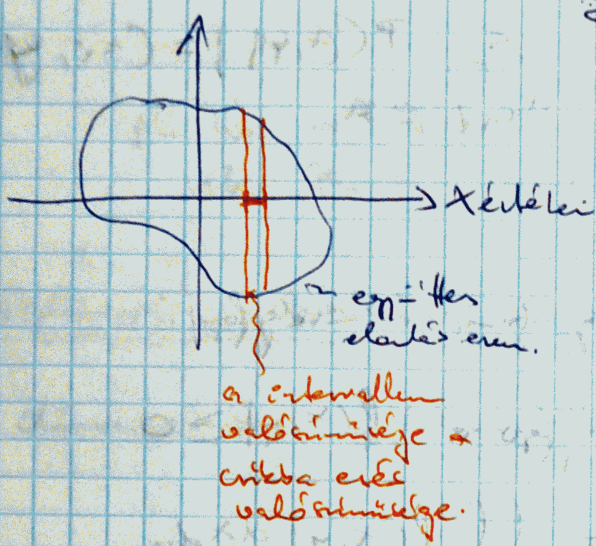
A marginális eloszlások kiszámolásával lehet eldönteni, hogy a valváltozók függetlenek-e.

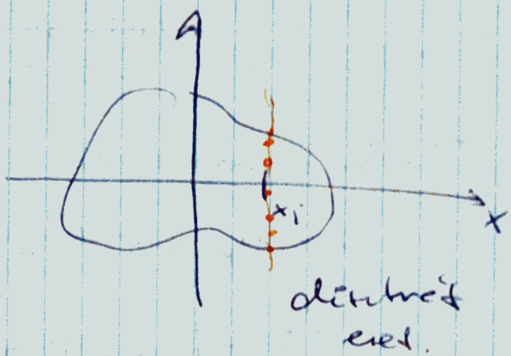
X sűrűségfgv-e:

$$f_X(x) = f^{(X)} = \int_{-\infty}^{\infty} f(x,y)\,dy$$ ha folytonos

Diszkrét esetre:

$$P(X=x_i) = \sum_{y_j} P(X=x_i, Y=y_j)$$





Y sűrűségfüggvénye hasonlóan.

__Példa:__

1, Profi ejtőernyős bele kell ugrani egy 5m sugarú körbe, ezen belül egyenletes eloszlással érkezik

↳ a tartományba esés valószínűsége csak a tartomány területétől függ.

Határozd meg, hogy az X és Y koordinátának mi az eloszlása (marginális eloszlás)

Független-e X az Y-tól?

Az együttes sűrűségfüggvény:

egyenletes eloszlás:

$A \subset D$ (legyen D az 5m sugarú körlap) vagy $P((X,Y)\in A) = \dfrac{\text{terület}(A)}{\text{terület}(D)}$

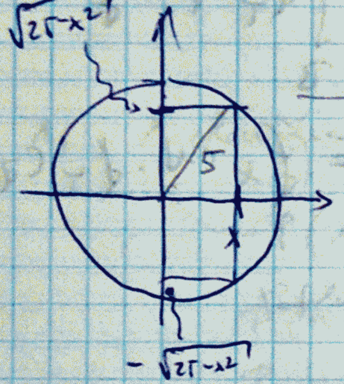
$\Updownarrow$

A sűrűségfgv. értékének $\forall$ pontban ugyanannyinak kell lennie, de integrálja 1 a D-n.

$\Updownarrow$

$$f(x,y) = \frac{1}{\text{terület}(D)} = \frac{1}{25\pi},\ \text{ha } (x,y)\in D$$

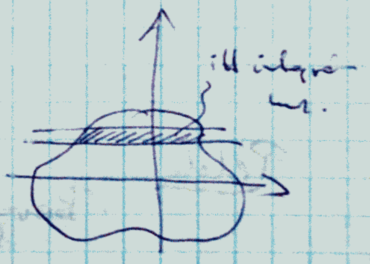
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy =$$



$$= \int_{-\sqrt{25-x^2}}^{\sqrt{25-x^2}} \frac{1}{25\pi}\,dy = \frac{2\sqrt{25-x^2}}{25\pi}$$

$$-5 \le x \le 5$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx =$$



$$= \frac{2\sqrt{25-y^2}}{25\pi}$$

adódik, mert minden szimmetrikus

$$-5 \le y \le 5$$

Független-e X az Y-tól? ~~[illegible]~~

(X, Y) esetén X ftlen. Y-tól, ha

• diszkrét esetben: $\forall x_i, y_j$-re $P((X,Y)=(x_i,y_j))$

a marginálisok szorzata $\longrightarrow$ $= P(X=x_i)\cdot P(Y=y_j)$

~~független~~ ez ugyanaz, ha: $P(X=x_i$ és $Y=y_j)$

ha • ha folytonos

↓ halmazokra esési valószínűségeket számolunk

$$P(X\in A \text{ és } Y\in B) = P(X\in A)\cdot P(Y\in B)$$

$\forall A, B \in \mathbb{R}$ esetén.

