

Valószínűesszámítás és statisztika – Többdimenziós eloszlások és regresszió

Tartalom

- ▶ Valószínűségi vektorváltozók, együttesen folytonos többdimenziós eloszlások, feltételes eloszlások
- ▶ Momentumok, kovariancia és korreláció
- ▶ Lineáris regresszió
- ▶ Feltételes várható érték/regresszió, tulajdonságai, teljes várható érték tétele
- ▶ Feltételes valószínűség és teljes valószínűség tétele folytonos esetben
- ▶ Többdimenziós normális eloszlás, polinomiális eloszlás

Előismeretek: Differenciál- és integrálszámítás több dimenzióban: parciális deriváltak, totális differenciálhatóság, integrálás normál tartományon.

Vektorok skaláris szorzása, mátrix determinánsa, inverze(, rangja).

A valszám előadás eddigi részei.

Valószínűségi vektorváltozók

Eddig: egy darab folytonos val. változó eloszlásával foglalkoztunk. Eloszlásfüggvény, sűrűségfüggvény, várható érték, eloszlástrafók...

Most: **több** folytonos val. változó együttes eloszlása (hasonlóan a diszkrét együttes eloszlásához) $\rightarrow$ valószínűségi vektorváltozók.

*A következő n -dimenziós definícióktól/állításoktól nem kell megijedni. Néha szükségünk lesz az általános n esetére, de a gyakorlatban legtöbbször **több=2**, és a kétdimenziós eset egyszerűbb.*

(7.1.1. Definíció.) Legyenek $X_1, \dots, X_n$ ugyanazon $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn definiált valószínűségi változók valamilyen n pozitív egész esetén. Ekkor az

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

függvényt valószínűségi vektorváltozónak hívjuk.

Az $\underline{X}$ (együttes) eloszlásfüggvénye az

$$F_{\underline{X}}: \mathbb{R}^n \rightarrow [0, 1], F_{\underline{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 < x_1, \dots, X_n < x_n)$$

vektorváltozós, skalárértékű függvény.

Az együttes eloszlásfüggvény tulajdonságai

Emlékeztető: egy darab X valószínűségi változó eloszlásfüggvényének tulajdonságai:

- ▶ $\lim_{x \rightarrow -\infty} F_X(x) = 0,$
- ▶ $\lim_{x \rightarrow \infty} F_X(x) = 1,$
- ▶ F_X balról folytonos, vagyis minden $x_1 \in \mathbb{R}$ -re $\lim_{x \uparrow x_1} F_X(x) = F_X(x_1),$
- ▶ F_X monoton növény, vagyis ha $x, y \in \mathbb{R}$ és $x \leq y$, akkor $F_X(x) \leq F_X(y).$

Mik ezeknek a tulajdonságoknak az analógjai egy valószínűségi vektorváltozó esetén?

Az együttes eloszlásfüggvény tulajdonságai

Ha $F_{\underline{X}}$ egy $\underline{X}$ valószínűségi vektorváltozó együttes eloszlásfüggvénye, akkor

- ▶ $\lim_{x_1 \rightarrow -\infty, x_2 \rightarrow -\infty, \dots, x_n \rightarrow -\infty} F_{\underline{X}}(x_1, \dots, x_n) = 0,$
- ▶ $\lim_{x_1 \rightarrow \infty, x_2 \rightarrow \infty, \dots, x_n \rightarrow \infty} F_{\underline{X}}(x_1, \dots, x_n) = 1,$
- ▶ $F_{\underline{X}}$ „minden változóban (parciálisan) balról folytonos”, azaz bármely $i \in \{1, \dots, n\}$ -re, $x_i \in \mathbb{R}$ -re és rögzített $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ valós számokra

$$\lim_{x \uparrow x_i} F_{\underline{X}}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) = F_{\underline{X}}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n).$$

- ▶ Ha egy kivételével minden változót rögzítünk, akkor (parciálisan) monoton növvő. Ugyanis legyen $i \in \{1, \dots, n\}$ és $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \in \mathbb{R}$ adott. Ekkor bármely $x \leq y$ valós számok esetén

$$F_{\underline{X}}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \leq F_{\underline{X}}(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n).$$

Az együttes eloszlásfüggvény tulajdonságai

Diszkrét eset $\rightarrow$ együttes eloszlás táblázata, peremeloszlások.

$Y \backslash X$	0	1	p_Y
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

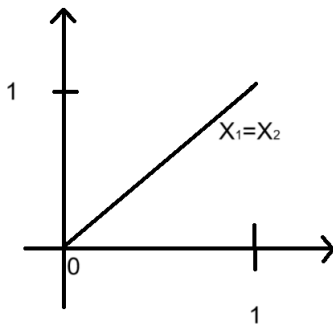
n dimenzióban ugyanez, csak n -dimenziós együttes súlyfüggvénnyel (táblázatot nehezebb rajzolni).

Egy változó peremeloszlása kiszámításához **az összes többi** változó szerint összegezni kell.

Folytonos eset?

Folytonos eset $\rightarrow$ mit értünk egyáltalán folytonos eset alatt?

Példa ($n = 2$): A síkon az $y = x$ egyenes $(0, 0)$ és $(1, 1)$ közé eső szakaszán választunk találomra (egyenletes eloszlással) egy pontot, jelölje ezt (X_1, X_2) .



Nyilvánvalóan $\mathbb{P}(X_2 = X_1) = 1$. Ekkor X_1 és X_2 is egyenletes eloszlású a $[0, 1]$ intervallumon, tehát külön-külön X_1 és X_2 is folytonos val. változó. Az (X_1, X_2) valószínűségi vektorváltozó

Folytonos eset?

Folytonos eset $\rightarrow$ mit értünk egyáltalán folytonos eset alatt?

Példa ($n = 2$): A síkon az $y = x$ egyenes $(0, 0)$ és $(1, 1)$ közé eső szakaszán választunk találmra (egyenletes eloszlással) egy pontot, jelölje ezt (X_1, X_2) .

Nyilvánvalóan $\mathbb{P}(X_2 = X_1) = 1$. Ekkor X_1 és X_2 is egyenletes eloszlású a $[0, 1]$ intervallumon, tehát külön-külön X_1 és X_2 is folytonos val. változó. Az (X_1, X_2) valószínűségi vektorváltozó értékkészlete viszont az $\{(x, y) \in \mathbb{R}^2 | y = x\} \leq \mathbb{R}^2$ egydimenziós altér egy részhalmaza.

Látni fogjuk: (X_1, X_2) nem **együttesen folytonos** valószínűségi vektorváltozó.

Egy **együttesen folytonos** (X, Y) 2-dim. valószínűségi vektorváltozó értékkészlete mindig egy pozitív területű halmaz.

Van **együttes sűrűségfüggvénye**: $f_{(X,Y)}: \mathbb{R}^2 \rightarrow [0, 1]$, aminek a kettős integrálja a teljes $\mathbb{R}^2$ -n 1.

Együttes sűrűségfüggvény

(7.1.3. Definíció.) Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Egy $f_{\underline{X}}: \mathbb{R}^n \rightarrow [0, \infty)$ függvény az $\underline{X}$ (együttes) sűrűségfüggvénye, ha $f_{\underline{X}}$ -nek létezik az improprius Riemann-integrálja $\mathbb{R}^n$ -en, és

$$\begin{aligned} \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_{\underline{X}}(z_1, \dots, z_n) dz_1, \dots, dz_n &= F_{\underline{X}}(x_1, \dots, x_n) \\ &= \mathbb{P}(X_1 < x_1, \dots, X_n < x_n). \end{aligned}$$

$\underline{X}$ -et folytonosnak hívjuk, ha létezik együttes sűrűségfüggvénye.

Megjegyzések:

- ▶ Egy dimenzióban ez ugyanaz, mint a folytonosság.
- ▶ Néha ahelyett, hogy $\underline{X}$ folytonos, azt mondjuk, hogy $X_1, \dots, X_n$ *együttesen folytonosak*.
- ▶ Néha lusták vagyunk és egy $\underline{X} = (X_1, X_2)$ kétdimenziós valószínűségi változó esetén $f_{\underline{X}} = f_{(X_1, X_2)}$ helyett f_{X_1, X_2} -t írunk.

Együttes sűrűségfüggvény

(7.1.3. Definíció.) Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Egy $f_{\underline{X}}: \mathbb{R}^n \rightarrow [0, \infty)$ függvény az $\underline{X}$ (együttes) sűrűségfüggvénye, ha $f_{\underline{X}}$ -nek létezik az improprius Riemann-integrálja $\mathbb{R}^n$ -en, és

$$\begin{aligned} \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_{\underline{X}}(z_1, \dots, z_n) dz_1, \dots, dz_n &= F_{\underline{X}}(x_1, \dots, x_n) \\ &= \mathbb{P}(X_1 < x_1, \dots, X_n < x_n). \end{aligned}$$

$\underline{X}$ -et folytonosnak hívjuk, ha létezik együttes sűrűségfüggvénye.

Legegyszerűbb példa: a $(0, 1)^2 \subset \mathbb{R}^2$ nyílt egységnégyzetlapon „találomra” (egyenletes eloszlással) választott $\underline{Y} = (Y_1, Y_2)$ pont sűrűségfüggvénye

$$f_{\underline{Y}}(y_1, y_2) = \begin{cases} 1, & \text{ha } 0 < y_1 < 1 \text{ és } 0 < y_2 < 1, \\ 0, & \text{különben.} \end{cases}$$

Együttes sűrűségfüggvény

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

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

(A deriválások tetszőleges sorrendben elvégezhetők.)

Megjegyzés: Az egyváltozós állítással ellentétben az állításnak feltétele, nem pedig következménye, hogy a sűrűségfüggvény létezzen (azaz $\underline{X}$ folytonos legyen).

Korábbi példa az $y = x$ egyenessel, ahol $X_2 = X_1$: $F_{(X_1, X_2)}$ néhány egyenes kivételével kétszer folytonosan (totálisan) differenciálható, de $\partial_{x_1} \partial_{x_2} F_{(X_1, X_2)}(x_1, x_2)$ mindenhol 0, ahol létezik, ami nem lehet együttes sűrűségfüggvény. (Világos?)

Így (X_1, X_2) nem folytonos, bár X_1 és X_2 peremeloszlásai folytonosak.

Együttes sűrűségfüggvény

Az együttes sűrűségfüggvények karakterizációja:

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

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(z_1, \dots, z_n) dz_1 \dots dz_n = 1$$

pontosan akkor teljesül, ha létezik olyan X valószínűségi vektorváltozó, amelynek f sűrűségfüggvénye.

Együttes sűrűségfüggvény

Egyváltozós eset $\rightarrow$ a sűrűségfüggvénnyel ki lehetett számolni pl. a $\mathbb{P}(a < X < b)$ alakú valószínűségeket. Többdimenziós eset:

(7.1.6. Állítás.) Legyen $H \subseteq \mathbb{R}^n$ Jordan-mérhető halmaz és $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó. Ekkor

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

Pl. $d = 2$ dimenzióban, egy $\underline{X} = (X_1, X_2)$ valószínűségi vektorváltozó esetén, ha $H = (a, b) \times (c, d)$ egy téglalap:

$$\begin{aligned} \mathbb{P}((X_1, X_2) \in H) &= \mathbb{P}(a < X_1 < b, c < X_2 < d) \\ &= \int_a^b \int_c^d f_{(X_1, X_2)}(x_1, x_2) dx_2 dx_1 \left(= \int_H f_{\underline{X}}(\underline{x}) d\underline{x} \right). \end{aligned}$$

Kétdimenziós egyenletes eloszlás

Kétdimenziós egyenletes eloszlás („találomra választott pont eloszlása”) általában: legyen $H \subseteq \mathbb{R}^2$ egy pozitív területű Jordan-mérhető halmaz és jelölje $T_H > 0$ a H területét.

Ekkor az $\underline{X} = (X, Y)$ val. vektorváltozó (kétdimenziós) egyenletes eloszlású H -n, ha

$$f_{(X,Y)}(x_1, x_2) = \begin{cases} \frac{1}{T_H}, & \text{ha } (x, y) \in H, \\ 0, & \text{ha } (x, y) \notin H. \end{cases}$$

A többdimenziós egyenletes eloszlást ugyanúgy az karakterizálja, hogy a sűrűségfüggvénye egy adott halmazon konstans, mindenhol máshol pedig 0, mint az egydimenziósat $\rightarrow$ geom. val. mező.

Példa: Legyen (X, Y) egyenletes eloszlású a $(0, 0)$, $(1, 0)$ és $(1, 1)$ csúcspontokkal rendelkező derékszögű, egyenlőszárú háromszögben. Ekkor $T_H = 1/2$ alapján a sűrűségfüggvény a háromszögön konstans 2.

Később erre a példára még visszatérünk.

Kicsit bonyolultabb példa: 8.6. feladat (lásd gyakorlat/megoldás).

Példa (nem egyenletes kétdimenziós eloszlásra)

Legyen $f: \mathbb{R}^2 \rightarrow [0, \infty)$,

$$f(x, y) = \begin{cases} 4xy, & \text{ha } 0 < x, y < 1, \\ 0, & \text{különben.} \end{cases}$$

- ▶ Van-e olyan (X, Y) val. vektorváltozó, amelynek sűrűségfüggvénye f ?
- ▶ Ha igen: $\mathbb{P}(X < Y) = ?$

(Megoldás: igen; $1/2$; lásd előadás)

Példa (nem egyenletes kétdimenziós eloszlásra)

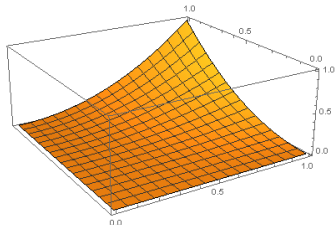
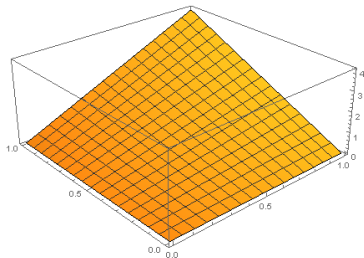
Legyen $f: \mathbb{R}^2 \rightarrow [0, \infty)$,

$$f(x, y) = \begin{cases} 4xy, & \text{ha } 0 < x, y < 1, \\ 0, & \text{különben.} \end{cases}$$

- ▶ Van-e olyan (X, Y) val. vektorváltozó, amelynek sűrűségfüggvénye f ?
- ▶ Ha igen: $\mathbb{P}(X < Y) = ?$

(Megoldás: igen; $1/2$; lásd előadás)

Együttes sűrűség- és eloszlásfüggvény a példában:



Az együttes sűrűségfüggvény szemléletes jelentése

Emlékeztető: ha X egydimenziós folytonos val. változó, akkor minden $x \in \mathbb{R}$ -re, ahol F_X differenciálható,

$$f_X(x) = \lim_{\Delta x \downarrow 0} \frac{F_X(x + \Delta x) - F_X(x - \Delta x)}{2\Delta x} = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))}{2\Delta x}.$$

Hasonlóan ha $\underline{X} = (X, Y)$ folytonos és az $F_{\underline{X}}$ együttes eloszlásfüggvény 2-szer totálisan differenciálható (x, y) -ban,

$$\begin{aligned} f_{\underline{X}}(x, y) &= \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{F_{\underline{X}}(x + \Delta x, y + \Delta y) - F_{\underline{X}}(x - \Delta x, y - \Delta y)}{2\Delta x \cdot 2\Delta y} \\ &= \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x), Y \in (y - \Delta y, y + \Delta y))}{2^2 \Delta x \cdot \Delta y}. \end{aligned}$$

Vagyis $f_{\underline{X}}(x, y)$ kb. annak a valószínűsége, hogy az $\underline{X} = (X, Y)$ vektor (x, y) egy kis környezetébe esik, osztva a környezet területével.

Az n -dimenziós eset hasonló, n -szer differenciálható $F_{\underline{X}}$ esetén, n -dimenziós térfogattal.

Marginális eloszlások ($n = 2$)

A következő pár dián tekintsük először a (kevésbé ijesztő) $n = 2$ esetet. Utána egy megjegyzés keretében rátérünk az általános n -re.

Ha (X, Y) egy valószínűségi vektorváltozó, akkor F_X -et és F_Y -t **perem-eloszlásfüggvényeknek (marginális eloszlásfüggvényeknek)** nevezzük. Hogyan kapjuk meg ezeket $F_{(X,Y)}$ -ből?

Az együttes eloszlásfüggvény argumentumában a másik változóval tartunk végtelenhez:

- ▶ $F_X(x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y), x \in \mathbb{R},$
- ▶ $F_Y(y) = \lim_{x \rightarrow \infty} F_{(X,Y)}(x, y), y \in \mathbb{R}.$

(Biz. vázlat: lásd előadás $\rightarrow$ val. mértékek folytonossága!)

Kérdés: hogyan kapjuk meg X marginális sűrűségfüggvényét?

- ▶ 1. válasz: határozzuk meg F_X -et, majd deriváljuk.
- ▶ 2. válasz: az együttes sűrűségfüggvényt integráljuk ki y szerint:

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy, \quad x \in \mathbb{R}.$$

Hasonlóan x szerint integrálva Y marginális sűrűségfüggvényét kapjuk.

Marginális eloszlások ($n = 2$)

Ha (X, Y) egy valószínűségi vektorváltozó, akkor F_X -et és F_Y -t perem-eloszlásfüggvényeknek (marginális eloszlásfüggvényeknek) nevezzük. Hogyan kapjuk meg ezeket $F_{(X,Y)}$ -ből?

Az együttes eloszlásfüggvény argumentumában a másik változóval tartsunk végtelenhez:

- ▶ $F_X(x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y), x \in \mathbb{R},$
- ▶ $F_Y(y) = \lim_{x \rightarrow \infty} F_{(X,Y)}(x, y), y \in \mathbb{R}.$

Kérdés: hogyan kapjuk meg X marginális sűrűségfüggvényét?

- ▶ 1. válasz: határozzuk meg F_X -et, majd deriváljuk.
- ▶ 2. válasz: az együttes sűrűségfüggvényt integráljuk ki y szerint:

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy, \quad x \in \mathbb{R}.$$

Hasonlóan x szerint integrálva Y marginális sűrűségfüggvényét kapjuk:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx, \quad y \in \mathbb{R}.$$

Marginális sűrűségfüggvény ($n = 2$)

Példa: egyenletes eloszlás/találomra választott $\underline{X} = (X, Y)$ pont a $(0, 1)^2$ négyzeten:

$$f_{(X,Y)}(x,y) = \begin{cases} 1, & \text{ha } 0 < x < 1 \text{ és } 0 < y < 1, \\ 0, & \text{különben.} \end{cases}$$

Így tehát $f_X(x) = \int_{-\infty}^{\infty} f_{\underline{X}}(x,y)dy = \int_0^1 1dy = 1$, ha $x \in (0, 1)$ és 0 különben.

Hasonlóan $f_Y(y) = 1$, ha $y \in (0, 1)$ és 0 különben.

Mind X , mind Y peremeloszlása egyenletes a $(0, 1)$ intervallumon.

Marginális sűrűségfüggvény ($n = 2$)

Eggyel érdekesebb példa: egyenletes eloszlás/találomra választott $\underline{X} = (X, Y)$ pont a $(0, 0), (1, 0), (1, 1)$ csúcspontokkal rendelkező háromszögön. Láttuk: $f_{(X,Y)}$ konstans 2 a háromszögön.

A háromszög: $\{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1, 0 < y < x\}$. Tehát

$$f_{(X,Y)}(x, y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben.} \end{cases}$$

Vagyis ($y < x$, figyeljünk az integrálási határookra!)

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy = \int_0^x 2 dz = 2x,$$

ha $x \in (0, 1)$ és $f_X(x) = 0$ egyébként; és

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx = \int_y^1 2 dx = 2 - 2y,$$

ha $y \in (0, 1)$ és $f_Y(y) = 0$ egyébként.

Sem X , sem Y peremeloszlása nem egyenletes a $(0, 1)$

intervallumon! Ez az $Y < X$ feltétel miatt van.

Marginális sűrűségfüggvény ($n = 2$)

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy, \quad x \in \mathbb{R}.$$

Miért igaz? $\rightarrow$ Láttuk:

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}(-\infty < X < x, -\infty < Y < \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f_{(X,Y)}(t,y)dydt.$$

Innen:

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Ez a magyarázat abban az esetben helyes, ha $F_{(X,Y)}$ minden pontban kétszer totálisan differenciálható (amiből az is következik, hogy a kétszeres parciális deriváltjai is léteznek).

Az általános esetben deriválás helyett integrálhelyettesítéssel kapjuk ugyanezt.

