STK1100 våren 2023

Simultanfordelinger for stokastiske variabler

Svarer til avsnitt 5.1 i læreboka

Matematisk institutt
Universitetet i Oslo

Eksempel – Fordeling av karakterer

Fordeling av karakterer i prosent ved eksamen i norsk og matematikk i videregående skole (konstruerte tall)

NORSK

MATEMATIKK	1	2	3	4	5	6	Sum
1	1	4	5	3	-	-	13
2	1	4	11	6	-	-	22
3	1	4	8	8	2	-	23
4	-	3	7	6	4	-	20
5	-	1	3	6	5	1	16
6	-	-	1	2	2	1	6
Sum	3	16	35	31	13	2	100

Simultanfordeling for diskrete stokastiske variabler

For to diskrete stokastiske variabler X og Y er den simultane punktsannsynligheten gitt ved

$$p(x,y) = P(X = x \text{ og } Y = y)$$

La A være en mengde av mulige verdier for (x,y). Da er

$$P\left[(X,Y) \in A\right] = \sum_{(x,y)\in A}\sum p(x,y)$$

Merk at $p(x,y) \geq 0$ og at $\sum\sum_{(x,y)} p(x,y) = 1$

Eksempel – Fordeling av karakterer

La X være karakteren i norsk og Y karakteren i matematikk for en tilfeldig valgt elev

Punktsannynligheten $p(x,y)$ er gitt ved tabellen (sannsynlighetene er gitt som prosent)

y \ x	1	2	3	4	5	6
1	1	4	5	3	-	-
2	1	4	11	6	-	-
3	1	4	8	8	2	-
4	-	3	7	6	4	-
5	-	1	3	6	5	1
6	-	-	1	2	2	1