A marginális sűrűségfüggvényekkel megfogalmazva ezt:

$$\int_A f_X(x)\,dx \cdot \int_B f_Y(y)\,dy$$

Vagyis $\forall (x,y)$-ra $f(x,y) = f_X(x)\cdot f_Y(y)$

A két koordináta nem lesz fgtl. egymástól

$$\frac{1}{2T\pi} \neq \frac{2\sqrt{2T-x^2}}{2\pi T}\cdot\frac{2\sqrt{2T-y^2}}{2T\pi}$$

vagyis nem függetlenek.

Példa

. Bergengóciában ↓ Y

	lehet: 1 gyerek	2 gyerek	3 gyerek
X → 0 autó	0,05	0,05	0,05
1 autó	0,15	0,25	0,2
2 autó	0	0,05	0,15
3 autó	0	0	0,05

ekkora valószínűséggel van a családban 1 autó és 2 gyerek.

Legyen X a gyerekszám
Y az autószám.

1., Határozzuk meg a marginális eloszlásokat!

2., Számoljuk ki annak a valószínűségét, hogy 1 autó van a családban, feltéve, hogy 2 gyerek.

3., Határozzuk meg Y feltételes eloszlását X-re nézve.

4., Határozzuk meg Y feltételes várható értékét X-re nézve

5., Határozzuk meg Y ~~teljes~~ várható értékét a teljes várható érték ~~tétel~~ segítségével.

1, X eloszlása:

$P(X=1) = 0{,}2$ (az oszlopösszeg)

$P(X=2) = 0{,}35$

$P(X=3) = 0{,}45$

← X marginális eloszlása

Y eloszlása:

$P(Y=0) = 0{,}15$ (sorösszeg)

$P(Y=1) = 0{,}6$

$P(Y=2) = 0{,}2$

$P(Y=3) = 0{,}05$

2., $$P(Y=1 \mid X=2) = \frac{P(X=2, Y=1)}{P(X=2)} = \frac{0{,}25}{0{,}35}$$

3., ~~Rögzít~~ Minden egyes rögzített i-re $(i=1,2,3)$ meg kell adnunk $P(Y=j \mid X=i)$ valószínűségeket $j=0,1,2,3$-ra:

pl $i = 2$ -re:

$$P(Y=0 \mid X=2) = ? = \frac{0{,}05}{0{,}35}$$

$$P(Y=1 \mid X=2) = ? = \frac{0{,}25}{0{,}35}$$

$$P(Y=2 \mid X=2) = ? = \frac{0{,}05}{0{,}35}$$

$$P(Y=3 \mid X=2) = ? = 0$$

$\Sigma = 1$ kell, hogy legyen.

Ezt minden i-re meg kell nézni.

4.) Y feltételes várható értéke X-re nézve:

↑ ez egy függvény

$E(Y \mid X=i)$ ahol $i = 1, 2, 3$.

pl $E(Y \mid X=2)$ $(i=2)$

↑ ez az előző feladatban kapott eloszlás várható értéke.

vagyis $E(Y \mid X=2) =$

$$= \sum_{j=0}^{3} j \cdot P(Y=j \mid X=2) =$$

$$= 1 \cdot \frac{0{,}25}{0{,}35} + 2 \cdot \frac{0{,}05}{0{,}35}$$

5.) $EY = \sum_i E(Y \mid X=i) \, P(X=i)$ a teljes valószínűség tételével rokon.

<u>Példa:</u> Bergengóc írószálban szennyeződések tömény van.

Annak a valószínűsége, hogy k gramm van 2^{-k+1} valószínűségű $k=2,3,4,\dots$

Ha k mg van az írószálban, akkor az író élettartama (órában) geometriai eloszlású $\frac{1}{k}$ paraméterrel.

Ha véletlenül kiválasztunk egy írót, akkor annak mennyi a várható élettartama?

X = A szennyeződés ~~mérete~~ nagysága

Y = az író élettartama

Kérdés: $EY=?$

Amit tudunk: Y eloszlása feltéve $X=k$-t geometriai $\frac{1}{k}$ paraméterrel

$\Downarrow$

Ha Z geometriai eloszlású p paraméterrel, akkor várható értéke $EZ = \frac{1}{p}$

egymás után teljesül és Z az első sikeres száma

Egy esemény bekövetkezését teljesül, p valószínűségű az esemény.

Z az első sikeres teszt száma

$\Downarrow$ Z geometriai eloszlású p paraméterrel.