Összefoglalás: együttes és marginális eloszlás- és sűrűségfüggvény, $n = 2$

Áttekintés: egy $\underline{X} = (X, Y)$ folytonos val. vektorváltozónál ($n = 2$):

$$\begin{array}{ccccc}
 F_X(x) & \xleftarrow{\lim_{y \rightarrow \infty}} & F_{(X,Y)}(x,y) & \xrightarrow{\lim_{x \rightarrow \infty}} & F_Y(y) \\
 \uparrow & & \uparrow & & \uparrow \\
 \int_{-\infty}^x \cdots & \left| \begin{array}{c} \partial_x \\ \downarrow \end{array} \right. & \int_{-\infty}^x \int_{-\infty}^y \cdots & \left| \begin{array}{c} \partial_x \partial_y \\ \downarrow \end{array} \right. & \partial_y \left| \begin{array}{c} \int_{-\infty}^y \cdots \\ \downarrow \end{array} \right. \\
 f_X(x) & \xleftarrow{\int_{-\infty}^{\infty} \cdots dy} & f_{(X,Y)}(x,y) & \xrightarrow{\int_{-\infty}^{\infty} \cdots dx} & f_Y(y)
 \end{array}$$

(Ha nincs vissza nyíl, az nem véletlen!)

Marginális sűrűségfüggvény: általános n esete

Általános definíció:

(7.1.8. Definíció.) Ha $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó, akkor az X_i valószínűségi változó eloszlását az $\underline{X}$ val. vektorváltozó i -edik **marginális eloszlásának** (vagy **peremeloszlásának**) hívjuk.

Tegyük fel, hogy $\underline{X}$ folytonos $\Rightarrow X_i$ sűrűségfüggvénye?

Ugyanúgy kapjuk, mint az $n = 2$ esetben, csak most x_i -n kívül **az összes többi** változó szerint integrálni kell:

(7.1.9. Állítás.) Ha $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, akkor X_i is folytonos, sűrűségfüggvénye:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n, \quad x_i \in \mathbb{R}.$$

A 7.1.9. Állítás bizonyítása (előadáson nem mondom el)

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n,$$

Biz.: A bizonyítást csak abban az esetben adjuk meg, ha F_{X_i} mindenütt deriválható. A bizonyítás ötlete ugyanaz, mint az intuitív magyarázat az $n = 2$ esetben, csak kicsit precízebben. Láttuk:

$$\begin{aligned} F_{X_i}(x_i) &= \mathbb{P}(X_i < x_i) = \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i < x_i, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\ &= \lim_{x_1 \uparrow \infty, \dots, x_{i-1} \uparrow \infty, x_{i+1} \uparrow \infty, \dots, x_n \uparrow \infty} F_{\underline{X}}(x_1, \dots, x_n), \end{aligned}$$

az utolsó lépés a val. mértékek folytonossága miatt igaz. Így

$$\begin{aligned} f_{X_i}(x_i) &= \frac{\partial}{\partial x_i} F_{X_i}(x_i) = \frac{\partial}{\partial x_i} \left[\lim_{x_1 \uparrow \infty, \dots, x_{i-1} \uparrow \infty, x_{i+1} \uparrow \infty, \dots, x_n \uparrow \infty} F_{\underline{X}}(x_1, \dots, x_n) \right] \\ &= \frac{\partial}{\partial x_i} \left[\lim_{x_1 \uparrow \infty, \dots, x_{i-1} \uparrow \infty, x_{i+1} \uparrow \infty, \dots, x_n \uparrow \infty} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\underline{X}}(z_1, \dots, z_n) dz_1 \dots dz_n \right] \\ &= \frac{\partial}{\partial x_i} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underline{X}}(z_1, \dots, z_n) dz_1 \dots dz_n \right] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underline{X}}(z_1, \dots, z_n) dz_1, \dots, dz_{i-1} dz_{i+1} \dots dz_n. \end{aligned}$$

Megjegyzés: többdimenziós marginálisok

Ha pl. (X_1, X_2, X_3) egy 3-dimenziós folytonos val. vektorváltozó, akkor (X_1, X_2) egy 2-dimenziós, szintén folytonos val. vektorvált.

Az előbbi elv szerint az (X_1, X_2) együttes sűrűségfüggvényét, $f_{(X_1, X_2, X_3)}$ -mat a többi változó (itt X_3) szerint $-\infty$ -tól ∞ -ig kiintegrálva kapjuk meg:

$$f_{(X_1, X_2)}(x_1, x_2) = \int_{-\infty}^{\infty} f_{(X_1, X_2, X_3)}(x_1, x_2, x_3) dx_3.$$

Ez egy kétdimenziós marginális sűrűségfüggvénye (X_1, X_2, X_3) -nak. A bizonyítás ugyanaz, mint az egydimenziós marginálisok esetében.

Alternatív módszer, ha az $F_{(X_1, X_2, X_3)}$ együttes eloszlásfüggvény adott, akkor az (X_1, X_2) kétdimenziós valószínűségi vektorváltozó együttes eloszlásfüggvénye

$$F_{(X_1, X_2)}(x_1, x_2) = \lim_{x_3 \rightarrow \infty} F_{(X_1, X_2, X_3)}(x_1, x_2, x_3),$$

majd ezt x_1 és x_2 szerint is (tetszőleges sorrendben) parciálisan deriválva megkaphatjuk $f_{(X_1, X_2)}(x_1, x_2)$ -t.

Többszdimenziós folytonos eloszlások és momentumok

Adott egy $\underline{X} = (X_1, \dots, X_n)$ folytonos val. vektorváltozó.

Kérdés: Hogyan határozzuk meg $\mathbb{E}(X_i)$ -t ($1 \leq i \leq n$)?

1. válasz: X_i sűrűségfüggvényét, f_{X_i} -t már meg tudjuk határozni.

Meghatározzuk $\rightarrow \mathbb{E}(X_i) = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i$ (ha létezik).

Ez jó válasz, de mi van, ha pl. $\mathbb{E}(X_i X_j)$ -t keressük valamely

$1 \leq i, j \leq n$ -re, ahol $i \neq j$?

Erre szolgál a következő állítás.

(10.1.2. Állítás.) Legyen $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g: \mathbb{R}^n \rightarrow \mathbb{R}$ olyan függvény, amire

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Ekkor

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

2. válasz: $g(x_1, \dots, x_n) = x_i$ választással

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Többdimenziós folytonos eloszlások és momentumok

Adott egy $\underline{X} = (X_1, \dots, X_n)$ folytonos val. vektorváltozó.

Kérdés: Hogyan határozzuk meg $\mathbb{E}(X_i)$ -t ($1 \leq i \leq n$)?

1. válasz: X_i sűrűségfüggvényét, f_{X_i} -t már meg tudjuk határozni.

Meghatározzuk $\rightarrow \mathbb{E}(X_i) = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i$ (ha létezik).

(10.1.2. Állítás.) Legyen $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g: \mathbb{R}^n \rightarrow \mathbb{R}$ olyan függvény, amire

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Ekkor

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

2. válasz: $g(x_1, \dots, x_n) = x_i$ választással

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Többdimenziós folytonos eloszlások és momentumok

Adott egy $\underline{X} = (X_1, \dots, X_n)$ folytonos val. vektorváltozó.

Kérdés: Hogyan határozzuk meg $\mathbb{E}(X_i)$ -t ($1 \leq i \leq n$)?

1. válasz: X_i sűrűségfüggvényét, f_{X_i} -t már meg tudjuk határozni.

Meghatározzuk $\rightarrow \mathbb{E}(X_i) = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i$ (ha létezik).

(10.1.2. Állítás.) Legyen $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g: \mathbb{R}^n \rightarrow \mathbb{R}$ olyan függvény, amire

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Ekkor

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

2. válasz: $g(x_1, \dots, x_n) = x_i$ választással

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

A két módszer **ugyanaz**, csak az 1. válasz esetén az x_i szerinti integrálást

később végezzük el a többinél.

Többszempontos folytonos eloszlások és momentumok

Adott egy $\underline{X} = (X_1, \dots, X_n)$ folytonos val. vektorváltozó.

Kérdés: Hogyan határozzuk meg $\mathbb{E}(X_i)$ -t ($1 \leq i \leq n$)?

1. válasz: X_i sűrűségfüggvényét, f_{X_i} -t már meg tudjuk határozni.

Meghatározzuk $\rightarrow \mathbb{E}(X_i) = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i$ (ha létezik).

(10.1.2. Állítás.) Legyen $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g: \mathbb{R}^n \rightarrow \mathbb{R}$ olyan függvény, amire

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Ekkor

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

2. válasz: $g(x_1, \dots, x_n) = x_i$ választással

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

(Pl.: $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{2}{3}$, $\mathbb{E}(XY) = \frac{4}{9}$ a $4xy$ -os példában, lásd ea.)

Várható értékek a háromszögön egyenletes eloszlásnál

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben.} \end{cases}$$

Azt is láttuk, hogy $f_X(x) = 2x$, ha $x \in (0, 1)$ és 0 egyébként, illetve $f_Y(y) = 2 - 2y$, ha $y \in (0, 1)$ és 0 egyébként. Számoljuk ki X , Y és XY várható értékét!

Várható értékek a háromszögön egyenletes eloszlásnál

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben.} \end{cases}$$

Azt is láttuk, hogy $f_X(x) = 2x$, ha $x \in (0, 1)$ és 0 egyébként, illetve $f_Y(y) = 2 - 2y$, ha $y \in (0, 1)$ és 0 egyébként. Számoljuk ki X , Y és XY várható értékét!

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 2x dx = [2x^3/3]_{x=0}^1 = \frac{2}{3}$$

$$\text{és } \mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \cdot (2 - 2y) dy = \left[\frac{2y^2}{2} - \frac{2y^3}{3} \right]_{y=0}^1 = 1 - \frac{2}{3} = \frac{1}{3}.$$

$$\text{Továbbá } \mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{(X,Y)}(x,y) dy dx =$$

$$\int_0^1 \int_0^x 2xy dy dx = \int_0^1 [2xy^2/2]_{y=0}^x dx = \int_0^1 x^3 dx = [x^4/4]_{x=0}^1 = \frac{1}{4}.$$

Függetlenség folytonos esetben

Emlékeztető: (6.1.1. Definíció.) Legyenek $X, Y: \Omega \rightarrow \mathbb{R}$ valószínűségi változók az $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn.

Azt mondjuk, hogy X és Y függetlenek, ha minden $x, y \in \mathbb{R}$ -re az $\{X < x\}$ és az $\{Y < y\}$ események függetlenek.

A fenti feltétel más szavakkal:

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y), \quad \forall x,y \in \mathbb{R}.$$

n darab val. változó függetlenségét hasonlóan definiáljuk, az események **együttes függetlensége** fogalmának megfelelően.

Valójában ekkor is elég a függetlenséghez, ha az együttes eloszlásfüggvény a perem-eloszlásfüggvények szorzatára bomlik:

(7.2.2. Állítás.) Az $X_1, \dots, X_n$ valószínűségi változók pontosan akkor függetlenek, ha

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

minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

Ez igaz diszkrét, folytonos és általános esetben is.

Függetlenség folytonos esetben

Diszkrét eset:

(6.1.3. Állítás átfogalmazása.) Két diszkrét valószínűségi változó, X és Y pontosan akkor független, ha az együttes sűrűségfüggvényük a perem-sűrűségfüggvényeik szorzata, azaz minden $x, y \in \mathbb{R}$ esetén

$$p_{(X,Y)}(x,y) = \mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y) = p_X(x)p_Y(y).$$

Folytonos analógia: $X_1, \dots, X_n$ függetlenek $\Leftrightarrow \underline{X} = (X_1, \dots, X_n)$ együttes sűrűségfüggvénye az $X_1, \dots, X_n$ marginális sűrűségfüggvényeinek szorzatára bomlik.

(7.2.3. Állítás.) Legyenek $X_1, \dots, X_n$ folytonos valószínűségi változók. Ezek pontosan akkor függetlenek, ha $(X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó és

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

teljesül minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

(4xy-os példa: X és Y függetlenek, lásd előadás.)

Függetlenség: ellenpélda

A $(0, 0)$, $(1, 0)$, $(1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén X és Y függetlensége nem áll fenn.

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben,} \end{cases}$$

és $f_X(x) = 2x$, ha $0 < x < 1$ és 0 különben, $f_Y(y) = 2 - 2y$, ha $0 < y < 1$ és 0 különben.

Például:

- ▶ $f_X(\frac{1}{4}) = 2 \cdot \frac{1}{4} = \frac{1}{2}$,
- ▶ $f_Y(\frac{1}{2}) = 2 - 2 \cdot \frac{1}{2} = 1$,
- ▶ de $f_{(X,Y)}(\frac{1}{4}, \frac{1}{2}) = 0$, mivel $\frac{1}{4} < \frac{1}{2}$,

így $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$, ha $x = 1/4$ és $y = 1/2$.

Téglalapon (pl. négyzeten) vett egyenletes eloszlásnál viszont függetlenek (és egyenletes eloszlásúak) a koordináták, **ha a téglalap oldalai a koordinátatengelyekkel párhuzamosak.**

$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, ha X és Y függetlenek

(7.2.3. Állítás.) Legyenek $X_1, \dots, X_n$ folytonos valószínűségi változók. Ezek pontosan akkor függetlenek, ha $(X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó és

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

teljesül minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

Volt: (6.1.4. Állítás.) Ha X és Y független valószínűségi változók, és $\mathbb{E}(XY)$, $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ létezik, akkor

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \quad (1)$$

Ezt annak idején csak egyszerű val. változókra láttuk be, most belátjuk arra az esetre, amikor (X, Y) folytonos val. vektorváltozó (bizonyítás: lásd előadás).

(Egy már említett) következmény: ha X és Y függetlenek, akkor $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.*

Ahogy a 6.1.4. Állítás, ez is igaz minden X, Y val. változóra.*

*Feltéve, hogy $\mathbb{E}(X^2) < \infty$ és $\mathbb{E}(Y^2) < \infty$.

Függetlenség hiánya és várható értékek a háromszögön egyenletes eloszlásnál

Láttuk: a $(0, 0)$, $(1, 0)$, $(1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén $\mathbb{E}(X) = \frac{2}{3}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathbb{E}(XY) = \frac{1}{4}$.

Innen egyből látszik, hogy X és Y nem lehet független, mert $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$.

(Ha viszont az jött volna ki, hogy $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, abból nem következne automatikusan a függetlenség.)

Független val. változók transzformáltjai is függetlenek

A következő állítás nem nyilvánvaló, de itt nem bizonyítjuk:

(10.1.5. Lemma.) Ha X és Y független valószínűségi változók, g és h folytonos, valós függvények, akkor $g(X)$ és $h(Y)$ is függetlenek.

Hamarosan (a feltételes várható érték témájánál) sokat fogjuk ezt használni.

Valószínűségi változók egymáshoz való viszonya

Láttuk: ha X és Y független val. változók, akkor nem befolyásolják egymás értékét. Függő val. változók esetén viszont az egyik val. változó értékéből következtethetünk a másik értékére.

Meteorológiai példa:

- ▶ A napi átlagos relatív páratartalom és a napi csapadékmennyiség között „pozitív kapcsolat” van: ha tudjuk, hogy az egyik magas, akkor valószínű, hogy a másik is, és ha tudjuk, hogy az egyik alacsony, akkor bizonyára a másik is.
- ▶ A hónap végi aszályindex (ami a talaj szárazságát fejezi ki a szárazság mértékével növekvő skálán) és az elmúlt hónap csapadékmennyisége között „negatív kapcsolat” van: ha az egyik magas, akkor a másik várhatóan alacsony, és viszont.

Két val. változó közötti **lineáris** összefüggőség mérése → kovariancia, korreláció.

Valószínűségi változók egymáshoz való viszonya

A 6.1.4. Állítás és a 10.1.5. Lemma szerint ha X és Y független, akkor

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

minden g, h folytonos, valós függvényre, amelyre mindkét oldal definiált.

Kérdés: $g(x) = h(x) = x$ (**lineáris** függvény) választással mikor teljesül

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

nem feltétlenül független X, Y esetén? Már egyszer említettük, hogy ez előfordulhat nem független val. változók esetén is.

Kovariancia

Kérdés: Mennyire tér el $\mathbb{E}(XY)$ -től $\mathbb{E}(X)\mathbb{E}(Y)$? Ennek mérésére vezetjük be a következő mennyiséget.

(6.3.1. Definíció.) Az X és Y valószínűségi változók **kovarianciáját** a következőképp definiáljuk:

$$\text{cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))),$$

feltéve, hogy $\mathbb{E}(X^2) < \infty$ és $\mathbb{E}(Y^2) < \infty$.

(6.3.2. Állítás) Ha $\text{cov}(X, Y)$ értelmes, akkor

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

(Bizonyítás)

Technikai megjegyzés: Előfordulhat, hogy $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$ véges akkor is, ha $\mathbb{E}(X^2) = \infty$ vagy $\mathbb{E}(Y^2) = \infty$, azonban a kovariancia további tulajdonságai általában csak akkor igazak, ha $\mathbb{E}(X^2) < \infty$ és $\mathbb{E}(Y^2) < \infty$.

Kovariancia

Kérdés: Mennyire tér el $\mathbb{E}(XY)$ -től $\mathbb{E}(X)\mathbb{E}(Y)$? Ennek mérésére vezetjük be a következő mennyiséget.

(6.3.1. Definíció.) Az X és Y valószínűségi változók **kovarianciáját** a következőképp definiáljuk:

$$\text{cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))),$$

feltéve, hogy $\mathbb{E}(X^2) < \infty$ és $\mathbb{E}(Y^2) < \infty$.

(6.3.2. Állítás) Ha $\text{cov}(X, Y)$ értelmes, akkor

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

(Bizonyítás)

Emlékeztető:

diszkrét eset: $\mathbb{E}(XY) = \sum_{k \in \text{Ran}(X)} \sum_{l \in \text{Ran}(Y)} kl \mathbb{P}(X = k, Y = l),$

együttesen folytonos eset:

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{(X,Y)}(x,y) dx dy.$$

Kovariancia: példa

$Y \backslash X$	X		p_Y
	0	1	
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Egy pénzérmét kétszer feldobunk, $X = \mathbb{1}_{\{\text{a 2. dobás fej}\}}$ $Y =$ a dobott fejek száma.

Ebben a példában egyszer már kiszámoltuk, hogy $\mathbb{E}(XY) = \frac{3}{4}$.
Továbbá $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = 1$, tehát $\text{cov}(X, Y) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$.

Lesz: pozitív kovariancia $\rightarrow$ pozitív lineáris kapcsolat X és Y között.

A kovariancia tulajdonságai I.

(6.3.3. Következmény.) Legyenek X és Y valószínűségi változók, amikre $\text{cov}(X, Y)$ értelmes.

1. Ha Y konstans, akkor $\text{cov}(X, Y) = 0$.
2. Ha X és Y függetlenek, akkor $\text{cov}(X, Y) = 0$.
3. Attól, hogy $\text{cov}(X, Y) = 0$, még nem feltétlenül teljesül, hogy X és Y független.

Bizonyítás:

1. Lásd előadás (várható érték linearitása)
2. Már láttuk, hogy $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Ezt átrendezve $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.
3. Egyszerű ellenpélda: legyen $\text{Ran}(X) = \{-1, 0, 1\}$, amely értékeket $\frac{1}{4}$, $\frac{1}{2}$ és $\frac{1}{4}$ valószínűségekkel veszi fel X .
Legyen $Y = |X|$. Kiszámolható (lásd előadás), hogy $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 \cdot \frac{1}{2} = 0$, de X és Y nem függetlenek, mert pl.

$$0 = \mathbb{P}(X = 1, Y = 0) \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{8}.$$

A kovariancia tulajdonságai II.

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

(Állítás.) „A kovariancia gyönyörű.” (Balázs Márton)

Legyen (X, Y, Z, W) valószínűségi vektorváltozó és $a, b, c, d \in \mathbb{R}$.

Ekkor teljesülnek az alábbiak, feltéve, hogy a bennük szereplő mennyiségek értelmezettek:

1. $\text{cov}(X, X) = \mathbb{D}^2(X)$. Következmény: $\text{cov}(X, X) \geq 0$, és $\text{cov}(X, X) = 0 \Leftrightarrow X$ egy valószínűséggel konstans.

2. Szimmetria: $\text{cov}(X, Y) = \text{cov}(Y, X)$.

3. Bilinearitás (linearitás mindkét változóban):

$$\text{cov}(X+Y, Z+W) = \text{cov}(X, Z) + \text{cov}(X, W) + \text{cov}(Y, Z) + \text{cov}(Y, W).$$

► A bilinearitás ekvivalens megfogalmazása:

$$\text{cov}(X, bY + cZ) = b \cdot \text{cov}(X, Y) + c \cdot \text{cov}(X, Z),$$

és ugyanez a másik változóban.

