

Matematisk Statistik och Diskret Matematik, MVE051/MSG810, VT22

Timo Vilkas

Chalmers - Göteborgs Universitet

Föreläsning 5
5 april 2022



Tvådimensionella stokastiska variabler

- En **tvådimensionell stokastisk variabel** (ibland kallad slumpvektor) är en funktion (X, Y) definierad på ett utfallsrum S som avbildar till $\mathbb{R}^2$. Båda koordinater X och Y är då (vanliga) slumpvariabler.
- En tvådimensionell s.v. kan vara **diskret** (om både X och Y är diskreta) eller **kontinuerlig** (om både X och Y är kontinuerliga).

Diskreta tvådimensionella s.v.

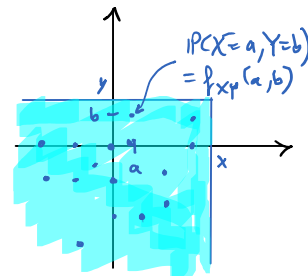
Låt X och Y vara två diskreta s.v.

- Sannolikhetsfunktionen f_{XY} av den tvådimensionella s.v. (X, Y) är definierad som

$$f_{XY}(x, y) = \mathbb{P}(X = x \text{ och } Y = y)$$

- Som förut är $f_{XY}(x, y)$ en sannolikhetsfunktion om och endast om
 - $f_{XY}(x, y) \geq 0$ för alla $(x, y) \in \mathbb{R}^2$
 - $\sum_{x \in X(S)} \sum_{y \in Y(S)} f_{XY}(x, y) = 1$
- Också fördelningsfunktionen har en tvådimensionell motsvarighet: $F(x, y)$ är definierad som

$$F(x, y) = \mathbb{P}(X \leq x \text{ och } Y \leq y) = \sum_{j \leq x} \sum_{k \leq y} f_{XY}(j, k)$$



Exempel

Låt X och Y stå för antalet pojkar resp. flickor i en slumpmässigt vald svensk familj. Sannolikhetsfunktion $f_{XY}(x, y)$ ges i tabellen:

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$\begin{aligned}
 F_{XY}(1, 2) &= \mathbb{P}(X \leq 1, Y \leq 2) \\
 &= f_{XY}(0, 0) + f_{XY}(0, 1) + f_{XY}(0, 2) \\
 &\quad + f_{XY}(1, 0) + f_{XY}(1, 1) + f_{XY}(1, 2) \\
 &= 0.38 + 0.17 + 0.05 + 0.16 + 0.08 + 0.02 = 0.86
 \end{aligned}$$

Marginella sannolikhetsfunktioner

Låt (X, Y) vara en två dimensionell diskret s.v. med täthetsfunktion f_{XY} . De **marginella** sannolikhetsfunktionerna f_X och f_Y för X resp. Y , ges av

$$f_X(x) = \sum_{y \in Y(S)} \underbrace{f_{XY}(x, y)}_{= \mathbb{P}(X=x, Y=y)}, \quad \text{för } x \in X(S)$$

Man delar upp händelsen $\{X=x\}$ med avseende på partitionen $\{Y=y\}, y \in Y(S)$

och

$$f_Y(y) = \sum_{x \in X(S)} f_{XY}(x, y), \quad \text{för } y \in Y(S).$$

$$\leadsto \mathbb{P}(X=x) = \sum_y \mathbb{P}(X=x, Y=y)$$

Exempel

(föregående exempel)

	X	0	1	2	3	4	
Y							f_Y
0		0.38	0.16	0.04	0.01	0.01	0.60
1		0.17	0.08	0.02			0.27
2		0.05	0.02	0.01			0.08
3		0.02	0.01				0.03
4		0.02					0.02
f_X		0.64	0.27	0.07	0.01	0.01	1

Kontinuerliga tvådimensionella s.v.

Låt X och Y vara två kontinuerliga s.v.