Vagyis $E(Y|X=k) = k$

A teljes várható érték tételt használva:

$$EY = \sum_i E(Y \mid X=i)\, P(X=i)$$

ami most:

$$EY = \sum_{k=2}^{\infty} E(Y \mid X=i)\, P(X=i) =$$

$$= \sum_{k=2}^{\infty} k \cdot 2 \cdot 2^{-k} = 3.$$

Várható-érték vektor és kovariancia mátrix

(X,Y) egy kétdimenziós valószínűségi változó

Ekkor várható érték vektora: (EX, EY)

A szórásnak megfelelő mennyiség $\to$ kovariancia mátrix

(X,Y) kovariancia mátrixa

$\downarrow$

ehhez kell tudni X és Y kovarianciáját

$$\operatorname{cov}(X,Y) = E\big((X-EX)(Y-EY)\big)$$

szimmetrikus $\to$

$$\operatorname{cov}(X,X) = E\big((X-EX)^2\big) = D^2(X)$$

A kovarianciamátrix:

$$\Sigma_{(X,Y)} = \begin{bmatrix} \operatorname{cov}(X,X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{cov}(Y,Y) \end{bmatrix}$$

A kovariancia lényege: $\dfrac{\operatorname{cov}(X,Y)}{D(X)\cdot D(Y)} = \operatorname{corr}(X,Y)$

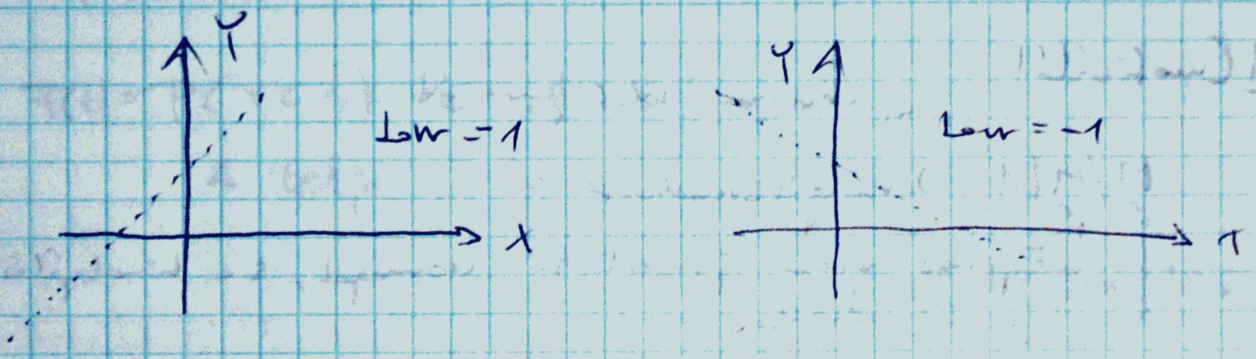
$\uparrow$ X és Y korrelációja

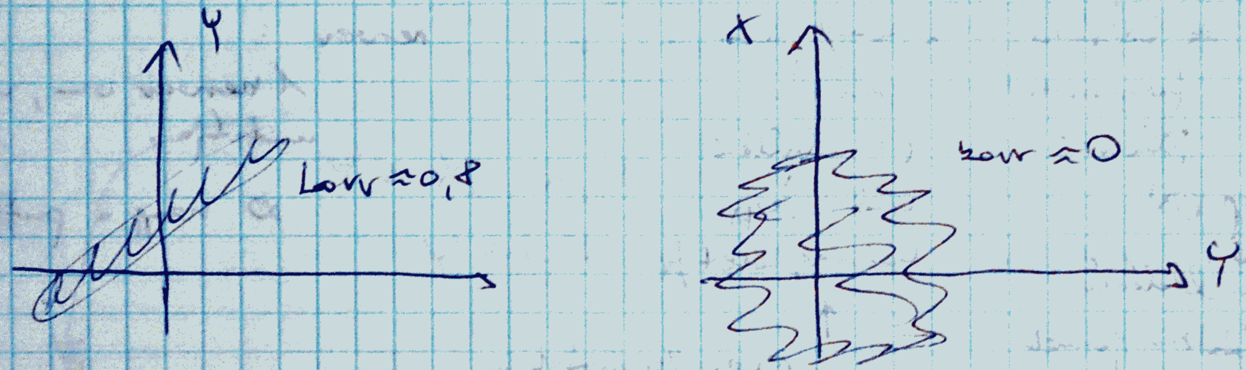
A korreláció -1 és 1 között van, X-nek az Y-tól való lineáris függését méri. úgy, hogy

ha $X = aY + b$ és $a > 0$, akkor

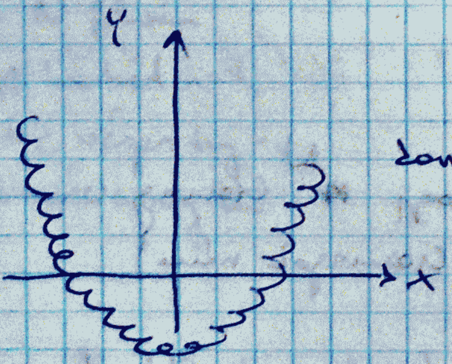
$korr = 1$

ha $a < 0$, akkor
$korr = -1$.





Csak a lineáris függést méri. pl.



korr = 0
pedig látható,
hogy erős a
függés. ← csak lin. függést tud
megadni.