► Köv.: „additív konstans elhagyható, konstans szorzó mindkét változóból kiemelhető”:

$$\text{cov}(aX + b, cY + d) = \text{cov}(aX, cY) = a \cdot c \cdot \text{cov}(X, Y).$$

Összeg szórásnégyzete

Megjegyzés: Azt, hogy „additív konstans elhagyható”, úgy is mondjuk, hogy a kovariancia **eltolás-invariáns**. Láttuk, hogy a szórás(négyzet) ugyanezzel a tulajdonsággal rendelkezik.

(Állítás.) Ha X, Y valószínűségi változók, amelyekre $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, akkor

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\text{cov}(X, Y).$$

(Biz.: $\mathbb{D}^2(X + Y) = \text{cov}(X + Y, X + Y)$, erre használjuk a kovariancia bilinearitását és szimmetriáját.)

Újra látjuk: ha $\text{cov}(X, Y) = 0$, akkor

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

Már ismert speciális eset: ha X és Y függetlenek, akkor

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

Ha $\text{cov}(X, Y) \neq 0$, ez nem teljesül!

Példa: ha $\text{cov}(X, Y) = 1$, $\mathbb{D}^2(X) = 2$ és $\mathbb{D}^2(Y) = 3$, akkor mennyi $\text{cov}(2X + Y - 42, 3Y + 100)$?

Analógia a vektorok skaláris szorzásával

A kovariancia a vektorok skaláris szorzatával (majdnem) analóg mennyiség.

Emlékeztető BSz-ből: ha $\underline{u} = (u_1, u_2, u_3)$, $\underline{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$, akkor $\underline{u}$ és $\underline{v}$ skaláris szorzata

$$\underline{u} \cdot \underline{v} = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

3 helyett n dimenzióban ugyanígy definiálható, n tagú összeg.

Tulajdonságok (csak részben voltak BSz-ből):

- ▶ Pozitív definit: $\underline{u} \cdot \underline{u} = u_1^2 + u_2^2 + u_3^2 \geq 0$ és $\underline{u} \cdot \underline{u} = 0 \Leftrightarrow \underline{u} = \underline{0}$.
- ▶ Szimmetrikus: $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$.
- ▶ Bilineáris: $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$ (ugyanígy a másik változóban, továbbá skalár szorzó mindkét változóból kiemelhető).

Analógia a vektorok skaláris szorzásával

A valószínűségi változók kovarianciája ugyanezekkel a tulajdonságokkal rendelkezik, azt leszámítva, hogy „additív konstans elhagyható”, azaz konstans val. változó bármivel vett kovarianciája 0.

- ▶ Pozitív **szemidefinit**: $\text{cov}(X, X) \geq 0$ és $\text{cov}(X, X) = 0 \Leftrightarrow \mathbb{P}(X = \mathbf{c}) = 1$ valamely $\mathbf{c} \in \mathbb{R}$ -re.
Ebből a $\text{cov}(X, X) \geq 0$ tulajdonságot nevezzük pozitív szemidefinitiségnek.
- ▶ Szimmetrikus: $\text{cov}(X, Y) = \text{cov}(Y, X)$.
- ▶ Bilineáris: $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$ (ugyanígy a másik változóban).

Várhatóérték-vektor, kovarianciamátrix

(12.3.4. Definíció.) Egy $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó **várhatóérték-vektora** az $(\mathbb{E}X_1, \dots, \mathbb{E}X_n) \in \mathbb{R}^n$ vektor. Jelölés: $\mathbb{E}\underline{X}$ (vagy $\mathbb{E}(\underline{X})$, $\mathbb{E}[\underline{X}]$).

(10.1.6. Definíció.) A fenti $\underline{X}$ **kovarianciamátrixa**

$$\text{cov}(\underline{X}) = \begin{pmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \dots & \dots & \ddots & \dots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Megjegyzés: oszlopvektorként kezelve a val. változókat,

$$\text{cov}(\underline{X}) = \mathbb{E}((\underline{X} - \mathbb{E}\underline{X})(\underline{X} - \mathbb{E}\underline{X})^T).$$

A kovarianciamátrix tulajdonságai

Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó.

$$\text{cov}(\underline{X}) = \begin{pmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \dots & \dots & \ddots & \dots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

(10.1.7. Állítás.) Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Ekkor

1. $\text{cov}(\underline{X})$ szimmetrikus, azaz $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$ minden $1 \leq i \leq j$ esetén.
2. $\text{cov}(\underline{X})$ pozitív szemidefinit mátrix, azaz minden $\underline{a} = (a_1, \dots, a_n)^T \in \mathbb{R}^n$ esetén

$$\underline{a}^T \text{cov}(\underline{X}) \underline{a} = \sum_{i,j=1}^n a_i \text{cov}(X_i, X_j) a_j \geq 0,$$

és “= 0” $\Leftrightarrow \sum_{i=1}^n a_i X_i$ egy valószínűséggel konstans.

(Bizonyítás: lásd előadás/jegyzet)

A kovarianciamátrix tulajdonságai

Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó.

$$\text{cov}(\underline{X}) = \begin{pmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \dots & \dots & \ddots & \dots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Megjegyzés (BSz-ből nem volt, de Anal2-ből talán ismerős?): Egy **szimmetrikus** $n \times n$ -es A mátrixnak mindig van (valós) sajátértéke. Sőt a $\det(A - \lambda E) = 0$ karakterisztikus egyenlet összes λ megoldása valós.

A pozitív szemidefinit $\Leftrightarrow A$ összes sajátértéke ≥ 0 .

A pénzürmés példában: $\text{cov}(X, Y) = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 1 \end{pmatrix}$

(számoljuk ki a sajátértékeit!).

CSB-egyenlőtlenség

A térbeli vektorok skaláris szorzatára igaz a

Cauchy–Schwarz–Bunyakovszkij-egyenlőtlenség (CSB):

$|\underline{u} \cdot \underline{v}| \leq \sqrt{|\underline{u} \cdot \underline{u}| \cdot |\underline{v} \cdot \underline{v}|}$, és egyenlőség akkor és csak akkor áll fenn, ha $\underline{u}$ és $\underline{v}$ párhuzamosak (azaz **lineárisan összefüggők**).

Más szavakkal: ha $\underline{u}, \underline{v} \neq \underline{0}$:

$$-1 \leq \frac{\underline{u} \cdot \underline{v}}{\sqrt{(\underline{u} \cdot \underline{u}) \cdot (\underline{v} \cdot \underline{v})}} \leq +1.$$

- ▶ $= +1 \Rightarrow \underline{u}$ és $\underline{v}$ párhuzamosak és **egyállásúak**, azaz $\underline{v} = a\underline{u}$ valamely $a > 0$ -ra.
- ▶ $= -1 \Rightarrow \underline{u}$ és $\underline{v}$ párhuzamosak és **ellenkező állásúak**, azaz $\underline{v} = a\underline{u}$ valamely $a < 0$ -ra.
- ▶ $= 0$ pontosan akkor, ha $\underline{u}$ és $\underline{v}$ merőlegesek.

Interpretáció: a hányados a „lineáris összefüggőség mértékét” mutatja (persze két nem párhuzamos, nem null vektor mindig lineárisan független).

Miért igazak ezek? Mert $\underline{u} \cdot \underline{v} = |\underline{u}||\underline{v}| \cos(\alpha)$, ahol α a két vektor bezárt szöge.

CSB-egyenlőtlenség és korreláció

$$-1 \leq \frac{\underline{u} \cdot \underline{v}}{\sqrt{|\underline{u} \cdot \underline{u}| \cdot |\underline{v} \cdot \underline{v}|}} \leq +1, \quad \underline{u}, \underline{v} \neq 0.$$

A tört kifejezés analógiája a kovariancia esetén a **korreláció**:

(6.5.1. Definíció.) Legyenek X és Y valószínűségi változók, amelyek értéke nem 1 valószínűséggel konstans. Ha $\mathbb{E}(X^2) < \infty$ és $\mathbb{E}(Y^2) < \infty$, akkor X és Y **korrelációja**:

$$\text{corr}(X, Y) \stackrel{\text{def.}}{=} \frac{\text{cov}(X, Y)}{\sqrt{\text{cov}(X, X)\text{cov}(Y, Y)}} = \frac{\text{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$$

Azért kell, hogy X és Y ne legyen 1 valószínűséggel konstans, hogy a szórásuk ne 0 legyen.

Korreláció

A korreláció is mindig -1 és 1 („ -100% és $+100\%$ ”) közé esik. A következő állítások a CSB-egyenlőtlenség analógiái, figyelembe véve az „**additív konstans elhagyható**” elvet (lásd 6.5.2. Állítás):

$$-1 \leq \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)} \leq 1.$$

- ▶ $\text{corr}(X, Y) = +1 \Rightarrow Y = aX + b$, 1 valószínűséggel, valamely $a > 0$ és $b \in \mathbb{R}$ esetén,
- ▶ $\text{corr}(X, Y) = -1 \Rightarrow Y = aX + b$, 1 valószínűséggel, valamely $a < 0$ és $b \in \mathbb{R}$ esetén,
- ▶ ha X és Y függetlenek, akkor $\text{corr}(X, Y) = 0$.
 $\text{corr}(X, Y) = 0$ -ból általában nem következik, hogy X, Y függetlenek, általában csak *lineáris* összefüggés nincs a két val. változó között.

Szóhasználat: $\text{corr}(X, Y) > 0$: X és Y **pozitívan korreláltak**,
 $\text{corr}(X, Y) = 0$: **korrelálatlanok**, $\text{corr}(X, Y) < 0$: **negatívan korreláltak**.

Korreláció: példa

$Y \backslash X$	X		p_Y
	0	1	
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Egy pénzérmét kétszer feldobunk, $X = \mathbb{1}_{\{\text{a 2. dobás fej}\}}$, $Y = \text{a dobott fejek száma}$. Láttuk: $\text{cov}(X, Y) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$.
Továbbá $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = 1$, tehát

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

($= p(1-p)$, $p = 1/2$, hiszen X egy $1/2$ valószínűségű esemény indikátora), és

$$\mathbb{D}^2(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{1}{2} \cdot 1^2 + \frac{1}{4} \cdot 2^2 - 1^2 = \frac{1}{2}.$$

$$\text{Így } \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{D}(X)\mathbb{D}(Y)}} = \frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{\sqrt{2}}} = \frac{\sqrt{2}}{2}.$$

X és Y pozitívan korreláltak (mi ennek az oka?).

Korreláció a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén $\mathbb{E}(X) = \frac{2}{3}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathbb{E}(XY) = \frac{1}{4}$. Tehát

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}.$$

Továbbá

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 2x dx = [2x^4/4]_{x=0}^1 = \frac{1}{2}.$$

Tehát $\mathbb{D}^2(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$. Hasonlóan

$$\mathbb{E}(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_0^1 y^2 (2-2y) dy = [2y^3/3 - 2y^4/4]_{y=0}^1 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

és így $\mathbb{D}^2(Y) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$. Ezért

$$\text{corr}(X, Y) = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}.$$

(Intuitív magyarázat arra, hogy miért pozitívan korreláltak $\rightarrow$ ea.)

(Magyarázat arra, hogy X és Y miért azonos szórásúak: könnyen látható,

hogy X és $1 - Y$ azonos eloszlásúak, $\mathbb{D}(1 - Y) = \mathbb{D}(Y)$.)

Korreláció a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén $\mathbb{E}(X) = \frac{2}{3}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathbb{E}(XY) = \frac{1}{4}$. Tehát

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}.$$

Továbbá

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 2x dx = [2x^4/4]_{x=0}^1 = \frac{1}{2}.$$

Tehát $\mathbb{D}^2(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$. Hasonlóan

$$\mathbb{E}(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_0^1 y^2 (2-2y) dy = [2y^3/3 - 2y^4/4]_{y=0}^1 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

és így $\mathbb{D}^2(Y) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$. Ezért

$$\text{corr}(X, Y) = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}.$$

(Intuitív magyarázat arra, hogy miért pozitívan korreláltak $\rightarrow$ ea.)

(Magyarázat arra, hogy X és Y miért azonos szórásúak: könnyen látható,

hogy X és $1 - Y$ azonos eloszlásúak, $\mathbb{D}(1 - Y) = \mathbb{D}(Y)$.)

Korreláció a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén $\mathbb{E}(X) = \frac{2}{3}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathbb{E}(XY) = \frac{1}{4}$. Tehát

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}.$$

Továbbá

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 2x dx = [2x^4/4]_{x=0}^1 = \frac{1}{2}.$$

Tehát $\mathbb{D}^2(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$. Hasonlóan

$$\mathbb{E}(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_0^1 y^2 (2-2y) dy = [2y^3/3 - 2y^4/4]_{y=0}^1 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

és így $\mathbb{D}^2(Y) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$. Ezért

$$\text{corr}(X, Y) = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}.$$

(Intuitív magyarázat arra, hogy miért pozitívan korreláltak $\rightarrow$ ea.)

(Magyarázat arra, hogy X és Y miért azonos szórásúak: könnyen látható, hogy X és $1 - Y$ azonos eloszlásúak, $\mathbb{D}(1 - Y) = \mathbb{D}(Y)$.)

Példa arra, amikor a korreláció –1

Korábbi példa a 6b. gyakorló feladatsorról:

Lajoska nagyon szereti a villamosokat, ezért a nyári szünet egy szép napján kimegy a Határ útra és megfigyel 3 db egymás után közlekedő 3-as villamost. Tegyük fel, hogy a villamosok egymástól függetlenül 70% valószínűséggel CAF és 30% valószínűséggel hannoveri típusúak. Jelölje X a Lajoska által megfigyelt CAF és Y az általa megfigyelt hannoveri villamosok számát.

Honnan látjuk számolás nélkül is, hogy $\text{corr}(X, Y) = -1$?

Példa arra, amikor a korreláció –1

Korábbi példa a 6b. gyakorló feladatsorról:

Lajoska nagyon szereti a villamosokat, ezért a nyári szünet egy szép napján kimegy a Határ útra és megfigyel 3 db egymás után közlekedő 3-as villamost. Tegyük fel, hogy a villamosok egymástól függetlenül 70% valószínűséggel CAF és 30% valószínűséggel hannoveri típusúak. Jelölje X a Lajoska által megfigyelt CAF és Y az általa megfigyelt hannoveri villamosok számát.

Honnan látjuk számolás nélkül is, hogy $\text{corr}(X, Y) = -1$?

Onnan, hogy mivel $X + Y$ értéke mindig 3, ezért $Y = 3 - X$ ($a = -1, b = 3$).

Arra, amikor $\text{corr}(X, Y) = +1$, a legegyszerűbb példa az, ha $X = Y$ ($a = 1, b = 0$).

(Ekkor $\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)} = \frac{\mathbb{D}^2(X)}{\mathbb{D}(X)\mathbb{D}(X)} = 1$.)

Másik diszkrét példa

Másik példa, szintén a 6b. gyakorló feladatsorról:

Egy okostelefonon 2 játékal alkalmazás és 3 közösségi média alkalmazás van. A telefon operációs rendszere először véletlenszerűen kiválaszt az öt alkalmazásból egyet, amit frissít, majd ha ez egy közösségi média alkalmazás volt, akkor a maradék négyből kiválaszt még egyet véletlenszerűen, amit frissít, ha pedig játékal alkalmazás volt, akkor a három közösségi média alkalmazás közül választ ki egyet, amit szintén frissít. Jelölje X a frissített játékal alkalmazások és Y a frissített közösségi média alk. számát. Határozzuk meg (lehetőleg számolás nélkül) $\text{corr}(X, Y)$ -t.

Másik diszkrét példa

Másik példa, szintén a 6b. gyakorló feladatsorról:

Egy okostelefonon 2 játékkalkalmazás és 3 közösségi média alkalmazás van. A telefon operációs rendszere először véletlenszerűen kiválaszt az öt alkalmazásból egyet, amit frissít, majd ha ez egy közösségi média alkalmazás volt, akkor a maradék négyből kiválaszt még egyet véletlenszerűen, amit frissít, ha pedig játékkalkalmazás volt, akkor a három közösségi média alkalmazás közül választ ki egyet, amit szintén frissít. Jelölje X a frissített játékkalkalmazások és Y a frissített közösségi média alk. számát. Határozzuk meg (lehetőleg számolás nélkül) $\text{corr}(X, Y)$ -t.

- ▶ Itt is igaz, hogy $\text{corr}(X, Y) = -1$, hiszen $X + Y$ összege 1 valószínűséggel 2, tehát $Y = 2 - X$ ($a = -1, b = 2$).
- ▶ Elsőre ez talán kevésbé feltűnő, mint a villamosos példában, mert X és Y értékkészlete különböző.
- ▶ Y értéke nem lehet 0, így Y peremeloszlása ezúttal nem binomiális. (X viszont 0–1-értékű val. változó, tehát Bernoulli-eloszlású, ami a binomális speciális esete $\rightarrow$ melyik esemény indikátora X ?)

Korreláció: interpretáció

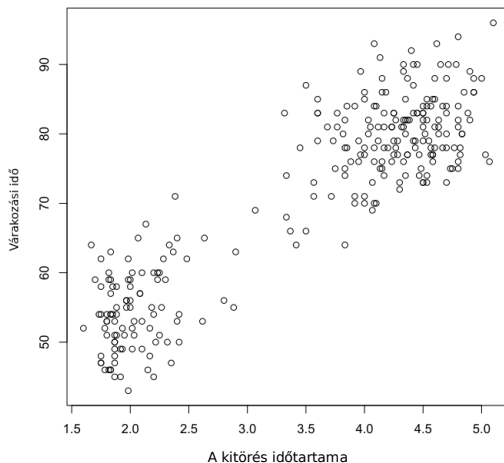
Az időjárásos példában:

Páratartalom és csapadék → pozitívan, aszályindex és csapadék → negatívan korreláltak.

Egyes társadalomtudományi területeken gyakori félreértés: két véletlen mennyiség között a nagy (1-hez közeli) korreláció általában nem ok-okozati, hanem **lineáris** összefüggést jelent. Lehet pl., hogy két pozitívan korrelált mennyiségből egyik sem a másik következménye, hanem mindkettő egy harmadik okból következik.

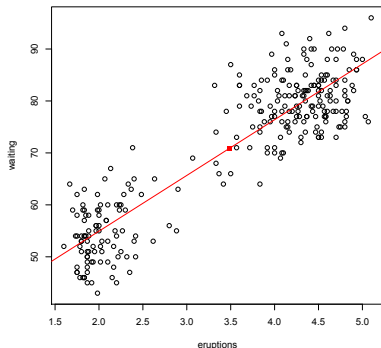
Lineáris regresszió

Példa (faithful adathalmaz R-ben): az ábrán egy gejzír kitöréseiig eltelt várakozási időket és a kitörések időtartamait ábráztuk. A két változó között pozitív korreláció mutatkozik.



Lineáris regresszió

Példa (`faithful` adathalmaz R-ben): az ábrán egy gejzír kitöréseig eltelt várakozási időket és a kitörések időtartamait ábrázoltuk. A két változó között pozitív korreláció mutatkozik.



Tegyük fel, hogy a változók együttes eloszlását ismerjük.

Cél: az y -adatpontokat az x -adatpontok egy lineáris függvényével közelíteni $\rightarrow$ a közelítés „átlagos” hibája legyen a lehető legkisebb.

Optimális egyenes? $\rightarrow$ Hogyan mérjük a hibát?

Lineáris regresszió

A következő definíció megadja, hogy hogyan mérjük a hibát, illetve mit tekintünk optimális lineáris függvénynek:

(10.2.1. Definíció.) Legyenek X és Y valószínűségi változók. Ekkor Y -nak az X -re vett **lineáris regresszióján** azt a $\beta X + \alpha$ valószínűségi változót értjük, ahol $\alpha, \beta \in \mathbb{R}$, és az

$$\mathbb{E}\left((Y - (\beta X + \alpha))^2\right)$$

menyiség minimális.