- En funktion f_{XY} som uppfyller
 - 1 $f_{XY}(x, y) \geq 0$ för alla $-\infty < x < +\infty$ och $-\infty < y < +\infty$,
 - 2 $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = 1$ samt
 - 3 $\mathbb{P}(a \leq X \leq b \text{ och } c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$
 kallas täthetsfunktion för slumpvektorn (X, Y)
- De marginella täthetsfunktionerna f_X och f_Y för X resp. Y ges av

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy \quad \text{och} \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

Exempel (Ex.8 i boken)

Låt X representera temperaturen (i grader) och Y tiden (i minuter) för hur lång tid en dieselmotor tar på sig för att värma upp. Täthetsfunktionen för (X, Y) ges av

$$f_{XY}(x, y) = c(4x + 2y + 1) \quad 0 \leq x \leq 40, \quad 0 \leq y \leq 2$$

- 1 Beräkna c .
- 2 Beräkna sannolikheten att temperaturen blir större än 20 och det tar minst 1 minut för att motoren värmer upp.
- 3 Beräkna de marginella täthetsfunktionerna f_X och f_Y .
- 4 Beräkna sannolikheten att det tar minst en minut för motoren att värma upp.
- 5 Beräkna sannolikheten att temperaturen är större än 20.

Exempel (Lösning)

1

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy &= \int_0^{40} \int_0^2 c(4x + 2y + 1) dy dx \\ &= c \cdot \int_0^{40} \left[y^2 + (4x+1)y \right]_0^2 dx = c \int_0^{40} 8x + 6 dx = c \left[4x^2 + 6x \right]_0^{40} \\ &= (6400 + 240 - 0) \cdot c = 6640 \cdot c \end{aligned}$$

Alltså är $c = \frac{1}{6640}$.

Exempel (lösning)

$$② \quad \mathbb{P}(X > 20, Y \geq 1) = \int_{20}^{40} \int_1^2 \frac{1}{6640} (4x + 2y + 1) dy dx \approx 0.3735$$

$$③ \quad f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^2 \frac{1}{6640} (4x + 2y + 1) dy = \frac{8x+6}{6640}, x \in [0, 40]$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_0^{40} \frac{1}{6640} (4x + 2y + 1) dx = \frac{3240+80y}{6640}, y \in [0, 2]$$

$$④ \quad \mathbb{P}(Y \geq 1) = \int_1^2 f_Y(y) dy = \left[\frac{3240y + 40y^2}{6640} \right]_1^2 = \frac{42}{83} \approx 0.506$$

$$⑤ \quad \mathbb{P}(X > 20) = \int_{20}^{40} f_X(x) dx = \left[\frac{4x^2 + 6x}{6640} \right]_{20}^{40} = \frac{123}{166} \approx 0.741$$

OBS: $\underbrace{\mathbb{P}(X > 20, Y \geq 1)}_{\approx 0.3735} \neq \mathbb{P}(Y \geq 1) \cdot \mathbb{P}(X > 20) \approx 0.506 \cdot 0.741 = 0.3749$

Oberoende stokastiska variabler

- Två händelser A och B är oberoende om $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ (jfr. föreläsning 1).
- Två stokastiska variabler X och Y är **oberoende** om

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{för alla } (x, y) \in \mathbb{R}^2.$$

OBS: För oberoende X, Y och godtyckliga mängder $C, D \subseteq \mathbb{R}$ gäller

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C) \cdot \mathbb{P}(Y \in D),$$

alltså t.ex. $F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = F_X(x) \cdot F_Y(y)$.

t.ex. $C = [a, b], D = [c, d]$ och (X, Y) kont.:

$$\begin{aligned} \mathbb{P}(X \in C, Y \in D) &= \int_a^b \int_c^d f_{XY}(x, y) dy dx \stackrel{\text{obero.}}{=} \int_a^b \int_c^d \underbrace{f_X(x)}_{\in \mathbb{R}} \cdot f_Y(y) dy dx = \int_a^b f_X(x) dx \cdot \int_c^d f_Y(y) dy \\ &= \int_a^b f_X(x) dx \cdot \int_c^d f_Y(y) dy = \mathbb{P}(X \in C) \cdot \mathbb{P}(Y \in D) \end{aligned}$$

Oberoende eller ej?