Itt:

$$\mathbb{E}\left(\underbrace{\left(\underbrace{Y}_{\text{„magyarázott/függő változó”}} - \underbrace{(\beta \underbrace{X}_{\text{„magyarázó változó”}} + \alpha)}_{\text{közelítés (lin. függvény)}}\right)^2}_{\text{közelítés hibája}}\right)$$

Lineáris regresszió

(10.2.2. Állítás.) Legyenek X és Y olyan valószínűségi változók, amire $\mathbb{E}(X^2)$ és $\mathbb{E}(Y^2)$ véges, továbbá $\mathbb{D}^2(X) \neq 0$. Ekkor az előbbi definícióban szereplő várható érték pontosan akkor minimális, ha

$$\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} \quad \text{és} \quad \alpha = \mathbb{E}(Y) - \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} \mathbb{E}(X).$$

(Bizonyítás: lásd előadás/jegyzet, anal 2 style)

(10.2.3. Definíció, kissé kiegészítve.) Az Y valószínűségi változó X -re vett **regressziós egyenese** az

$$\{(x, y) \in \mathbb{R}^2 : y = \beta x + \alpha\}$$

egyenes a síkon, Y -nak X -re vett **lineáris regressziója** pedig a $\beta X + \alpha$ valószínűségi változó, ahol α és β értéke a fenti állításban szerepel.

Lineáris regresszió

(10.2.2. Állítás.) Legyenek X és Y olyan valószínűségi változók, amire $\mathbb{E}(X^2)$ és $\mathbb{E}(Y^2)$ véges, továbbá $\mathbb{D}^2(X) \neq 0$. Ekkor az előbbi definícióban szereplő várható érték pontosan akkor minimális, ha

$$\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} \quad \text{és} \quad \alpha = \mathbb{E}(Y) - \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} \mathbb{E}(X).$$

(Bizonyítás: lásd előadás/jegyzet, anal 2 style)

Megjegyzés? Y -nak és $\beta X + \alpha$ -nak ugyanaz a várható értéke és az X -szel vett kovarianciája (lásd előadás). Ha ezt tudjuk, β és α ezekből az egyenletekből is kifejezhető.

Lineáris regresszió: megjegyzések

Emlékeztető: korreláció: $\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\mathbb{D}X\mathbb{D}Y}$.

Állítás:

$$\frac{(\beta X + \alpha) - \mathbb{E}(Y)}{\mathbb{D}(Y)} = \frac{X - \mathbb{E}(X)}{\mathbb{D}X} \text{corr}(X, Y).$$

Más szavakkal: ha Y standardizáltjába ($= \frac{Y - \mathbb{E}(Y)}{\mathbb{D}(Y)}$) az első Y helyére a

$\beta X + \alpha$ lin. regressziót helyettesítjük, akkor az eredmény X standardizáltjának korreláció-szorosa.

(Egyszerű átrendezéssel belátható.)

Lineáris regresszió: példa

$Y \backslash X$	0	1	p_Y
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Egy pénzérmét kétszer feldobunk, $X = \mathbb{1}_{\{\text{a 2. dobás fej}\}}$, $Y = \text{a dobott fejek száma}$. Láttuk: $\text{cov}(X, Y) = \frac{1}{4}$, $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = 1$, $\mathbb{D}^2(X) = \frac{1}{4}$, $\mathbb{D}^2(Y) = \frac{1}{2}$. Ezért

$$\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = 1, \quad \alpha = \mathbb{E}(Y) - \beta \mathbb{E}(X) = 1 - 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

A regressziós egyenes $\{(x, y) \in \mathbb{R}^2 : y = x + \frac{1}{2}\}$.

Lineáris regresszió hibájának szórásnégyzete

(10.2.5. Állítás.) Legyen az Y valószínűségi változó X -re vett lineáris regressziója

$\beta X + \alpha$. Ekkor az eltérés/hiba szórásnégyzete:

$$\mathbb{D}^2(Y - (\beta X + \alpha)) = \mathbb{D}^2(Y) - \frac{\text{cov}(X, Y)^2}{\mathbb{D}^2(X)}.$$

(Bizonyítás: szórásnégyzet tulajdonságai és $\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)}$ alapján)

Megjegyzés: Ilyen képleteket nem feltétlenül érdemes megjegyezni, sokkal inkább a szórásnégyzet azonosságait érdemes tudni. És persze β -t, α -t sem árt.

A fenti egyenletet átrendezve:

$$\mathbb{D}^2(Y - (\beta X + \alpha)) = \mathbb{D}^2(Y) \left(1 - \text{corr}(X, Y)^2\right).$$

Vagyis minél nagyobb $|\text{corr}(X, Y)|$, annál kisebb a lineáris regresszió hibája.

$\text{corr}(X, Y) = \pm 1$ esetén a hiba 0, ami nem meglepő, hiszen tudjuk, hogy ebben az esetben Y egy lineáris függvénye, és β, α definíciója alapján $Y = \beta X + \alpha$.

Lineáris regresszió hibája: példa

$\begin{array}{c} Y \\ \backslash X \end{array}$	0	1	p_X
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_Y	$\frac{1}{2}$	$\frac{1}{2}$	1

Láttuk: $\text{cov}(X, Y) = \frac{1}{4}$, $\mathbb{D}^2(X) = \frac{1}{4}$, $\mathbb{D}^2(Y) = \frac{1}{2}$. Ezért

$$\mathbb{D}^2(Y - (\beta X + \alpha)) = \mathbb{D}^2(Y) - \frac{\text{cov}(X, Y)^2}{\mathbb{D}^2(X)} = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Persze ugyanezt kiszámolhatjuk kézzel is ($\beta = 1$, $\alpha = \frac{1}{2}$):

$$\begin{aligned}\mathbb{D}^2(Y - (\beta X + \alpha)) &= \mathbb{D}^2\left(Y - \left(X + \frac{1}{2}\right)\right) = \mathbb{D}^2(Y - X) \\ &= \mathbb{D}^2(Y) + \mathbb{D}^2(X) - 2\text{cov}(X, Y) = \frac{1}{2} + \frac{1}{4} - \frac{2}{4} = \frac{1}{4}.\end{aligned}$$

Lineáris regresszió a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén $\text{cov}(X, Y) = \frac{1}{36}$, $\mathbb{D}^2(X) = \frac{1}{18}$, $\mathbb{E}(Y) = \frac{1}{3}$, $\mathbb{E}(X) = \frac{2}{3}$. Tehát

$$\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = \frac{1}{2}, \quad \alpha = \mathbb{E}(Y) - \beta\mathbb{E}(X) = 0.$$

A lineáris regresszió tehát $\frac{1}{2}X$, a regressziós egyenes pedig

$$\{(x, y) \in \mathbb{R}^2 : y = \frac{1}{2}x\}.$$

Később (a feltételes várható értéknél) lesz rá intuitív magyarázat, hogy miért ennek kell kijönnie.

Feltételes eloszlások: diszkrét eset

Egy kicsit térjünk vissza a diszkrét együttes eloszlásos példánkra:

$Y \backslash X$	X		p_Y
	0	1	
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Kérdés: Mennyi $\mathbb{P}(Y < 2|X = 1)$?

Válasz: mivel $\{X = 1\}$ egy nem 0 valószínűségű esemény, ezt a feltételes valószínűség szokásos képletével lehet meghatározni:

$$\begin{aligned}\mathbb{P}(Y < 2|X = 1) &= \frac{\mathbb{P}(Y < 2, X = 1)}{\mathbb{P}(X = 1)} = \frac{\mathbb{P}(Y = 0, X = 1) + \mathbb{P}(Y = 1, X = 1)}{\mathbb{P}(X = 1)} \\ &= \frac{0 + 1/4}{1/2} = \frac{1}{2}.\end{aligned}$$

Ha $x \in \mathbb{R}$ -re $\mathbb{P}(X = x) > 0$, az $y \mapsto \mathbb{P}(Y < y|X = x)$ **feltételes eloszlásfüggvény** úgy viselkedik, mint egy sima eloszlásfüggvény.

Feltételes eloszlások: diszkrét eset

Egy kicsit térjünk vissza a diszkrét együttes eloszlásos példánkra:

$Y \backslash X$	X		p_Y
	0	1	
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Kérdés: Mennyi $\mathbb{P}(Y < 2|X = 1)$?

Válasz: mivel $\{X = 1\}$ egy nem 0 valószínűségű esemény, ezt a feltételes valószínűség szokásos képletével lehet meghatározni:

$$\begin{aligned}\mathbb{P}(Y < 2|X = 1) &= \frac{\mathbb{P}(Y < 2, X = 1)}{\mathbb{P}(X = 1)} = \frac{\mathbb{P}(Y = 0, X = 1) + \mathbb{P}(Y = 1, X = 1)}{\mathbb{P}(X = 1)} \\ &= \frac{0 + 1/4}{1/2} = \frac{1}{2}.\end{aligned}$$

Ha $x \in \mathbb{R}$ -re $\mathbb{P}(X = x) > 0$, az $y \mapsto \mathbb{P}(Y < y|X = x)$ **feltételes eloszlásfüggvény** úgy viselkedik, mint egy sima eloszlásfüggvény.

Feltételes eloszlások folytonos esetben

Kérdés: Mi a helyzet a feltételes eloszlásfüggvénnyel egy (X, Y) folytonos valószínűségi vektorváltozó esetén?

Probléma: Hogyan definiáljuk $\mathbb{P}(Y < y | X = x)$ -et, ha $x, y \in \mathbb{R}$?
Mivel (X, Y) folytonos, X peremeloszlása is folytonos, így $\{X = x\}$ minden x -re **nulla valószínűségű esemény**, amire feltételezni nem lehet.

Klasszikus feltételes valószínűségekkel nem fogjuk tudni $\mathbb{P}(Y < y | X = x)$ -et definiálni. Viszont van értelmes **formális** definíciója, amit az úgynevezett **feltételes sűrűségfüggvény** segítségével adunk meg.

Feltételes sűrűség- és eloszlásfüggvény

Legyen (X, Y) folytonos valószínűségi vektorváltozó. Együttes sűrűségfüggvényük: $f_{X,Y}$, X marginális sűrűségfüggvénye: f_X .

(11.2.2. Definíció.) Legyen (X, Y) folytonos valószínűségi vektorváltozó, és jelölje együttes sűrűségfüggvényét $f_{X,Y}$. Ekkor Y -nak az X -re vett **feltételes sűrűségfüggvénye**:

$$f_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} f_{X,Y}(x, u) du},$$

olyan $x, y \in \mathbb{R}$ számokra értelmezve, amire $f_X(x) > 0$, és $f_{Y|X}(y|x) = 0$, ha $f_X(x) = 0$.

Példák és intuitív jelentés: hamarosan. Előtte még egy definíció:

(Definíció.) Az előző definíció szerinti szituációban Y -nak az X -re vett **feltételes eloszlásfüggvénye**:

$$F_{Y|X}(y|x) \stackrel{\text{def}}{=} \mathbb{P}(Y < y | X = x) \stackrel{\text{def}}{=} \int_{-\infty}^y f_{Y|X}(z|x) dz,$$

ha $f_X(x) > 0$, és $F_{Y|X}(y|x) = 0$ különben.

Feltételes sűrűség- és eloszlásfüggvény

Legyen (X, Y) folytonos valószínűségi vektorváltozó. Együttes sűrűségfüggvényük: $f_{X,Y}$, X marginális sűrűségfüggvénye: f_X .

(11.2.2. Definíció.) Legyen (X, Y) folytonos valószínűségi vektorváltozó, és jelölje együttes sűrűségfüggvényét $f_{X,Y}$. Ekkor Y -nak az X -re vett **feltételes sűrűségfüggvénye**:

$$f_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} f_{X,Y}(x, u) du},$$

olyan $x, y \in \mathbb{R}$ számokra értelmezve, amire $f_X(x) > 0$, és $f_{Y|X}(y|x) = 0$, ha $f_X(x) = 0$.

Példák és intuitív jelentés: hamarosan. Előtte még egy definíció:

(Definíció.) Az előző definíció szerinti szituációban Y -nak az X -re vett **feltételes eloszlásfüggvénye**:

$$F_{Y|X}(y|x) \stackrel{\text{def}}{=} \mathbb{P}(Y < y | X = x) \stackrel{\text{def}}{=} \int_{-\infty}^y f_{Y|X}(z|x) dz,$$

ha $f_X(x) > 0$, és $F_{Y|X}(y|x) = 0$ különben.

Feltételes sűrűségfüggvény: példa

Az (X, Y) val. vektorváltozó együttes sűrűségfüggvénye

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{10x}e^{-x/10}, & \text{ha } 0 < x < y < 2x, \\ 0, & \text{különben.} \end{cases}$$

Adjuk meg $f_{Y|X}(y|x)$ -et minden $x, y \in \mathbb{R}$ esetén.

Először ki kell számolnunk $f_X(x)$ -et $x \in \mathbb{R}$ -re. Láthatóan

$f_X(x) = 0$, ha $x \leq 0$. $x > 0$ esetén pedig

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,u) du = \int_x^{2x} \frac{1}{10x} e^{-x/10} du = \frac{1}{10x} e^{-x/10} \left[u \right]_x^{2x} \\ &= \frac{1}{10} e^{-x/10}. \end{aligned}$$

Vagyis X peremeloszlása exponenciális $\lambda = 1/10$ paraméterrel.

Ez alapján a fenti definíció szerint

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{1}{x}, & \text{ha } 0 < x < y < 2x, \\ 0, & \text{különben.} \end{cases}$$

Feltételes eloszlásfüggvény: példa

Azt kaptuk, hogy

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{1}{x}, & \text{ha } 0 < x < y < 2x, \\ 0, & \text{különben.} \end{cases}$$

Ez alapján (figyeljünk az integrálási határookra!) ha $x > 0$, akkor

$$F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(z|x) dz = \int_x^y \frac{1}{x} dy = \frac{y-x}{x},$$

ha $y \in (x, 2x)$, valamint $F_{Y|X}(y|x) = 0$, ha $y \leq x$ és

$F_{Y|X}(y|x) = 1$, ha $y \geq 2x$,

és ha $x \leq 0$ (vagyis $f_X(x) = 0$), akkor definíció szerint

$F_{Y|X}(y|x) = 0$ minden $y \in \mathbb{R}$ -re.

Vegyük észre, hogy minden olyan x -re, amelyre $f_X(x) > 0$,

$y \mapsto F_{Y|X}(y|x)$ rendelkezik az eloszlásfüggvény tulajdonságaival (jó határérték $\pm\infty$ -ben, monoton növekvő, balról folyt.),

$y \mapsto f_{Y|X}(y|x)$ pedig a sűrűségfüggvény tulajdonságaival (nemnegatív, integrálja $-\infty$ -től ∞ -ig 1).

Feltételes sűrűségfüggvény a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén

$$f_{X,Y}(x, y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben,} \end{cases}$$

és

$$f_X(x) = \begin{cases} 2x, & \text{ha } 0 < x < 1, \\ 0, & \text{különben.} \end{cases}$$

Tehát

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x},$$

ha $0 < x < 1$ és $0 < y < x$ és $f_{Y|X}(y|x) = 0$ különben.

(Itt $f_{Y|X}(y|x)$ két különböző okból lehet 0: ha $x \in (0, 1)$, vagyis $f_X(x) > 0$, de $y \geq x$, akkor azért, mert $f_{X,Y}(x, y) = 0$. Ha viszont $x \notin (0, 1)$, vagyis $f_X(x) = 0$, akkor $f_{Y|X}(y|x)$ definíció szerint 0.)

Feltételes sűrűségfüggvény a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögon egyenletes eloszlású (X, Y) esetén

$$f_{X,Y}(x,y) = \begin{cases} 2, & \text{ha } 0 < x < 1 \text{ és } 0 < y < x, \\ 0, & \text{különben,} \end{cases}$$

és

$$f_X(x) = \begin{cases} 2x, & \text{ha } 0 < x < 1, \\ 0, & \text{különben.} \end{cases}$$

Tehát

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x},$$

ha $0 < x < 1$ és $0 < y < x$ és $f_{Y|X}(y|x) = 0$ különben.

Vegyük észre a felt. sűrűv. alakja alapján: ha tudjuk, hogy $X = x$, akkor Y egyenletes eloszlású $(0, x)$ -en (hiszen a feltételes sűrűv. a $(0, x)$ intervallumon konstans és mindenhol máshol 0).

Bátrabban fogalmazva: X ismeretében $Y \sim U(0; X)$.

Feltételes sűrűségfüggvény: interpretáció

Tegyük fel, hogy $F_{X,Y}$ (totálisan) differenciálható (x,y) -ban.

Emlékeztető: együttes sűrűv. szemléletes jelentése:

$$f_{X,Y}(x,y) = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x), Y \in (y - \Delta y, y + \Delta y))}{2\Delta x \cdot 2\Delta y}$$

A marginálisé:

$$f_X(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))}{2\Delta x}.$$

Így a feltételesé, ha $f_X(x) > 0$:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x), Y \in (y - \Delta y, y + \Delta y))}{2\Delta x \cdot 2\Delta y}}{\frac{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))}{2\Delta x}} \\ &= \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{1}{2\Delta y} \frac{\mathbb{P}(\{X \in (x - \Delta x, x + \Delta x)\} \cap \{Y \in (y - \Delta y, y + \Delta y)\})}{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))} \\ &= \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y) | X \in (x - \Delta x, x + \Delta x))}{2\Delta y}. \end{aligned}$$

Feltételes sűrűségfüggvény: interpretáció

Azt kaptuk, hogy $f_X(x) > 0$ esetén

$$f_{Y|X}(y|x) = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y) | X \in (x - \Delta x, x + \Delta x))}{2\Delta y}.$$

- ▶ Vagyis a feltételes sűrűv. **feltételes valószínűségek** Δy -nal skálázott határértéke, amint $\Delta x, \Delta y \downarrow 0$.
- ▶ $f_{Y|X}(y|x)$ kb. annak a valószínűsége, hogy Y az y egy kis környezetébe esik, feltéve, hogy X az x egy kis környezetébe esik, osztva az y kis környezetének méretével.
- ▶ Ha X folytonos, $\{X = x\}$ mindig 0 valószínűségű esemény, de ha $f_X(x) > 0$, akkor általában $\Delta x > 0$ -ra $\{X \in (x - \Delta x, x + \Delta x)\}$ már pozitív valószínűségű, amire lehet feltételezni.
- ▶ A feltételes sűrűv. hasonló, mint Y marginális sűrűségfv.-e:

$$f_Y(y) = \lim_{\Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y))}{2\Delta y},$$

de figyelembe veszi az X -re való feltételezést is.

Megjegyzés: események megfeltételezése folytonos X -re

Láttuk: $f_X(x) > 0$ esetén minden $y \in \mathbb{R}$ -re

$$f_{Y|X}(y|x) = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y) | X \in (x - \Delta x, x + \Delta x))}{2\Delta y}.$$

Ezáltal a feltételes **eloszlásfüggvény** interpretációja is világos:

$$\mathbb{P}(Y < y | X = x) = \int_{-\infty}^y f_{Y|X}(z|x) dz = \lim_{\Delta x \downarrow 0} \mathbb{P}(Y < y | X \in (x - \Delta x, x + \Delta x)).$$

Olyan, mint Y eloszlásfüggvénye, csak azzal a feltétellel, hogy X nagyon közel van x -hez.

Hasonlóan nemcsak az $\{Y < y\}$ eseménynek, hanem bármilyen $A \in \mathcal{F}$ eseménynek is értelmezhetjük ugyanilyen feltételes valószínűségét, ha $f_X(x) > 0$:

$$\mathbb{P}(A | X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A | X \in (x - \Delta x, x + \Delta x)).$$

Megfelelő feltételek teljesülése esetén (amiket itt nem részletezünk), $A \mapsto \mathbb{P}(A | X = x)$ egy valószínűségi mérték.

Lásd hamarosan: feltételes várható érték, TVT folytonos esetben.

Megjegyzés: események megfeltételezése folytonos X -re

Láttuk: $f_X(x) > 0$ esetén minden $y \in \mathbb{R}$ -re

$$f_{Y|X}(y|x) = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y) | X \in (x - \Delta x, x + \Delta x))}{2\Delta y}.$$

Ezáltal a feltételes **eloszlásfüggvény** interpretációja is világos:

$$\mathbb{P}(Y < y | X = x) = \int_{-\infty}^y f_{Y|X}(z|x) dz = \lim_{\Delta x \downarrow 0} \mathbb{P}(Y < y | X \in (x - \Delta x, x + \Delta x)).$$

Olyan, mint Y eloszlásfüggvénye, csak azzal a feltétellel, hogy X nagyon közel van x -hez.

Hasonlóan nemcsak az $\{Y < y\}$ eseménynek, hanem bármilyen $A \in \mathcal{F}$ eseménynek is értelmezhetjük ugyanilyen feltételes valószínűségét, ha $f_X(x) > 0$:

$$\mathbb{P}(A | X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A | X \in (x - \Delta x, x + \Delta x)).$$

Megfelelő feltételek teljesülése esetén (amiket itt nem részletezünk), $A \mapsto \mathbb{P}(A | X = x)$ egy valószínűségi mérték.