Exempel

- I det diskreta exemplet ovan är X och Y inte oberoende då t.ex.

$$0 = f_{XY}(3, 1) \neq f_X(3) \cdot f_Y(1) = 0.01 \cdot 0.27 = 0.0027.$$

- I det kontinuerliga exemplet ovan är X och Y inte heller oberoende eftersom t.ex.

$$0.3735 \approx \mathbb{P}(X > 20, Y \geq 1) \neq \mathbb{P}(X > 20) \cdot \mathbb{P}(Y \geq 1) \approx 0.3749$$

Väntevärde

Låt (X, Y) vara en tvådimensionell s.v. (med täthetsfunktion $f_{XY}(x, y)$) och $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ en reellvärdig funktion.

Väntevärdet $\mathbb{E}[H(X, Y)]$ defineras för diskret (X, Y) genom

$$\mathbb{E}[H(X, Y)] = \sum_{x \in X(S)} \sum_{y \in Y(S)} H(x, y) \cdot f_{XY}(x, y)$$

och för kontinuerlig (X, Y) genom

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

(I båda fall: givet att $\mathbb{E}[|H(X, Y)|] < \infty$).

Marginella väntevärden

Observera att man utan vidare kan räkna ut väntevärdet av en koordinat i slumpvektorn (genom att välja t.ex. $H(x, y) = x$):

Om (X, Y) är diskret gäller

$$\mathbb{E}(X) = \sum_{x \in X(S)} x \cdot \underbrace{\sum_{y \in Y(S)} f_{XY}(x, y)}_{= f_X(x)}.$$

För kontinuerlig (X, Y) är t.ex.

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot \underbrace{\int_{-\infty}^{\infty} f_{XY}(x, y) dx}_{= f_Y(y)} dy.$$

Precis som för endimensionella slumpvariabler (vilket $H(X, Y)$ faktiskt är) får vi en hel del enkla räkneregler:

- Så gäller t.ex. fortfarande att väntevärdet är linjärt, dvs. för två funktioner $G, H : \mathbb{R}^2 \rightarrow \mathbb{R}$ och en konstant $c \in \mathbb{R}$:

$$\mathbb{E}[c H(X, Y)] = c \mathbb{E}[H(X, Y)] \text{ och}$$

$$\mathbb{E}[H(X, Y) + G(X, Y)] = \mathbb{E}[H(X, Y)] + \mathbb{E}[G(X, Y)].$$

- Om X och Y är oberoende, så får vi

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y], \text{ t.ex. när } (X, Y) \text{ är diskret, } H(x, y) = x \cdot y$$

$$\begin{aligned} \mathbb{E}[H(X, Y)] &= \sum_{x \in X(S)} \sum_{y \in Y(S)} H(x, y) \cdot f_{XY}(x, y) = \sum_{x \in X(S)} \sum_{y \in Y(S)} x \cdot f_X(x) \cdot y \cdot f_Y(y) \\ &= \sum_{x \in X(S)} x \cdot f_X(x) \cdot \underbrace{\sum_{y \in Y(S)} y \cdot f_Y(y)}_{= \mathbb{E}Y} = \mathbb{E}X \cdot \mathbb{E}Y \end{aligned}$$

vilket medför

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &\stackrel{\text{linj.}}{=} \mathbb{E}(X^2) + 2\mathbb{E}(X \cdot Y) + \mathbb{E}(Y^2) - ((\mathbb{E}X)^2 + 2\mathbb{E}X \cdot \mathbb{E}Y + (\mathbb{E}Y)^2) \\ &\stackrel{X, Y \text{ oberoende}}{=} \mathbb{E}(X^2) - (\mathbb{E}X)^2 + \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = \text{var}(X) + \text{var}(Y). \end{aligned}$$

Kovarians

Låt X och Y vara två s.v. med $\mathbb{E}[X] = \mu_X$ och $\mathbb{E}[Y] = \mu_Y$.

- **Kovariansen** $\text{Cov}[X, Y]$ eller σ_{XY} är ett mått på samvariation mellan X och Y och defineras som

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

- Med väntevärdets linjäritet följer direkt

$$\text{Cov}[X, Y] = \mathbb{E}(XY) - \mu_X \mathbb{E}Y - \mu_Y \mathbb{E}X + \mu_X \mu_Y = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y].$$
- Specialfall: $\text{Cov}(X, X) = \text{Var}(X)$ och om X och Y är oberoende s.v. så gäller $\text{Cov}[X, Y] = 0$.
- Vidare är kovariansen symmetrisk och linjär i båda argument:
 - 1 $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
 - 2 $\text{Cov}[cX, Y] = c \cdot \text{Cov}[X, Y] = \text{Cov}[X, cY]$
 - 3 $\text{Cov}[X + Y, Z] = \text{Cov}[X, Z] + \text{Cov}[Y, Z].$

Korrelation

- Pearsons **korrelationskoefficient** ρ_{XY} är ett mått som beskriver huruvida det finns en linjär relation mellan två utfall/stokastiska variabler X och Y .
- ρ_{XY} definieras som

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}}.$$

OBS: ρ_{XY} är odefinierad om antingen X eller Y är konstant.

- Man kan visa att $-1 \leq \rho_{XY} \leq 1$.
- $|\rho_{XY}| = 1$ om och endast om $Y = a + bX$ för några tal $a, b \in \mathbb{R}, b \neq 0$.
- $\rho_{XY} = 0$ betyder att det inte finns någon linjär relation mellan X och Y (okorrelerade) vilket dock är svagare än oberoende!

okorrelerade men inte oberoende

Exempel

Låt X vara likformigt fördelad på $\{-1, 0, 1\}$ och $Y = X^2$. Då har (X, Y) följande sannolikhetsfunktion $f_{XY}(x, y)$:

X	-1	0	1
Y			
0	0	$\frac{1}{3}$	0
1	$\frac{1}{3}$	0	$\frac{1}{3}$

$$\mathbb{E}(X) = 0, \quad \mathbb{E}(Y) = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3} \rightarrow \mathbb{E}X \cdot \mathbb{E}Y = 0$$

$$\mathbb{E}(X \cdot Y) = \frac{1}{3}(-1 \cdot 1) + \frac{1}{3}(0 \cdot 0) + \frac{1}{3}(1 \cdot 1) = 0$$

$$\text{cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = 0 - 0 = 0 \rightarrow \rho_{XY} = 0$$

ävs. X, Y okorrelerade.

Doch inte oberoende: $\mathbb{P}(X=0, Y=0) = \frac{1}{3} \neq \mathbb{P}(X=0) \cdot \mathbb{P}(Y=0) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

Betingade fördelningar

- Om A och B är händelser och $\mathbb{P}(B) > 0$, defineras sannolikheten av A betingad på B som (jfr. föreläsning 1)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- Den betingade sannolikhetsfunktionen/täthetsfunktionen för den stokastiska variabeln X **givet** $Y = y$ definieras som:

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

och den betingade sannolikhetsfunktionen/täthetsfunktionen för Y **givet** $X = x$ definieras som:

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

Betingade fördelningar

OBS: Om X, Y är oberoende slumpvariabler sammanfaller de betingade sannolikhets-/täthetsfunktioner med de marginella:

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} \stackrel{\text{obs.}}{=} \frac{f_X(x) \cdot \cancel{f_Y(y)}}{\cancel{f_Y(y)}} = f_X(x) \quad (\text{på samma sätt: } f_{Y|X}(y) = f_Y(y))$$

Exempel

I det diskreta exemplet ovan ges den betingade sannolikhetsfunktionen av X givet $Y = 2$ som

$$f_{X|2}(x) = \frac{f_{XY}(x, 2)}{f_Y(2)},$$

alltså

$$f_{X|2}(0) = \frac{0.05}{0.08} = \frac{5}{8}, \quad f_{X|2}(1) = \frac{0.02}{0.08} = \frac{1}{4}, \quad f_{X|2}(2) = \frac{0.01}{0.08} = \frac{1}{8},$$

samt $f_{X|2}(3) = f_{X|2}(4) = 0$.