Lásd hamarosan: feltételes várható érték, TVT folytonos esetben.

Függetlenség és feltételes sűrűségfüggvény

Tegyük fel, hogy X és Y független folytonos valószínűségi változók. Tudjuk: ekkor $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$, minden $x, y \in \mathbb{R}$ esetén.

Ha $x \in \mathbb{R}$ olyan, hogy $f_X(x) > 0$, akkor minden $y \in \mathbb{R}$ -re

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y).$$

Vagyis független X és Y esetén a feltételes sűrűségfüggvény megegyezik a marginálissal, nincs hatása a feltételezésnek.

Folytonos (X, Y) esetén ha ez a feltétel **minden** olyan x -re teljesül, amire $f_X(x) > 0$, abból következik az is, hogy X és Y függetlenek.

Függetlenség: ellenpélda

Az előbbi példában:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{10x}e^{-x/10}, & \text{ha } 0 < x < y < 2x, \\ 0, & \text{különben,} \end{cases}$$

a függetlenség nem áll fenn. Pl.

- ▶ $f_X(50) > 0$,
- ▶ $f_Y(42) > 0$,
- ▶ de $f_{(X,Y)}(50, 42) = 0$.

Az együttes sűrűségfüggvény képlete „látszólag csak x -től függ”, de az, hogy hol nem 0, függ x és y viszonyától!

Feltételes várható érték

(Egy nehéz téma, olykor könnyű vizsgafeladatokkal)

Feltételes várható érték: motiváció

Emlékeztető: feltételes valószínűség, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, ha $\mathbb{P}(B) > 0$.

Most: feltételes várható érték, kezdetnek két kérdés.

1. példa (diszkrét): Két kockával dobunk. Feltéve, hogy az első dobás értéke $k \in [6]$, mennyi a két dobás összegének várható értéke?

Az eddigi ismereteink alapján a kérdésben szereplő „feltételes várható érték” még nem jól definiált. Próbáljunk meg ennek ellenére *intuitíve helyes* választ adni $\rightarrow$ motiváció a későbbi definíciók megadásához is.

Feltételes várható érték: diszkrét bevezető példa

1. példa (diszkrét): Két kockával dobunk. Feltéve, hogy az első dobás értéke $k \in [6]$, mennyi a két dobás összegének várható értéke?

Válasz:

Mivel a két kockadobás eredménye független, az, hogy az első dobás értéke konstans k , semmit nem árul el arról, hogy a második dobásnak mi az eredménye. Ezért a kért **feltételes várható érték** k és egy szabályos kockadobás várható értékének összege, azaz $k + 3,5$.

Figyeljük meg: itt nem elég a várható érték linearitására hivatkozni, kell a két dobás **függetlensége** is.

Ugyanis ha a két kockadobás nem független, akkor az eredmény nem feltétlenül $k + 3,5$. Pl. ha a második kockadobásnak ugyanazt a kockadobást választjuk, mint az első, akkor: feltéve, hogy az 1. dobás eredménye k , az összeg várható értéke $2k$.

Feltételes várható érték: diszkrét bevezető példa

1. példa: Két kockával dobunk. Legyen X az első, Y a második dobás eredménye. Feltéve, hogy az $X = k \in [6]$, $X + Y$ várható értéke $k + 3,5$.

Feltételes várható érték definíciója, ami ezt az eredményt adja?

Láttuk: $\mathcal{F} \rightarrow [0, 1]$, $A \mapsto \mathbb{P}(A|X = k)$ egy valószínűségi mérték, ha $\mathbb{P}(X = k) > 0$. Egy Z valószínűségi változóra legyen

$$\mathbb{E}(Z|X = k) \stackrel{\text{def}}{=} \sum_{i \in \text{Ran}(Z)} i \mathbb{P}(Z = i|X = k), \quad k \in \text{Ran}(X).$$

Vegyük észre, hogy $l = 1, \dots, 6$ esetén

$$\mathbb{P}(X + Y = k + l|X = k) = \mathbb{P}(Y = l|X = k) = \mathbb{P}(Y = l),$$

ahol az utolsó lépés azért igaz, mert X és Y függetlenek. Így

$$\begin{aligned} \mathbb{E}(X + Y|X = k) &= \sum_{i \in \text{Ran}(X+Y)} i \mathbb{P}(X + Y = i|X = k) \\ &= \sum_{l \in \text{Ran}(Y)} (l + k) \mathbb{P}(X + Y = l + k|X = k) = \sum_{l \in \text{Ran}(Y)} (l + k) \mathbb{P}(Y = l) \\ &= \sum_{l \in \text{Ran}(Y)} l \mathbb{P}(Y = l) + k \sum_{l \in \text{Ran}(Y)} \mathbb{P}(Y = l) = \mathbb{E}(Y) + k = 3,5 + k. \end{aligned}$$

Feltételes várható érték: diszkrét bevezető példa

1. példa: Két kockával dobunk. Legyen X az első, Y a második dobás eredménye. Feltéve, hogy az $X = k \in [6]$, $X + Y$ várható értéke $k + 3,5$.

Feltételes várható érték definíciója, ami ezt az eredményt adja?

Láttuk: $\mathcal{F} \rightarrow [0, 1]$, $A \mapsto \mathbb{P}(A|X = k)$ egy valószínűségi mérték, ha $\mathbb{P}(X = k) > 0$. Egy Z valószínűségi változóra legyen

$$\mathbb{E}(Z|X = k) \stackrel{\text{def}}{=} \sum_{i \in \text{Ran}(Z)} i \mathbb{P}(Z = i|X = k), \quad k \in \text{Ran}(X).$$

Vegyük észre, hogy $l = 1, \dots, 6$ esetén

$$\mathbb{P}(X + Y = k + l|X = k) = \mathbb{P}(Y = l|X = k) = \mathbb{P}(Y = l),$$

ahol az utolsó lépés azért igaz, mert X és Y függetlenek. Így

$$\begin{aligned} \mathbb{E}(X + Y|X = k) &= \sum_{i \in \text{Ran}(X+Y)} i \mathbb{P}(X + Y = i|X = k) \\ &= \sum_{l \in \text{Ran}(Y)} (l + k) \mathbb{P}(X + Y = l + k|X = k) = \sum_{l \in \text{Ran}(Y)} (l + k) \mathbb{P}(Y = l) \\ &= \sum_{l \in \text{Ran}(Y)} l \mathbb{P}(Y = l) + k \sum_{l \in \text{Ran}(Y)} \mathbb{P}(Y = l) = \mathbb{E}(Y) + k = 3,5 + k. \end{aligned}$$

Feltételes várható érték: diszkrét eset

A példában látottakat és definiáltakat most fogalmazzuk meg általában a diszkrét esetben.

(11.1.1. Definíció.) Legyen Y valószínűségi változó és A olyan esemény, amire $\mathbb{P}(A) > 0$. Ekkor az Y -nak az A -ra vett **feltételes várható értéke** az Y változó $\mathbb{P}(\cdot|A)$ valószínűségi mérték szerinti várható értéke. Jelölés: $\mathbb{E}(Y|A)$.

Interpretáció: $\mathbb{E}(Y|A)$ jelentése az Y átlagos értéke a teljes eseménytér helyett az A eseményre szorítkozva.

(11.1.2. Lemma.) Legyen Y egyszerű (azaz véges értékészletű) valószínűségi változó és A esemény, amelyre $\mathbb{P}(A) > 0$. Ekkor

$$\mathbb{E}(Y|A) = \sum_{k \in \text{Ran}(Y)} k \mathbb{P}(Y = k|A).$$

(A bizonyítás hasonló, mint $\mathbb{E}(Y)$ képletéé.)

Feltételes várható érték: diszkrét eset

Feltételes várható érték visszavezetése a várható érték képletére:

(11.1.3. Állítás.) Legyen Y valószínűségi változó, és $\mathbb{P}(A) > 0$.

Ekkor $\mathbb{E}(Y|A) = \frac{\mathbb{E}(Y \cdot \mathbb{1}_A)}{\mathbb{P}(A)}$.

Itt az $Y \cdot \mathbb{1}_A$ valószínűségi változót a valószínűségi változók szokásos ω -nkénti szorzásával definiáljuk:

$$(Y \cdot \mathbb{1}_A)(\omega) = Y(\omega)\mathbb{1}_A(\omega) = \begin{cases} Y(\omega), & \text{ha } \omega \in A, \\ 0, & \text{különben.} \end{cases}$$

Példa

Legyen újra X és Y két független kockadobás eredménye.

$\mathbb{P}(Y = k | X + Y < 7) = ? \quad (k = 1, \dots, 6)$

$$\begin{aligned}\mathbb{P}(Y = k | X + Y < 7) &= \frac{\mathbb{P}(Y = k, X + Y < 7)}{\mathbb{P}(X + Y < 7)} = \frac{\mathbb{P}(Y = k, X < 7 - k)}{\frac{15}{36}} \\ &= \frac{\frac{1}{6} \cdot \frac{7-k-1}{6}}{\frac{15}{36}} = \frac{6-k}{15}.\end{aligned}$$

Ezért

$$\begin{aligned}\mathbb{E}(Y | X + Y < 7) &= \sum_{k=1}^6 k \mathbb{P}(Y = k | X + Y < 7) = \sum_{k=1}^6 k \frac{6-k}{15} \\ &= \frac{6}{15}(1 + \dots + 6) - \frac{1}{15}(1 + 4 + \dots + 36) = \frac{6 \cdot 21}{15} - \frac{91}{15} = \frac{35}{15} = \frac{7}{3}.\end{aligned}$$

Példa

Legyen újra X és Y két független kockadobás eredménye.

$\mathbb{P}(Y = k | X + Y < 7) = ?$ ($k = 1, \dots, 6$)

$$\begin{aligned}\mathbb{P}(Y = k | X + Y < 7) &= \frac{\mathbb{P}(Y = k, X + Y < 7)}{\mathbb{P}(X + Y < 7)} = \frac{\mathbb{P}(Y = k, X < 7 - k)}{\frac{15}{36}} \\ &= \frac{\frac{1}{6} \cdot \frac{7-k-1}{6}}{\frac{15}{36}} = \frac{6-k}{15}.\end{aligned}$$

Ezért

$$\begin{aligned}\mathbb{E}(Y | X + Y < 7) &= \sum_{k=1}^6 k \mathbb{P}(Y = k | X + Y < 7) = \sum_{k=1}^6 k \frac{6-k}{15} \\ &= \frac{6}{15}(1 + \dots + 6) - \frac{1}{15}(1 + 4 + \dots + 36) = \frac{6 \cdot 21}{15} - \frac{91}{15} = \frac{35}{15} = \frac{7}{3}.\end{aligned}$$

Példa, folytatás

Mivel $X + Y$ diszkrét, bizonyos $\{X + Y = k\}$ alakú események valószínűsége is pozitív. Ezért például $\mathbb{E}(Y|X + Y = 5)$ is értelmes, értéke

$$\begin{aligned}\mathbb{E}(Y|X + Y = 5) &= \sum_{k=1}^6 k \mathbb{P}(Y = k|X + Y = 5) \\&= \sum_{k=1}^6 k \frac{\mathbb{P}(Y = k, X + Y = 5)}{\mathbb{P}(X + Y = 5)} \\&= \sum_{k=1}^6 k \frac{\mathbb{P}(Y = k) \mathbb{P}(X = 5 - k)}{\mathbb{P}(X + Y = 5)} \\&= \sum_{k=1}^4 k \frac{\frac{1}{6} \cdot \frac{1}{6}}{\frac{4}{36}} = \frac{10}{4} = 2,5.\end{aligned}$$

Az eredmény nem meglepő: feltéve, hogy $X + Y = 5$, X átlagosan 2,5, mivel X és Y független, azonos eloszlású val. változók.

Feltételes várható érték: diszkrét eset

A példára visszatérve: most, hogy tudjuk, hogy

$$\mathbb{E}(X + Y|X = k) = k + 3,5, \quad k = 1, \dots, 6,$$

kérdezzünk egy kicsit még durvábbat.

Mennyi $X + Y$ várható értéke, **feltéve X -et** (jelölés: $\mathbb{E}(X + Y|X)$)?

Mivel X véletlen, erre már nem egy szám lesz a válasz. A válasz egy **val. változó** lesz, " $X + Y$ legjobb közelítése X ismeretében".

$X = k$ ismeretében $X + Y$ feltételes várható értéke $k + 3,5$.

Tehát a válasz $\mathbb{E}(X + Y|X) = X + 3,5$ lesz.

Feltételes várható érték diszkrét esetben

(11.1.5. Definíció.) Legyenek X és Y valószínűségi változók, ahol X diszkrét. Jelölje S_X az X *lényeges értékeinek* a halmazát, azaz

$$S_X \stackrel{\text{def.}}{=} \{x \in \mathbb{R} | \mathbb{P}(X = x) > 0\}.$$

Tekintsük a $g: S_X \rightarrow \mathbb{R}, g(x) = \mathbb{E}(Y|X = x)$ valós függvényt.

Ekkor az Y -nak az X -re vett **regressziója** (**feltételes várható értéke**) a $g(X)$ valószínűségi változó. Jelölése: $\mathbb{E}(Y|X)$.

Feltételes várható érték: diszkrét eset

A példára visszatérve: most, hogy tudjuk, hogy

$$\mathbb{E}(X + Y|X = k) = k + 3,5, \quad k = 1, \dots, 6,$$

kérdezzünk egy kicsit még durvábbat.

Mennyi $X + Y$ várható értéke, **feltéve X -et** (jelölés: $\mathbb{E}(X + Y|X)$)?

Mivel X véletlen, erre már nem egy szám lesz a válasz. A válasz egy **val. változó** lesz, “ $X + Y$ legjobb közelítése X ismeretében” .
 $X = k$ ismeretében $X + Y$ feltételes várható értéke $k + 3,5$.

Tehát a válasz $\mathbb{E}(X + Y|X) = X + 3,5$ lesz.

Feltételes várható érték diszkrét esetben

(11.1.5. Definíció.) Legyenek X és Y valószínűségi változók, ahol X diszkrét. Jelölje S_X az X lényeges értékeinek a halmazát, azaz

$$S_X \stackrel{\text{def.}}{=} \{x \in \mathbb{R} | \mathbb{P}(X = x) > 0\}.$$

Tekintsük a $g: S_X \rightarrow \mathbb{R}, g(x) = \mathbb{E}(Y|X = x)$ valós függvényt.

Ekkor az Y -nak az X -re vett **regressziója** (**feltételes várható értéke**) a $g(X)$ valószínűségi változó. Jelölése: $\mathbb{E}(Y|X)$.

Feltételes várható érték: még egy diszkrét példa

(11.1.6. Példa.) Dobunk egy szabályos kockával, majd az eredménynek megfelelő számú szabályos érmével.

Legyen X a kockadobás értéke, Y pedig a fejek száma az érmedobások között. Mi Y regressziója X -re?

Ha X értékét tudjuk, akkor Y értéke binomiális eloszlású X és $1/2$ paraméterrel.

Tehát: $\mathbb{E}(Y|X = x) = x \cdot \frac{1}{2} =: g(x)$, vagyis $\mathbb{E}(Y|X) = g(X) = \frac{X}{2}$.

Általában: regresszió=feltételes várható érték (egy val. változóra nézve) $\neq$ lineáris regresszió.

Lin. regresszió: Y legjobb közelítése X egy lineáris függvényével.

Regresszió: nem szorítkozik lineáris függvényekre, vagyis azt a $g(X)$ függvényt adja vissza, ami Y -t X ismeretében „a legjobban közelíti”. Milyen értelemben? $\rightarrow$ köv. dia.

Feltételes várható érték: még egy diszkrét példa

(11.1.6. Példa.) Dobunk egy szabályos kockával, majd az eredménynek megfelelő számú szabályos érmével.

Legyen X a kockadobás értéke, Y pedig a fejek száma az érmedobások között. Mi Y regressziója X -re?

Ha X értékét tudjuk, akkor Y értéke binomiális eloszlású X és $1/2$ paraméterrel.

Tehát: $\mathbb{E}(Y|X = x) = x \cdot \frac{1}{2} =: g(x)$, vagyis $\mathbb{E}(Y|X) = g(X) = \frac{X}{2}$.

Általában: **regresszió=feltételes várható érték (egy val. változóra nézve) $\neq$ lineáris regresszió.**

Lin. regresszió: Y legjobb közelítése X egy **lineáris** függvényével.

Regresszió: nem szorítkozik lineáris függvényekre, vagyis azt a $g(X)$ függvényt adja vissza, ami Y -t X ismeretében „a legjobban közelíti”. Milyen értelemben? $\rightarrow$ köv. dia.

A regresszió mint legjobb közelítés

Intuitívan fogalmazva, az $\mathbb{E}(Y|X)$ valószínűségi változó „mindent tud az Y -ról, amit X ismeretében tudni lehet”, és ezzel a „lehető legtöbb” információval ad közelítést Y -ra.

Pecízen megfogalmazva: ha további, X -re vonatkozó feltétel esetén nézzük Y feltételes várható értékét, akkor az már kiszámolható $\mathbb{E}(Y|X)$ regresszióból is, nem kell hozzá az eredeti Y -t ismernünk.

(11.1.7. Állítás.) Legyen X diszkrét valószínűségi változó. Tegyük fel, hogy $\mathbb{P}(X < x) > 0$ és $\mathbb{E}(|Y|) < \infty$. Ekkor

$$\mathbb{E}[g(X)|X < x] = \mathbb{E}(Y|X < x)$$

minden $x \in \mathbb{R}$ esetén, ahol $g(X) = \mathbb{E}(Y|X)$.

(Bizonyítás: lásd előadás)

Regresszió folytonos esetben: első példa

2. példa (folytonos): Számítógéppel szimulálunk egy $X \sim U(0; 1)$ eloszlású val. változót. Valaki megmondja nekünk, hogy az eredmény nagyobb lett, mint $1/2$. Ezt feltéve mennyi a val. változó várható értéke?

Heurisztikus válasz: mivel X egyenletes eloszlású, a sűrűségfüggvénye a $(0, 1)$ intervallumon konstans 1. Ezért feltéve, hogy $X > 1/2$, még mindig konstans lesz a sűrűségfüggvénye, most az $(1/2, 1)$ intervallumon (most konstans 2, hogy az integrálja 1 legyen). Így $\mathbb{E}(X|X > 1/2)$ az intervallum közepe, azaz $3/4$.

Rendes válasz: mivel $\mathbb{P}(X > 1/2) = \frac{1}{2} > 0$, az előbbi definíció szerint

$$\begin{aligned}\mathbb{E}(X|X > 1/2) &= \frac{\mathbb{E}(X \cdot \mathbb{1}_{\{X \in (1/2, 1)\}})}{\mathbb{P}(X > 1/2)} = \frac{\int_{1/2}^1 x f_X(x) dx}{1/2} = \frac{\int_{1/2}^1 x dx}{1/2} \\ &= \frac{1/2 - 1/8}{1/2} = \frac{3}{4}.\end{aligned}$$

Regresszió folytonos esetben: első példa

2. példa (folytonos): Számítógéppel szimulálunk egy $X \sim U(0; 1)$ eloszlású val. változót. Valaki megmondja nekünk, hogy az eredmény nagyobb lett, mint $1/2$. Ezt feltéve mennyi a val. változó várható értéke?

Heurisztikus válasz: mivel X egyenletes eloszlású, a sűrűségfüggvénye a $(0, 1)$ intervallumon konstans 1. Ezért feltéve, hogy $X > 1/2$, még mindig konstans lesz a sűrűségfüggvénye, most az $(1/2, 1)$ intervallumon (most konstans 2, hogy az integrálja 1 legyen). Így $\mathbb{E}(X|X > 1/2)$ az intervallum közepe, azaz $3/4$.

Rendes válasz: mivel $\mathbb{P}(X > 1/2) = \frac{1}{2} > 0$, az előbbi definíció szerint

$$\begin{aligned}\mathbb{E}(X|X > 1/2) &= \frac{\mathbb{E}(X \cdot \mathbb{1}_{\{X \in (1/2, 1)\}})}{\mathbb{P}(X > 1/2)} = \frac{\int_{1/2}^1 x f_X(x) dx}{1/2} = \frac{\int_{1/2}^1 x dx}{1/2} \\ &= \frac{1/2 - 1/8}{1/2} = \frac{3}{4}.\end{aligned}$$

Regresszió folytonos esetben: első példa

2. példa (folytonos): Számítógéppel szimulálunk egy $X \sim U(0; 1)$ eloszlású val. változót. Valaki megmondja nekünk, hogy az eredmény nagyobb lett, mint $1/2$. Ezt feltéve mennyi a val. változó várható értéke?

Heurisztikus válasz: mivel X egyenletes eloszlású, a sűrűségfüggvénye a $(0, 1)$ intervallumon konstans 1. Ezért feltéve, hogy $X > 1/2$, még mindig konstans lesz a sűrűségfüggvénye, most az $(1/2, 1)$ intervallumon (most konstans 2, hogy az integrálja 1 legyen). Így $\mathbb{E}(X|X > 1/2)$ az intervallum közepe, azaz $3/4$.

Rendes válasz: mivel $\mathbb{P}(X > 1/2) = \frac{1}{2} > 0$, az előbbi definíció szerint

$$\begin{aligned}\mathbb{E}(X|X > 1/2) &= \frac{\mathbb{E}(X \cdot \mathbb{1}_{\{X \in (1/2, 1)\}})}{\mathbb{P}(X > 1/2)} = \frac{\int_{1/2}^1 x f_X(x) dx}{1/2} = \frac{\int_{1/2}^1 x dx}{1/2} \\ &= \frac{1/2 - 1/8}{1/2} = \frac{3}{4}.\end{aligned}$$

Regresszió folytonos esetben

Tegyük fel most, hogy X folytonos val. változó. Egy Y val. változóra szeretnénk definiálni $\mathbb{E}(Y|X)$ -et, mint „ Y legjobb közelítését X ismeretében”, a diszkrét esethez hasonlóan.

Diszkrét Z esetén a pár diával ezelőtti definíció szerint:

$$\mathbb{E}[Y|Z = z] = \sum_{k \in \text{Ran}(Y)} k \mathbb{P}(Y = k | Z = z),$$

ha $\mathbb{P}(Z = z) > 0$.

Probléma: Folytonos X esetén minden $x \in \mathbb{R}$ -re $\mathbb{P}(X = x) = 0$.

Emiatt a fenti képlettel, illetve klasszikus feltételes valószínűségekkel $\mathbb{E}(Y|X = x)$ **nincs definiálva!**

Viszont ugyanezért vezettük be a **feltételes sűrűségfüggvényt**, abban az esetben, amikor (X, Y) folytonos val. vektorváltozó:

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)},$$

ha $f_X(x) > 0$ és 0 különben.

Regresszió folytonos esetben és feltételes sűrűségfüggvény

Feltételes sűrűségfüggvény:

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)},$$

ha $f_X(x) > 0$ és 0 különben.

Korábban (informálisan) azt mondtuk, hogy ez olyan, mint $f_Y(y)$, csak arra feltételezve, hogy X közel van x -hez.

Tudjuk: $\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$. Ezért intuitíve azt várjuk, hogy

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

jó definíció lesz. (Ez ismét **formális** jelölés, nem igazi feltételes valószínűség).

Vezessük be a $g(x) = \mathbb{E}(Y|X = x)$ jelölést. A diszkrét esetben azt láttuk, hogy $\mathbb{E}(Y|X) = g(X)$.

Kérdés: Ugyanez igaz általánosabb (pl. folytonos) esetben is? Mi ilyen esetben egyáltalán a regresszió definíciója?

Regresszió általános esetben

(11.2.1. Definíció.) Legyenek X és Y **tetszőleges** valószínűségi változók. Az Y -nak az X -re vett **regressziója** (**feltételes várható értéke**) az a $g(X)$ alakú valószínűségi változó, amire

$$\mathbb{E}(g(X)|X < x) = \mathbb{E}(Y|X < x) \quad (2)$$

minden olyan $x \in \mathbb{R}$ esetén, amire $\mathbb{P}(X < x) > 0$.

A $g(X)$ valószínűségi változó jelölése: $\mathbb{E}(Y|X)$.

- ▶ A $\mathbb{P}(X < x) > 0$ feltétel azért kell, hogy $\mathbb{E}(\cdot|X < x)$, illetve $\mathbb{E}(g(X)|X < x)$ definiált legyen a 11.1.1. Definíció szerint.
- ▶ g -nek teljesítenie kell bizonyos mértékelméleti feltételeket (hogy $\mathbb{E}(g(X)|X < x)$ definiált legyen), de ezzel itt nem foglalkozunk.
- ▶ „A regresszió 1 valószínűséggel egyértelmű.”

Állítás: ha g mellett egy $\tilde{g}$ függvényre is teljesül (2), akkor $\mathbb{P}(g(X) = \tilde{g}(X)) = 1$. (Nem bizonyítjuk.)

$g(x)$ -et „büntetlenül” megváltoztathatjuk „néhány” (pl. megszámlálható sok) olyan x pontban, amelyre $\mathbb{P}(X = x) = 0$. Olyan x pontban, ahol $\mathbb{P}(X = x) > 0$, nem változtatható meg.

Regresszió általános esetben

(11.2.1. Definíció.) Legyenek X és Y **tetszőleges** valószínűségi változók. Az Y -nak az X -re vett **regressziója** (**feltételes várható értéke**) az a $g(X)$ alakú valószínűségi változó, amire

$$\mathbb{E}(g(X)|X < x) = \mathbb{E}(Y|X < x) \quad (2)$$

minden olyan $x \in \mathbb{R}$ esetén, amire $\mathbb{P}(X < x) > 0$.

A $g(X)$ valószínűségi változó jelölése: $\mathbb{E}(Y|X)$.

Jelölés: $g(x) := \mathbb{E}(Y|X = x)$. g neve: **regressziós függvény**.

Ez egy **jelölés**.

- ▶ Diszkrét X esetén megfelel az eddig tárgyalt $g(x)$ függvénynek, tehát $g(x)$ az Y -nak az $A \mapsto \mathbb{P}(A|X = x)$ módon definiált valószínűségi mérték szerinti várható értéke.
- ▶ Folytonos esetben $\mathbb{P}(A|X = x)$ -nek nincs értelme feltételes valószínűségként, de **jelölésként** korábban már bevezettük: ha $f_X(x) > 0$,

$$\mathbb{P}(A|X = x) = \lim_{\Delta x \downarrow 0} \mathbb{P}(A \cap \{X \in (x - \Delta x, x + \Delta x)\}).$$

Feltételes sűrűségfüggvény és regresszió

$$f_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,u)du}, \quad f_X(x) > 0.$$

Most kimondjuk a már beharangozott állítást arról, hogy segítségével hogyan adható meg a regressziós függvény.

(11.2.3. Állítás.) Legyen (X, Y) folytonos valószínűségi vektorváltozó. Ekkor

$$g(x) = \mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)dy$$

az Y -nak az X -re vett regressziós függvénye.

(Bizonyítás: lásd jegyzet.)

Így tehát Y -nak az X -re vett regressziója valóban $\mathbb{E}(Y|X) = g(X)$.

Feltételes sűrűv. és regresszió: interpretáció

$$g(x) = \mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) dy.$$

Vázlatos indoklás arra, hogy ez miért igaz: $f_X(x) > 0$ esetén

$$f_{Y|X}(y|x) = \lim_{\Delta x \downarrow 0, \Delta y \downarrow 0} \frac{\mathbb{P}(Y \in (y - \Delta y, y + \Delta y) | X \in (x - \Delta x, x + \Delta x))}{2\Delta y}.$$

Ha $f_X(x) > 0$ és használjuk a korábban már bevezetett

$$\mathbb{P}(A|X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x))$$

jelölést, akkor megmutatható:

- ▶ $A \mapsto \mathbb{P}(A|X = x)$ tényleg egy valószínűségi mérték.
- ▶ Ha ezt a val. mértéket tekintjük (az eredeti $\mathbb{P}$ helyett), akkor Y folytonos val. változó, és sűrűségfüggvénye $y \mapsto f_{Y|X}(y|x)$.
- ▶ Tehát az Y val. változó ezen val. mérték szerinti **várható értéke** $g(x) = \mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$ lesz (ezt mondja ki gyakorlatilag a 11.2.3. Állítás).
- ▶ Ebből az $x \mapsto g(x)$ függvény megkapható, végül Y regressziója X -re $g(X)$ lesz.

Példa

Egy rossz állapotú vasúti mellékvonalon a motorvonatra $X \sim \text{Exp}(1/10)$ percig kell várunk, és $X = x$ ismeretében a célállomásunkig tartó utazás $U(x, 2x)$ ideig tart.

- ▶ Határozzuk meg $f_{Y|X}$ -et és $f_{X,Y}$ -t.
- ▶ Határozzuk meg Y -nak az X -re való g regressziós függvényét.

Mivel $X \sim \text{Exp}(1/10)$,

$$f_X(x) = \frac{1}{10} \cdot e^{-x/10},$$

ha $x > 0$ és $f_X(x) = 0$ különben. Továbbá $X = x$ ismeretében $Y \sim U(x, 2x)$, vagyis a feltételes sűrűségfüggvény

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2x-x} = \frac{1}{x}, & \text{ha } x < y < 2x, \\ 0, & \text{különben.} \end{cases}$$

Így $f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x)$ (ha $f_X(x) > 0$) miatt

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{10x} e^{-x/10}, & \text{ha } 0 < x < y < 2x, \\ 0, & \text{különben.} \end{cases}$$

Példa: folytatás

Mivel

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2x-x} = \frac{1}{x}, & \text{ha } x < y < 2x, \\ 0, & \text{különben,} \end{cases}$$

azt kapjuk, hogy

$$g(x) = \mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \int_x^{2x} \frac{y}{x} dy = \frac{1}{x} \left[y^2/2 \right]_x^{2x} = \frac{3}{2}x.$$

Tehát

$$\mathbb{E}(Y|X) = g(X) = \frac{3}{2}X.$$

Interpretáció: Mivel X ismeretében $Y \sim U(X, 2X)$, $\mathbb{E}(Y|X)$ az $(X, 2X)$ intervallum középpontja, azaz $\frac{3}{2}X$.

Az ilyen intuitív érvelés jó, de könnyen hiányos lehet – vizsgán inkább csak ellenőrzésnek használjuk és számoljuk ki rendesen is.

Regresszió a háromszöges példában

Láttuk: a $(0, 0), (1, 0), (1, 1)$ csúcspontú háromszögön egyenletes eloszlású (X, Y) esetén

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x},$$

ha $0 < x < 1$ és $0 < y < x$ és $f_{Y|X}(y|x) = 0$ különben.

Legyen most x olyan, hogy $f_X(x) > 0$, vagyis $0 < x < 1$. Ekkor (megint figyeljünk az integrálási határookra!)

$$g(x) = \mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \int_0^x y \cdot \frac{1}{x} dy = \frac{1}{x} [y^2/2]_{y=0}^x = \frac{x}{2}.$$

Ez nem meglepő, hiszen már láttuk: $X = x$ ismeretében Y egyenletes eloszlású a $(0, x)$ intervallumon. Ennek pedig a várható értéke az intervallum középpontja, $\frac{x}{2}$.

Így tehát $\mathbb{E}(Y|X) = g(X) = \frac{X}{2}$. (X ismeretében $Y \sim U(0; X)$.)

A feltételes várható érték tulajdonságai

A most következő tulajdonságok diszkrét és folytonos (és bármilyen) val. változók esetén is igazak.

(11.3.1. Állítás.) Legyenek X, Y, Z valószínűségi változók. Ekkor teljesülnek a következők, feltéve, hogy a jobb oldalukon szereplő mennyiségek léteznek és végesek:

1. **Linearitás:** Tetszőleges $a, b \in \mathbb{R}$ esetén
$$\mathbb{E}(aY + bZ|X) = a\mathbb{E}(Y|X) + b\mathbb{E}(Z|X).$$
2. **Az ismert val. változó függvényei kiemelhetők**, *taking out what is known*: Tetszőleges h folytonos függvény (vagy bármilyen olyan h függvény, amelyre $h(X)$ is val. változó) esetén

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

3. Ha X és Y függetlenek, akkor $\mathbb{E}(Y|X) = \mathbb{E}(Y)$.

A két független kockadobás, X és Y példájában:

$$\mathbb{E}(X + Y|X) = \mathbb{E}(X \cdot 1 + Y|X) = X + \mathbb{E}(Y) = X + 3,5.$$

Példa: egy vizsgafeladat (össznépi gyak., 10.8. feladat)

2022. december 22-i vizsga 5. feladata:

Legyenek $X \sim N(2; 4)$ és $Y \sim U(0, 2)$ független valószínűségi változók, és legyen

$$Z = 2X \ln Y + X^2 Y^2 - 3Y$$

Számoljuk ki az $\mathbb{E}(Z \mid Y)$ regressziót.

(Megoldás: lásd előadás)

Tipp: mielőtt elkezdenénk a vizsgán bármilyen feltételes sűrűségfüggvényt meghatározni, gondoljuk át, hogy ez tényleg szükséges-e. Itt nem szükséges (és a felt. sűrűv. túl csúnya ahhoz, hogy a vizsga ideje alatt kiszámoljuk), más feladatokban igen (és ott szebb is).

Regresszió: optimalitási kritérium és lin. regresszió

(11.3.2. Állítás.) Tegyük fel, hogy $\mathbb{E}(Y^2)$ véges. Ekkor az

$$\mathbb{E}((Y - g(X))^2)$$

várható érték pontosan akkor minimális a g függvényben, ha $g(X)$ és $\mathbb{E}(Y|X)$ majdnem biztosan (azaz 1 valószínűséggel) megegyeznek.

Emlékeztető: lineáris regresszió esetén ugyanezt a várható értéket minimalizáltuk g -ben, csak ott kizárólag a lineáris függvények körében.

Előfordulhat-e, hogy a regresszió megegyezik a lin. regresszióval? Néha igen. Pl. 11.1.6. Példa: dobunk egy szabályos kockával, majd annyi szabályos érmével, ahányat dobtunk. X : dobás értéke, Y : fejek száma.

Láttuk: $\mathbb{E}(Y|X) = X/2$. Ez lineáris függvény, így a fenti tulajdonságok miatt ez a lin. regresszió is ($\beta = 1/2, \alpha = 0$).

(Aki nem hiszi, járjon utána!)

Másik példa: a háromszögön vett egyenletes eloszlásnál $\mathbb{E}(Y|X)$ -re és $\beta X + \alpha$ -ra is $X/2$ adódott. **Ellenpélda:** 10.5. feladat.

Teljes várható érték tétele

Példa: A vonatos példában hogyan számoljuk ki Y várható értékét?
Persze ki lehet számolni $f_{X,Y}$ vagy f_Y ismeretében a szokásos módon.

Egyszerűbb/elegánsabb megoldás: az alábbi tétel segítségével.

(11.3.4. Állítás, **teljes várható érték tétele**.) Legyen X és Y valószínűségi változó, amikre $\mathbb{E}(|X|)$ és $\mathbb{E}(|Y|)$ véges. Ekkor

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

Az állítás ismert **toronyszabály** néven is (más nyelveken is gyakran *torony*-, de inkább *tulajdonság*, mint *szabály*, pl. angolul **tower property**, németül **Turmeigenschaft**).

A példában: $\mathbb{E}(Y|X) = \frac{3}{2}X$, így mivel $X \sim \text{Exp}(1/10)$,
 $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = 3/2 \cdot 10 = 15$ perc.

Teljes várható érték tétele: diszkrét eset

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

Kérdés: mi köze ennek a teljes valószínűség tételéhez?

Diszkrét esetben a külső valószínűségi változót kifejezve a TVT-vel analóg alakot kapunk:

(11.3.5. Állítás, teljes várható érték tétele, diszkrét eset.) Legyen X diszkrét valószínűségi változó, és $\{x_1, \dots, x_n, \dots\}$ az X értékkészletének azon pontjai, amelyekre $\mathbb{P}(X = x_i) > 0$. Ha $\mathbb{E}(|Y|) < \infty$, akkor

$$\mathbb{E}(Y) = \sum_i \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i).$$

Teljes eseményrendszerre kimondva még inkább TVT-szerű:

(11.3.7. Állítás: teljes várható érték tétele, teljes eseményrendszerrel.) Legyen $A_1, \dots, A_n$ teljes eseményrendszer Ω -n, amire $\mathbb{P}(A_i) > 0$ minden i -re. Ha $\mathbb{E}(|Y|) < \infty$, akkor

$$\mathbb{E}(Y) = \sum_{i=1}^n \mathbb{E}(Y|A_i)\mathbb{P}(A_i)$$

Teljes várható érték tétele: diszkrét eset

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

Kérdés: mi köze ennek a teljes valószínűség tételéhez?

Diszkrét esetben a külső valószínűségi változót kifejezve a TVT-vel analóg alakot kapunk:

(11.3.5. Állítás, teljes várható érték tétele, diszkrét eset.) Legyen X diszkrét valószínűségi változó, és $\{x_1, \dots, x_n, \dots\}$ az X értékkészletének azon pontjai, amelyekre $\mathbb{P}(X = x_i) > 0$. Ha $\mathbb{E}(|Y|) < \infty$, akkor

$$\mathbb{E}(Y) = \sum_i \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i).$$

Teljes eseményrendszerre kimondva még inkább TVT-szerű:

(11.3.7. Állítás: teljes várható érték tétele, teljes eseményrendszerrel.) Legyen $A_1, \dots, A_n$ teljes eseményrendszer Ω -n, amire $\mathbb{P}(A_i) > 0$ minden i -re. Ha $\mathbb{E}(|Y|) < \infty$, akkor

$$\mathbb{E}(Y) = \sum_{i=1}^n \mathbb{E}(Y|A_i)\mathbb{P}(A_i).$$

Teljes várható érték tétele: diszkrét és folytonos eset

Diszkrét eset:

$$\mathbb{E}(Y) = \sum_i \mathbb{E}(Y|X = x_i) \mathbb{P}(X = x_i).$$

(Példa: geometriai eloszlás szórása egyszerűen, 11.3.6. Példa)

Folytonos esetben a $p_X(x_i) = \mathbb{P}(X = x_i)$ súlyfüggvényt az f_X sűrűségfüggvény, a $g(x_i) = \mathbb{E}(Y|X = x_i)$ -t, azaz a

$\mathbb{P}(Y = \cdot | X = x_i)$ **feltételes súlyfüggvény** szerinti várható értéket pedig a feltételes sűrűségfüggvény szerinti v.é. helyettesíti:

(11.3.8. Állítás, teljes várható érték tétele, folytonos eset.) Legyen X folytonos valószínűségi változó, $\mathbb{E}(|Y|) < \infty$, és jelölje X sűrűségfüggvényét f_X . Ekkor

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) f_X(x) dx.$$

Az együttesen folytonos esetben (azaz ha (X, Y) folytonos) tehát

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \right) f_X(x) dx.$$

Teljes valószínűség tétele folytonos esetben

A TVT-t eddig csak teljes eseményrendszerre fogalmaztuk meg. Diszkrét val. változóra is megfogalmazható: legyen X diszkrét val. változó és $S_X = \{k \in \mathbb{R} : \mathbb{P}(X = k) > 0\}$, ekkor $\forall A$ eseményre

$$\mathbb{P}(A) = \sum_{k \in S_X} \mathbb{P}(X = k) \mathbb{P}(A|X = k).$$

Ez speciális esete az eredeti TVT-nek (ami megszámlálhatóan végtelen sok eseményre ugyanúgy működik, mint véges sokra).

Szeretnénk a TVT-t folytonos X esetére is általánosítani.

Emlékeztető: $\mathbb{P}(A) = \mathbb{E}(\mathbf{1}_A)$ bármilyen A eseményre.

Folytonos eset: mi $\mathbb{P}(A|X = x)$ (formális def.)?

(12.1.1. Definíció). Legyen X valószínűségi változó és A esemény.

Ekkor A -nak az X -re vett feltételes valószínűsége az

$x \mapsto \mathbb{E}(\mathbf{1}_A|X = x)$ regressziós függvény. Jelölése: $\mathbb{P}(A|X = x)$.

Ez a definíció jogos kételyeket ébreszthet $\rightarrow$ lásd 2–3 dia múlva.

Teljes valószínűség tétele folytonos esetben

Diszkrét X esete:

$$\mathbb{P}(A) = \sum_{k \in S_X} \mathbb{P}(X = k) \mathbb{P}(A|X = k).$$

A TVT folytonos analógiája az, hogy az X súlyfüggvényét a sűrűségfüggvényére, a feltételes valószínűséget a (formális) feltételes valószínűségre és a szummát integrálra cseréljük:

(12.1.2. Tétel, teljes valószínűség tétele folytonos esetben, röviden „folytonos TVT”.) Legyen X folytonos valószínűségi változó és A esemény. Ekkor

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x) f_X(x) dx,$$

ahol f_X az X sűrűségfüggvénye.

Bizonyítás (itt a diasoron, mert nincs benne a jegyzetben)

Tekintsük az $x \mapsto g(x) = \mathbb{P}(A|X = x) = \mathbb{E}(\mathbb{1}_A|X = x)$ regressziós függvényt. Ekkor a regresszió(s függvény) definíciója/jelölése szerint $g(X) = \mathbb{E}(\mathbb{1}_A|X)$.

A transzformált várható értékére vonatkozó képlet szerint

$$\int_{-\infty}^{\infty} \mathbb{P}(A|X = x) f_X(x) dx = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \mathbb{E}(g(X)) = \mathbb{E}(\mathbb{E}(\mathbb{1}_A|X)).$$

Végül a teljes várható érték tétele szerint

$$\mathbb{E}(\mathbb{E}(\mathbb{1}_A|X)) = \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A),$$

amivel kész a bizonyítás.

Teljes valószínűség tétele folytonos esetben: példa

Még nem volt szó arról, hogy $\mathbb{P}(A|X = x)$ -et általában hogyan számolhatjuk ki.

Egyes feladatokban a feladat szövege megadja. Pl.: össznépi gyakorlat, 11.20. feladat. (A 19., 21. is hasonló.)

Cimucka bácsi túrkevei telkén egy tő kukoricát nevelget. Túrkeve Magyarország legszárazabb területei közé tartozik, a júliusi csapadékmennyiséget milliméterben mérve egy X folytonos valószínűségi változó írja le, amelynek eloszlásfüggvénye

$$F_X(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{x^2}{5000}, & \text{ha } 0 < x \leq 50, \\ \frac{1}{2} + \frac{x-50}{50} - \frac{(x-50)^2}{5000}, & \text{ha } 50 < x < 100, \\ 1, & \text{ha } x \geq 100. \end{cases}$$

Feltéve, hogy $x \in (0, 100)$ milliméter eső esik júliusban, a kukoricatő $1 - \frac{x}{100}$ valószínűséggel szárad el a hónap végére. Mi a valószínűsége, hogy a kukoricatő nem szárad el július végére?

(Vizsgakurzus vizsgája, 2024.06.10., egy viszonylag könnyű 6. feladat.)

A folytonos valószínűségi változóra vett feltételes valószínűség kétféle definíciója

(12.1.1. Definíció). Legyen X valószínűségi változó és A esemény. Ekkor A -nak az X -re vett **feltételes valószínűsége** az $x \mapsto \mathbb{E}(\mathbb{1}_A | X = x)$ regressziós függvény. Jelölése: $\mathbb{P}(A | X = x)$.

Ez a definíció szerepel Mészáros Szabolcs jegyzetében, azonban előadáson korábban már elhangzott egy másik definíció is:
Ha $f_X(x) > 0$, akkor

$$\mathbb{P}(A | X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A | X \in (x - \Delta x, x + \Delta x)).$$

Zavaró kérdés: A kettő tényleg ugyanaz?

Ha igen, akkor a második a jelen ismereteink alapján hasznosabb. Ugyanis az $\mathbb{E}(Y | X)$ regressziót ugyan bármilyen X, Y val. változók esetén definiáltuk, de kiszámolni csak két esetben tanultuk:

- ▶ ha (X, Y) folytonos val. vektorváltozó (és így Y is folytonos),
- ▶ vagy ha X diszkrét.

Itt viszont X folytonos, míg $Y = \mathbb{1}_A$ diszkrét (sőt 0-1-értékű).

A folytonos valószínűségi változóra vett feltételes valószínűség kétféle definíciója

(12.1.1. Definíció). Legyen X valószínűségi változó és A esemény. Ekkor A -nak az X -re vett **feltételes valószínűsége** az $x \mapsto \mathbb{E}(1_A | X = x)$ regressziós függvény. Jelölése: $\mathbb{P}(A | X = x)$.

Ez a definíció szerepel Mészáros Szabolcs jegyzetében, azonban előadáson korábban már elhangzott egy másik definíció is:
Ha $f_X(x) > 0$, akkor

$$\mathbb{P}(A | X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A | X \in (x - \Delta x, x + \Delta x)).$$

Zavaró kérdés: A kettő tényleg ugyanaz?

Ezen diászor végén **informális** bizonyítást adok arra, hogy a két definíció tényleg ekvivalens (ez nem vizsgaanyag, előadáson nem mondom el).

Most: feltesszük, hogy a kettő ugyanaz $\rightarrow$ megnézzük, hogyan lehet ennek segítségével számolni egy speciális esetben.

Teljes valószínűség tétele folytonos esetben: egy fontos példatípus

Speciális eset: ha (X, Y) folytonos val. vektorváltozó és A valamilyen Y -tól függő esemény, pl. $A = \{Y < y\}$, $y \in \mathbb{R}$.

Volt: $\mathbb{P}(Y < y|X = x) \stackrel{\text{def}}{=} F_{Y|X}(y|x) \stackrel{\text{def}}{=} \int_{-\infty}^y f_{Y|X}(z|x)dz$. Ez felel meg az előbbi definíciónak:

$$\mathbb{P}(A|X = x) \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x)).$$

Ezzel a definícióval $A = \{Y < y\}$ -ra igaz a folytonos TVT, vagyis

$$\mathbb{P}(Y < y) = \int_{-\infty}^{\infty} \mathbb{P}(Y < y|X = x)f_X(x)dx, \text{ ugyanis}$$

$$\begin{aligned}\mathbb{P}(Y < y) &= \mathbb{P}(-\infty < X < \infty, -\infty < Y < y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^y f_{X,Y}(x, z)dzdx = \int_{-\infty}^{\infty} \int_{-\infty}^y \frac{f_{X,Y}(x, z)}{f_X(x)}dz f_X(x)dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^y f_{Y|X}(z|x)dz f_X(x)dx = \int_{-\infty}^{\infty} \mathbb{P}(Y < y|X = x)f_X(x)dx.\end{aligned}$$

(Ebből még nem derül ki, hogy az alternatív 12.1.1. Definícióval is igaz. Lásd hátul.) 97 / 123

Megjegyzés: folytonos TVT differenciális alakja

Legyen (X, Y) folytonos valószínűségi vektorváltozó.

Az előző dián láttuk: ha $y \in \mathbb{R}$, akkor

$$\mathbb{P}(Y < y) = \int_{-\infty}^{\infty} \mathbb{P}(Y < y | X = x) f_X(x) dx.$$

Eloszlásfüggvényekkel kifejezve:

$$F_Y(y) = \int_{-\infty}^{\infty} F_{Y|X}(y|x) f_X(x) dx.$$

Deriváljuk ezt y szerint:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx.$$

Ezt az alakot is szokták néha folytonos TVT-nek (pontosabban a folytonos TVT differenciális avagy infinitezimális alakjának) hívni, vizsgán 6-os feladatban hasznos lehet.

Vegyük észre azonban, hogy a feltételes sűrűségfüggvény definíciója alapján nem más, mint az alábbi, ismert azonosság:

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

Teljes valószínűség tétele folytonos esetben: példák

A feladatsoron szereplő folytonos TVT-s feladatokból szinte az összes az alábbi két típus egyikébe tartozik:

1. ahol $\mathbb{P}(A|X = x)$ -et gyakorlatilag csak ki kell olvasni a feladat szövegéből (11.19., 20. és 21. feladat),
 2. vagy ahol (X, Y) folytonos valószínűségi vektorváltozó, f_X és $f_{Y|X}$ van megadva, és A valami Y -től függő esemény (11.15., 16. és 17. feladat – a látszat ellenére ezek közül a 3. a legkönnyebb!).
- ▶ Az utóbbinak lehet olyan, trükkösebb változata is, ahol nem ilyen egyszerű rájönni, hogy a folytonos TVT-ről van szó.
 - ▶ Lehet olyan is, ahol X ismeretében Y diszkrét (pl. 11.18. feladat, ami egyébként az első típusba tartozik).
 - ▶ Az integrálások könnyen elbonyolódhatnak (lásd 11.16., 11.19. feladat), ezt a vizsgafeladatok kitűzésekor igyekszünk elkerülni.

Feladatok sorszámai frissítve.

Emlékeztető: normális eloszlás

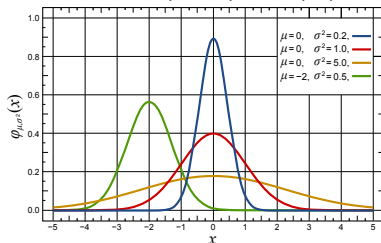
(8.1.1. Definíció.) Egy Y folytonos valószínűségi változó **normális eloszlású** $\mu \in \mathbb{R}$ és $\sigma^2 > 0$ paraméterekkel ($Y \sim N(\mu; \sigma^2)$), ha sűrűségfüggvénye

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Speciális eset: standard normális:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \int_{-\infty}^x \varphi(x) dx.$$

$X \sim N(0; 1) \Rightarrow Y = \sigma X + \mu \sim N(\mu; \sigma^2)$ és $\mathbb{E}(Y) = \mu$, $\mathbb{D}^2(Y) = \sigma^2$.



Független normálisok összege normális, és a paraméterek összeadódnak

Korábban egyszer már ígértük az alábbi állítást:

(Állítás.) Legyen $\mu_1, \mu_2 \in \mathbb{R}$ és $\sigma_1, \sigma_2 > 0$. Ha $X \sim N(\mu_1; \sigma_1^2)$ és $Y \sim N(\mu_2; \sigma_2^2)$ függetlenek, akkor $X + Y \sim N(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$.

Ebben csak az a meglepő, hogy az összeg is normális eloszlású.

Ugyanis a várható érték és szórásnégyzet tulajdonságai alapján:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_1 + \mu_2$$

és mivel függetlenek,

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) = \sigma_1^2 + \sigma_2^2.$$

Normalitás bizonyítása? Pl. konvolúciós képlettel: ahogy már említettük, ha X és Y függetlenek és folytonosak, akkor $X + Y$ sűrűségfüggvénye $f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(z)f_Y(x-z)dz$, $x \in \mathbb{R}$.

Ez az új TAD szerint már nem vizsgaanyag. Összetett számolás, teljes négyzetté kiegészítés a kitevőben... → akinek kedve van, próbálja meg belátni az állítást $\mu_1 = \mu_2 = 0$ -ra és $\sigma_1^2 = \sigma_2^2 = 1$ -re.

Független normálisok összege normális, és a paraméterek összeadódnak – feladatok

(Állítás.) Legyen $\mu_1, \mu_2 \in \mathbb{R}$ és $\sigma_1, \sigma_2 > 0$. Ha $X \sim N(\mu_1; \sigma_1^2)$ és $Y \sim N(\mu_2; \sigma_2^2)$ függetlenek, akkor $X + Y \sim N(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$.

Így természetesen nemcsak kettő, hanem n független normális összege is normális és a paramétereik összeadódnak.

Ez alapján az összeg a szokásos módon standardizálható, a normális eloszlás táblázat segítségével az összegre vonatkozó valószínűségek is kiszámíthatóak.

Ezt és a kovariancia/korreláció azonosságainak alkalmazását kombinálja a 12. feladatsoron az 1., 2. b), c) [régi vizsgafeladat, nem 6-os!], 3., 4. a), b), c) feladat.

A 3. talán kissé trükkös, ezt idén megnézzük össznépiént:

3. Legyenek $X \sim N(5; 2)$ és $Y \sim N(4; 3)$ függetlenek. Adjuk meg a $\mathbb{P}(X < Y)$ valószínűséget.

Függetlenség és korrelátlanság ugyanazon független normálisok különböző lineáris kombinációi esetén

Volt: Ha az X és Y valószínűségi változók (véges szórással) függetlenek, akkor a kovarianciájuk 0.

Szintén volt: Ha az X, Y (véges szórású) valószínűségi változók kovarianciája 0, abból nem feltétlenül következik, hogy X és Y függetlenek.

Független normális eloszlású val. változókból képzett lineáris kombinációk esetén azonban **a korrelátlanságból következik a függetlenség:**

(Állítás, idén nem bizonyítjuk.) Legyenek $Z_1, \dots, Z_n$ független, normális eloszlású valószínűségi változók, $X = c + \sum_{i=1}^n a_i Z_i$ és $Y = d + \sum_{i=1}^n b_i Z_i$ valamilyen $a_1, \dots, a_n, b_1, \dots, b_n, c, d$ valós számokra.

Ekkor ha $\text{cov}(X, Y) = 0$, akkor X és Y függetlenek.

(Példa: 12.4. feladat $\rightarrow$ lásd gyakorlat.)

Regresszió és lineáris regresszió

A regresszió és a lineáris regresszió általában nem ugyanaz. De:

(11.3.3. Állítás, nem biz.) Legyenek $Z_1, \dots, Z_n$ független, normális eloszlású valószínűségi változók, $X = c + \sum_{i=1}^n a_i Z_i$ és

$Y = d + \sum_{i=1}^n b_i Z_i$ valamilyen $a_1, \dots, a_n, b_1, \dots, b_n, c, d$ valós számokra.

Ekkor $\mathbb{E}(Y|X)$ megegyezik az Y változó X -re vett lineáris regressziójával.

A konkrétság kedvéért: ha $\mathbb{E}(X) = \mu_1$, $\mathbb{E}(Y) = \mu_2$,

$$\text{cov} \left(\begin{pmatrix} X \\ Y \end{pmatrix} \right) = \begin{pmatrix} \mathbb{D}^2(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \mathbb{D}^2(Y) \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \text{ akkor}$$

$$\mathbb{E}(Y|X) = \beta X + \alpha,$$

ahol $\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = \frac{b}{a}$ és $\alpha = \mathbb{E}(Y) - \beta \mathbb{E}(X) = \mu_2 - \frac{b}{a} \mu_1$, tehát

$$\mathbb{E}(Y|X) = \frac{b}{a} X + \mu_2 - \frac{b}{a} \mu_1.$$

(Példa: 12.5. feladat, lásd [részben] össznépi gyakorlat. Másik

példa: 2024.01.18. vizsga 6. feladat $\rightarrow$ tárgyhonlap.)

Regresszió és lineáris regresszió

A regresszió és a lineáris regresszió általában nem ugyanaz. De:

(11.3.3. Állítás, nem biz.) Legyenek $Z_1, \dots, Z_n$ független, normális eloszlású valószínűségi változók, $X = c + \sum_{i=1}^n a_i Z_i$ és

$Y = d + \sum_{i=1}^n b_i Z_i$ valamilyen $a_1, \dots, a_n, b_1, \dots, b_n, c, d$ valós számokra.

Ekkor $\mathbb{E}(Y|X)$ megegyezik az Y változó X -re vett lineáris regressziójával.

A konkrétság kedvéért: ha $\mathbb{E}(X) = \mu_1$, $\mathbb{E}(Y) = \mu_2$,

$$\text{cov} \left(\begin{pmatrix} X \\ Y \end{pmatrix} \right) = \begin{pmatrix} \mathbb{D}^2(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \mathbb{D}^2(Y) \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \text{ akkor}$$

$$\mathbb{E}(Y|X) = \beta X + \alpha,$$

ahol $\beta = \frac{\text{cov}(X, Y)}{\mathbb{D}^2(X)} = \frac{b}{a}$ és $\alpha = \mathbb{E}(Y) - \beta \mathbb{E}(X) = \mu_2 - \frac{b}{a} \mu_1$, tehát

$$\mathbb{E}(Y|X) = \frac{b}{a} X + \mu_2 - \frac{b}{a} \mu_1.$$

(Példa: 12.5. feladat, lásd [részben] össznépi gyakorlat. Másik példa: 2024.01.18. vizsga 6. feladat $\rightarrow$ tárgyhonlap.)

Standard n -dimenziós normális eloszlás

Kérdés: Hogyan általánosítható a normális eloszlás több dimenzióra? A standard eset egyszerű, a nem standardot idén csak érintőlegesen tárgyaljuk.

(Definíció.) Egy $\underline{X} = (X_1, \dots, X_n)$ val. vektorváltozó **standard n -dimenziós normális eloszlású**, ha koordinátái független, standard normális eloszlású val. változók.

A függetlenség miatt $\underline{X}$ sűrűségfüggvénye n darab standard normális sűrűv. szorzata (lásd 12.3.2. Definíció):

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \right) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}.$$

Nem-standard n -dimenziós normális eloszlás

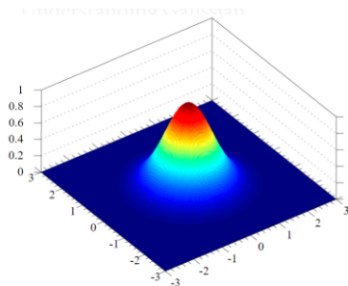
(12.3.3. Definíció.) Az $\underline{Y} = (Y_1, \dots, Y_n)$ valószínűségi vektorváltozó **n -dimenziós normális eloszlású**, ha létezik $A \in \mathbb{R}^{n \times n}$ mátrix, $\underline{\mu} \in \mathbb{R}^n$ vektor és $\underline{X}$ n -dimenziós standard normális eloszlású valószínűségi vektorváltozó, amire

$$\underline{Y} = A \cdot \underline{X} + \underline{\mu},$$

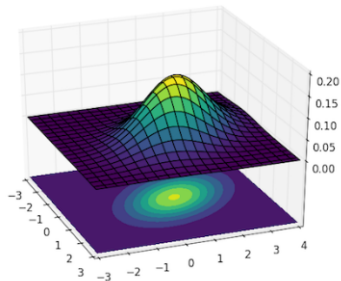
$\underline{X}$ -et oszlopvektorként kezelve.

$\underline{Y}$ eloszlását **nemelfajuló**nak hívjuk, ha $\det A \neq 0$.

Standard normális



Nem-standard normális



Nem-standard normális koordinátái normálisak

(Állítás.) Ha $\underline{Y}$ **remelfajuló** n -dimenziós normális eloszlású, akkor minden koordinátája normális eloszlású.

Biz.: legyen $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$, $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix}$ és $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix}$

olyan, hogy $\underline{Y} = A \cdot \underline{X} + \underline{\mu}$.

Belátjuk, hogy Y_i (ahol $1 \leq i \leq n$) normális eloszlású.

A mátrixszorzás szabályai alapján

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{in}X_n + \mu_i.$$

Tudjuk, hogy minden j -re X_j standard normális eloszlású. Tehát ha $a_{ij} \neq 0$, akkor $a_{ij}X_j$ is normális eloszlású.

Mivel feltettük, hogy **$\det A \neq 0$** , A -nak nem lehet csupa 0 sora, így létezik olyan j , amelyre $a_{ij} \neq 0$.

Emiatt Y_i független normálisok összege plusz $\mu_i \Rightarrow$ a korábbi állítás szerint normális eloszlású.

Nem-standard normális koordinátái normálisak

(Állítás.) Ha $\underline{Y}$ **nemelfajuló** n -dimenziós normális eloszlású, akkor minden koordinátája normális eloszlású.

Biz.: legyen $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$, $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix}$ és $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix}$

olyan, hogy $\underline{Y} = A \cdot \underline{X} + \underline{\mu}$.

Belátjuk, hogy Y_i (ahol $1 \leq i \leq n$) normális eloszlású.

A mátrixszorzás szabályai alapján

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{in}X_n + \mu_i.$$

Tudjuk, hogy minden j -re X_j standard normális eloszlású. Tehát ha $a_{ij} \neq 0$, akkor $a_{ij}X_j$ is normális eloszlású.

Mivel feltettük, hogy **$\det A \neq 0$** , A -nak nem lehet csupa 0 sora, így létezik olyan j , amelyre $a_{ij} \neq 0$.

Emiatt Y_i független normálisok összege plusz $\mu_i \Rightarrow$ a korábbi állítás szerint normális eloszlású.

Nem-standard normális eloszlás várhatóérték-vektora és kovarianciamátrixa

Az előbbi $\underline{Y} = A\underline{X} + \underline{\mu}$, azaz

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}}_{=:\underline{Y}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}}_{=:A} \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix}}_{=:\underline{X}} + \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix}}_{=:\underline{b}}$$

esetben ha feltesszük, hogy az $X_1, \dots, X_n$ független val. változók **standard** normális eloszlásúak (azaz $\mathbb{E}(X_i) = 0, \mathbb{D}^2(X_i) = 1$), akkor egyszerű lineáris algebrai műveletekkel belátható:

- ▶ $\underline{Y} = A\underline{X} + \underline{\mu}$ várhatóérték-vektora $\mathbb{E}(\underline{Y}) = \underline{\mu}$
- ▶ és $\underline{Y} = A\underline{X} + \underline{\mu}$ kovarianciamátrixa $\text{cov}(\underline{Y}) = AA^T$.

A bizonyítást nem részletezzük.

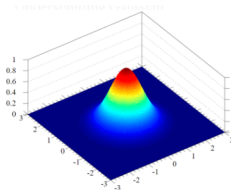
Kitekintés: nemelfajuló többdimenziós normális eloszlás

Az alábbiakat idéntől nem részletezzük. Az érdeklődnek ajánlom a leendő Mesterséges intelligencia MSc haladó valszámos tárgyát.

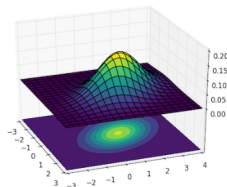
- ▶ Láttuk: $\underline{Y} = A\underline{X} + \underline{\mu}$ várhatóérték-vektora $\underline{\mu}$, kovarianciamátrixa $\underline{\Sigma} := AA^T$.
- ▶ Belátható: $\underline{Y}$ sűrűségfüggvénye $\underline{\mu}$ és $\underline{\Sigma}$ segítségével kifejezhető. Ez a két paraméter meghatározza az eloszlást. (Ez nem triviális, mert $\underline{\Sigma} = AA^T$ többféle A -ból is előállhat!)
- ▶ Többdimenziós normális eloszlás további lineáris transzformáltja is többdimenziós normális. Ha $B \in \mathbb{R}^{n \times n}$ és $\det B \neq 0$, akkor $B\underline{Y}$ várhatóérték-vektora $B\underline{\mu}$, kovarianciamátrixa $B^T \underline{\Sigma} B$.
- ▶ A korábbi állítások átfogalmazása: Többdimenziós normális eloszlás esetén a függetlenségből következik a korrelálatlanság, és a regresszió lineáris.
- ▶ Láttuk: $\underline{Y} = A\underline{X} + \underline{\mu}$ peremeloszlásai normálisak. A peremeloszlások normalitásából azonban nem következik a vektor n -dimenziós normalitása.

Vizualizálás még egyszer

Standard normális



Nem-standard normális



Standard normális: a szintvonalak körök, forgásszimmetrikus.

Nem standard, $\underline{\mu} = 0$: a szintvonalak ellipszisek.

Létezik olyan $\underline{U} \in \mathbb{R}^{2 \times 2}$ ortogonális transzformáció (ortogonális = $\underline{U}^{-1} = \underline{U}^T$), amely az ellipszis főtengelyeit a koordinátatengelyekbe viszi.

Ekkor $\underline{U} \cdot \underline{Y} \sim N(0; D)$, ahol D diagonális mátrix. $\underline{U} \cdot \underline{Y}$ koordinátái függetlenek, $\det(D) = \det(\Sigma)$.

Standard normális eloszlás szimulációja (kieg. anyag)

Volt: ha X folytonos v.v. F_X eloszlásfüggvénnyel és $U \sim U(0; 1)$, akkor $F_X^{-1}(U)$ ugyanolyan eloszlású, mint X .

Tehát ha egyenletes eloszlású (pszeudo)véletlen számokat tudunk szimulálni, akkor X -szel azonos eloszlásúakat is.

Példaként láttuk: exponenciális eloszlású X .

Kérdés: Hogyan működik ez **standard normális eloszlású** X esetén?

Probléma: a Φ eloszlásfüggvényre nincs zárt képlet (csak táblázat), $\Phi^{-1}(U)$ -t körülményes kiszámolni.

Állítás: ha U és V két független $U(0; 1)$ eloszlású v.v., akkor $X = \sqrt{-2 \ln U} \cos(2\pi V)$ és $Y = \sqrt{-2 \ln U} \sin(2\pi V)$ két független, standard normális eloszlású v.v.

Más szavakkal: ha (U, V) egyenletes eloszlású $(0, 1) \times (0, 1)$ -en, akkor (X, Y) standard kétdimenziós normális.

Biz.: polárkoordinátákkal, nem túl durva, 3 éve volt IMSc-s házi. Elsőre talán meglepő, hogy két függetlent könnyebb szimulálni, mint egy darabot.

A standard kétdimenziós normális néha szebben viselkedik, mint az egydimenziós $\rightarrow$ lásd idei IMSc-s házi!

Standard normális eloszlás szimulációja (kieg. anyag)

Volt: ha X folytonos v.v. F_X eloszlásfüggvénnyel és $U \sim U(0; 1)$, akkor $F_X^{-1}(U)$ ugyanolyan eloszlású, mint X .

Tehát ha egyenletes eloszlású (pszeudo)véletlen számokat tudunk szimulálni, akkor X -szel azonos eloszlásúakat is.

Példaként láttuk: exponenciális eloszlású X .

Kérdés: Hogyan működik ez **standard normális eloszlású** X esetén?

Probléma: a Φ eloszlásfüggvényre nincs zárt képlet (csak táblázat), $\Phi^{-1}(U)$ -t körülményes kiszámolni.

Állítás: ha U és V két független $U(0; 1)$ eloszlású v.v., akkor $X = \sqrt{-2 \ln U} \cos(2\pi V)$ és $Y = \sqrt{-2 \ln U} \sin(2\pi V)$ két független, standard normális eloszlású v.v.

Más szavakkal: ha (U, V) egyenletes eloszlású $(0, 1) \times (0, 1)$ -en, akkor (X, Y) standard kétdimenziós normális.

Biz.: polárkoordinátákkal, nem túl durva, 3 éve volt IMSc-s házi. Elsőre talán meglepő, hogy két függetlent könnyebb szimulálni, mint egy darabot.

A standard kétdimenziós normális néha szebben viselkedik, mint az egydimenziós $\rightarrow$ lásd idei IMSc-s házi!

Példa: többdimenziós binomiális eloszlás

(12.2.1. Példa.) Szabályos dobókocka, címkéi: 1 db 1-es, 2 db 2-es, 3 db 3-as.

13-szor dobunk. Mi a valószínűsége, hogy 3 db 1-est, 4 db 2-est és 6 db 3-ast látunk?

$X_i := i$ értékű dobások száma.

Klasszikus valószínűség:

$$\mathbb{P}(X_1 = 3, X_2 = 4, X_3 = 6) = \frac{13!}{3!4!6!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{3}\right)^4 \left(\frac{1}{2}\right)^6 \approx 0,05364 \approx 5\%.$$

Példa: többdimenziós binomiális eloszlás

(12.2.1. Példa.) Szabályos dobókocka, címkéi: 1 db 1-es, 2 db 2-es, 3 db 3-as.

13-szor dobunk. Mi a valószínűsége, hogy 3 db 1-est, 4 db 2-est és 6 db 3-ast látunk?

$X_i := i$ értékű dobások száma.

Klasszikus valószínűség:

$$\mathbb{P}(X_1 = 3, X_2 = 4, X_3 = 6) = \frac{13!}{3!4!6!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{3}\right)^4 \left(\frac{1}{2}\right)^6 \approx 0,05364 \approx 5\%.$$

Polinomiális eloszlás

(12.2.2. Definíció.) Az $\underline{X} = (X_1, \dots, X_m)$ val. vektorváltozó **polinomiális** (avagy **multinomiális**) **eloszlású**, n és $(p_1, \dots, p_m) \in [0, 1]^m$ paraméterekkel, ha $p_1 + \dots + p_m = 1$ és

$$\mathbb{P}(X_1 = k_1, \dots, X_m = k_m) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}$$

minden $0 \leq k_i \leq n$ ($i = 1, \dots, m$) esetén, ha $k_1 + \dots + k_m = n$.

Megjegyzés:

- ▶ A peremeloszlások binomiális eloszlásúak.
- ▶ A koordináták nem függetlenek.
- ▶ A peremeloszlások nem határozzák meg az együttes eloszlást.

Dobókockás példa: $p_1 = 1/6, p_2 = 1/3, p_3 = 1/2$ (tehát $m = 3$), $n = 13$.

Példa: polinomiális eloszlás

A polinomiális/multinomiális eloszlás fontos szerepet játszik pl. a matematikai populációbiológiában.

Példa: Egy baktériumpopuláció ivartalanul szaporodik és minden reggel n egyedből áll, akik az előző reggeli populáció utódai.

Tudjuk, hogy minden egyed azonos valószínűséggel lehet bármelyik előző napi egyed utódja, a többi egyedtől függetlenül (az előző napi egyedek eddigre elpusztulnak).

Legyen X_i az előző napi i -edik egyed utódainak száma.

$(X_1, \dots, X_n)$ együttes eloszlása?

Példa: polinomiális eloszlás

A polinomiális/multinomiális eloszlás fontos szerepet játszik pl. a matematikai populációbiológiában.

Példa: Egy baktériumpopuláció ivartalanul szaporodik és minden reggel n egyedből áll, akik az előző reggeli populáció utódai.

Tudjuk, hogy minden egyed azonos valószínűséggel lehet bármelyik előző napi egyed utódja, a többi egyedtől függetlenül (az előző napi egyedek eddigre elpusztulnak).

Legyen X_i az előző napi i -edik egyed utódainak száma.
($X_1, \dots, X_n$) együttes eloszlása?

Ha $k_i \geq 0$ ($1 \leq i \leq n$) és $k_1 + \dots + k_n = n$, akkor

$$\mathbb{P}(X_1 = k_1, \dots, X_n = k_n) = \frac{n!}{k_1! \dots k_n!} \left(\frac{1}{n}\right)^{k_1} \dots \left(\frac{1}{n}\right)^{k_n}.$$

Tehát $(X_1, \dots, X_n)$ polinomiális eloszlású n és $(1/n, \dots, 1/n)$ paraméterekkel (azaz $m = n$ és $p_1 = p_2 = \dots = p_n = 1/n$).

Ezt az esetet (amikor $m = n$ és a p_i -k egyenlők) **szimmetrikus** polinomiális/multinomiális eloszlásnak is nevezik.

Bővebben lásd: **Wright–Fisher-modell**.

Függelék

Példa a valószínűségi módszerre: egy színezési probléma (kiegészítő anyag)

Adott az n csúcsú teljes gráf, K_n , ahol $n \geq 1$.

Kérdés: Adott $k \geq 1$ esetén ki lehet-e színezni K_n éleit két színnel, **pirossal** és **kékkel** úgy, hogy ne legyen benne se piros, se kék k csúcsú teljes részgráf (azaz klikk)?

Most: Ha $k \geq 3$ és $n < 2^{k/2}$, akkor igen.

Bizonyítás ötlete (Erdős Pál): színezzük az éleket véletlenszerűen és egymástól függetlenül, $1/2$ - $1/2$ valószínűséggel pirosra vagy kékre. Mutassuk meg, hogy **pozitív valószínűséggel** olyan színezést kapunk, amelyben nincs sem piros, sem kék k -as klikk.

Ez garantálja azt is, hogy kell lennie egy olyan (determinisztikus) színezésnek, amelyre ugyanez igaz.

Egy színezési probléma

Cél: Ha $3 \leq n < 2^{k/2}$, akkor K_n élei kiszínezhetők 2 színnel úgy, hogy ne legyen a gráfban se piros, se kék k -as klikk.

Jelölje g_n a K_n összes lehetséges színezéseinek számát. Ekkor, mivel K_n -nek $\binom{n}{2}$ éle van, $g_n = 2^{\binom{n}{2}}$.

Jelölje $g_{n,k}$ a K_n összes olyan színezéseinek számát, amelyben van teljes **piros** k -as klikk. Ekkor

$$g_{n,k} \leq \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}}.$$

Ez egy durva felső becslés. Az olyan színezéseket, amelyekben van piros $k+1$ -es klikk, $k+1$ -szer számoltuk, a teljesen pirosra színezett K_n -t pedig $\binom{n}{k}$ -szor.

($j > k$ -ra az olyan színezést, amiben van j -es piros klikk, $\binom{j}{k}$ -szor.)
Tehát a rossz (kedvezőtlen) színezések (ahol van piros vagy kék k -as klikk) számára, amit r_n -nel jelölünk, teljesül, hogy

$$r_n \leq 2 \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}} = \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2} + 1}.$$

Egy színezési probléma

Cél: Ha $3 \leq n < 2^{k/2}$, akkor K_n élei kiszínezhetők 2 színnel úgy, hogy ne legyen a gráfban se piros, se kék k -as klikk.

Jelölje g_n a K_n összes lehetséges színezéseinek számát. Ekkor, mivel K_n -nek $\binom{n}{2}$ éle van, $g_n = 2^{\binom{n}{2}}$.

Jelölje $g_{n,k}$ a K_n összes olyan színezéseinek számát, amelyben van teljes **piros** k -as klikk. Ekkor

$$g_{n,k} \leq \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}}.$$

Ez egy durva felső becslés. Az olyan színezéseket, amelyekben van piros $k+1$ -es klikk, $k+1$ -szer számoltuk, a teljesen pirosra színezett K_n -t pedig $\binom{n}{k}$ -szor.

($j > k$ -ra az olyan színezést, amiben van j -es piros klikk, $\binom{j}{k}$ -szor.)
Tehát a rossz (kedvezőtlen) színezések (ahol van piros vagy kék k -as klikk) számára, amit r_n -nel jelölünk, teljesül, hogy

$$r_n \leq 2 \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}} = \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2} + 1}.$$

Egy színezési probléma

Rossz színezések száma:

$$r_n \leq 2 \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}} = \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2} + 1}.$$

Cél: $\mathbb{P}(\text{rossz színezés}) < 1$.

A Poincaré-formula szerint

$$\begin{aligned} \mathbb{P}(\text{rossz színezés}) &= \frac{r_n}{g_n} = \mathbb{P}(\text{van piros } k\text{-as klikk}) + \mathbb{P}(\text{van kék } k\text{-as klikk}) \\ &\quad - \mathbb{P}(\text{van piros és kék } k\text{-as klikk is}) \leq 2 \frac{g_{n,k}}{g_n}. \end{aligned}$$

Így elég belátni: $\frac{g_{n,k}}{g_n} < 1/2$.

Egy színezési probléma

Rossz színezések száma:

$$r_n \leq 2 \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}} = \binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2} + 1}.$$

Cél: $\mathbb{P}(\text{rossz színezés}) < 1$.

A Poincaré-formula szerint

$$\begin{aligned} \mathbb{P}(\text{rossz színezés}) &= \frac{r_n}{g_n} = \mathbb{P}(\text{van piros } k\text{-as klikk}) + \mathbb{P}(\text{van kék } k\text{-as klikk}) \\ &\quad - \mathbb{P}(\text{van piros és kék } k\text{-as klikk is}) \leq 2 \frac{g_{n,k}}{g_n}. \end{aligned}$$

Így elég belátni: $\frac{g_{n,k}}{g_n} < 1/2$. Mivel $n < 2^{k/2}$,

$$\frac{g_{n,k}}{g_n} \leq \frac{\binom{n}{k} 2^{\binom{n}{2} - \binom{k}{2}}}{2^{\binom{n}{2}}} = \binom{n}{k} 2^{-\binom{k}{2}} < \frac{n^k}{k!} 2^{-\binom{k}{2}} < \frac{2^{k^2/2}}{k!} 2^{-k(k-1)/2} = \frac{2^{k/2}}{k!}.$$

A jobb oldal $1/2$ -nél kisebb, ha $k \geq 3$, és 0 -hoz tart, amint $k \rightarrow \infty$.

A valószínűségi módszer filozófiája általában

Valószínűségi módszer: adott valamilyen kombinatorikai/gráfelméleti probléma, kérdés, hogy létezik-e bizonyos fajta objektumok közül egy bizonyos tulajdonsággal rendelkező.

Pl. az esetünkben az objektumok a K_n kék-piros színezései, a keresett tulajdonság az, hogy nincs sem k méretű piros, sem k méretű kék klikk a gráfban.

Egzisztenciabizonyítás: azzal igazoljuk, hogy az adott tulajdonságú objektum létezik (esetünkben $n < 2^{k/2}$ esetén), hogy belátjuk, hogy *véletlenszerűen* választva az objektumok közül, *pozitív valószínűséggel* ilyen tulajdonságút kapunk.

Ez egy **nemkonstruktív** módszer: általában semmit nem tudunk mondani arról, hogy az adott tulajdonságú objektum hogy néz ki. Pl. a módszer nem ad determinisztikus algoritmust adni arra, hogy hogyan konstruáljunk meg (véletlen használata nélkül) egy sem piros, sem kék k -as klikket nem tartalmazó színezést, ha n nagy.

Feltételes valószínűség folytonos esetben

Most megválaszoljuk a **zavaró kérdést**: a feltételes valószínűség folytonos esetre vonatkozó két definíciója ekvivalens egymással.

(Lemma.) Ha $A \in \mathcal{F}$ esemény és X folytonos valószínűségi vektorváltozó, akkor $x \in \mathbb{R}$ -re legyen

$$g(x) = \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x)),$$

ha $f_X(x) > 0$ és 0 különben.

Ekkor $x \mapsto g(x)$ az $\mathbb{1}_A$ indikátorváltozó X -re vonatkozó regressziós függvénye.

Tehát a lemma alapján tényleg igaz, hogy $g(x) = \mathbb{E}(\mathbb{1}_A|X = x) = \mathbb{P}(A|X = x)$ minden x -re, ahol $f_X(x) > 0$.

Most vázlatos bizonyítást adunk a Lemmára.

Feltételes valószínűség folytonos esetben

Most megválaszoljuk a **zavaró kérdést**: a feltételes valószínűség folytonos esetre vonatkozó két definíciója ekvivalens egymással.

(Lemma.) Ha $A \in \mathcal{F}$ esemény és X folytonos valószínűségi vektorváltozó, akkor $x \in \mathbb{R}$ -re legyen

$$g(x) = \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x)),$$

ha $f_X(x) > 0$ és 0 különben.

Ekkor $x \mapsto g(x)$ az $\mathbb{1}_A$ indikátorváltozó X -re vonatkozó regressziós függvénye.

Tehát a lemma alapján tényleg igaz, hogy $g(x) = \mathbb{E}(\mathbb{1}_A|X = x) = \mathbb{P}(A|X = x)$ minden x -re, ahol $f_X(x) > 0$.

Most vázlatos bizonyítást adunk a Lemmára.

Felt. valószínűség folytonos esetben, biz. (kieg. anyag)

(Lemma biz.) A cél tehát azt belátni, hogy a g függvény, ahol

$$g(x) = \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x)),$$

ha $f_X(x) > 0$ és 0 különben, regressziós függvénye 1_A -nak X -re.
A regresszió definíciója szerint ehhez az alábbiak kell teljesülnie:

$$\mathbb{E}(1_A|X < x) = \mathbb{E}(g(X)|X < x),$$

minden olyan x -re, amelyre $\mathbb{P}(X < x) > 0$. Átírva:

$$\begin{aligned} g(x) &= \lim_{\Delta x \downarrow 0} \mathbb{P}(A|X \in (x - \Delta x, x + \Delta x)) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(A \cap \{X \in (x - \Delta x, x + \Delta x)\})}{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))} \\ &= \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(A \cap \{X \in (x - \Delta x, x + \Delta x)\})}{2\Delta x} \cdot \frac{2\Delta x}{\mathbb{P}(X \in (x - \Delta x, x + \Delta x))} \\ &= \left(\frac{d}{dx} \mathbb{P}(A \cap \{X < x\}) \right) \cdot \frac{1}{f_X(x)}, \end{aligned}$$

ha $x \mapsto \mathbb{P}(A \cap \{X < x\})$ differenciálható. Innentől feltesszük, hogy az, és a derivált Riemann-integrálható $-\infty$ -től ∞ -ig.

Felt. valószínűség folytonos esetben, biz. (kieg. anyag)

Azt kaptuk, hogy

$$g(x) = \left(\frac{d}{dx} \mathbb{P}(A \cap \{X < x\}) \right) \cdot \frac{1}{f_X(x)}.$$

Ezért (emlékeztető: $\mathbb{E}(Y|A) = \mathbb{E}(Y\mathbb{1}_A)/\mathbb{P}(A)$, ha A esemény, Y val. változó)

$$\begin{aligned} \mathbb{E}(g(X)|X < x) &= \frac{\mathbb{E}(g(X)\mathbb{1}_{\{X < x\}})}{\mathbb{P}(X < x)} = \frac{\int_{-\infty}^x g(t)f_X(t)dt}{\mathbb{P}(X < x)} \\ &= \frac{\int_{-\infty}^x \frac{d}{dt} \mathbb{P}(A \cap \{X < t\})dt}{\mathbb{P}(X < x)} = \frac{\mathbb{P}(A \cap \{X < x\})}{\mathbb{P}(X < x)}. \end{aligned}$$

Az eredmény pedig valóban megegyezik a következővel:

$$\mathbb{E}(\mathbb{1}_A|X < x) = \frac{\mathbb{P}(A \cap \{X < x\})}{\mathbb{P}(X < x)}.$$

Tehát $x \mapsto g(x)$ az $\mathbb{1}_A$ regressziós függvénye X -re, ahogy állítottuk.