$$P(X+Y \geq 9) = 0.22 \quad (22\%)$$

Marginalfordeling for diskrete stokastiske variabler

Den marginale punktsannsynligheten for X er

$$p_X(x) = P(X = x)$$

$$= P((X = x) \text{ og } (Y \text{ har en eller annen verdi}))$$

$$= P\left[\bigcup_y (X = x \text{ og } Y = y)\right]$$

$$= \sum_y P(X = x \text{ og } Y = y)$$

$$= \sum_y p(x,y)$$

Den marginale punktsannsynligheten for Y er

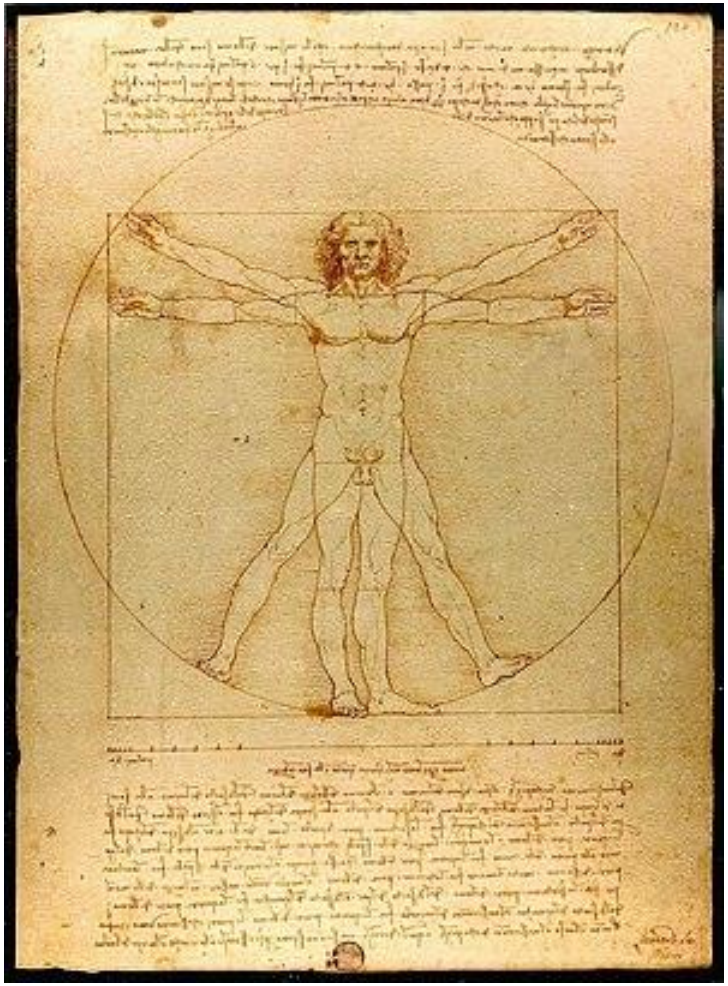
$$p_Y(y) = \sum_x p(x,y)$$

Eksempel – Fordeling av karakterer

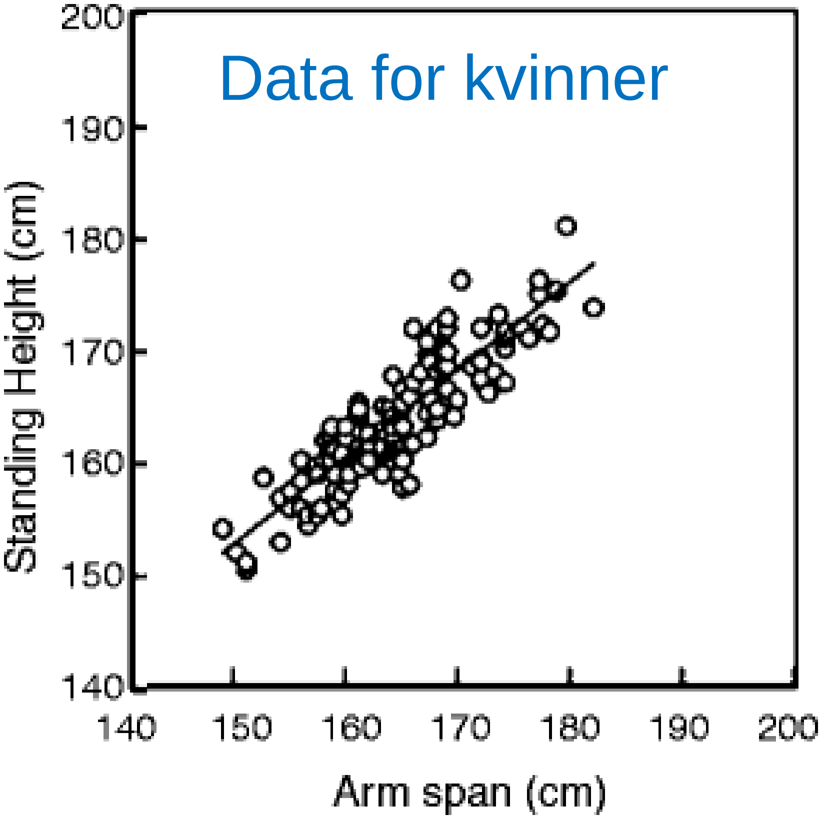
Den marginale punktsannsynligheten til X er i nederste rad av tabellen (sannsynlighetene er gitt som prosent)

Den marginale punktsannsynligheten til Y er i høyre kolonne av tabellen

y \ x	1	2	3	4	5	6	$P(Y=y)$
1	1	4	5	3	-	-	**13**
2	1	4	11	6	-	-	**22**
3	1	4	8	8	2	-	**23**
4	-	3	7	6	4	-	**20**
5	-	1	3	6	5	1	**16**
6	-	-	1	2	2	1	**6**
$P(X=x)$	**3**	**16**	**35**	**31**	**13**	**2**	**100**



Leonardo da Vinci ca 1490



Vi har to kontinuerlige stokastiske variabler

X = rekkevidde

Y = høyde

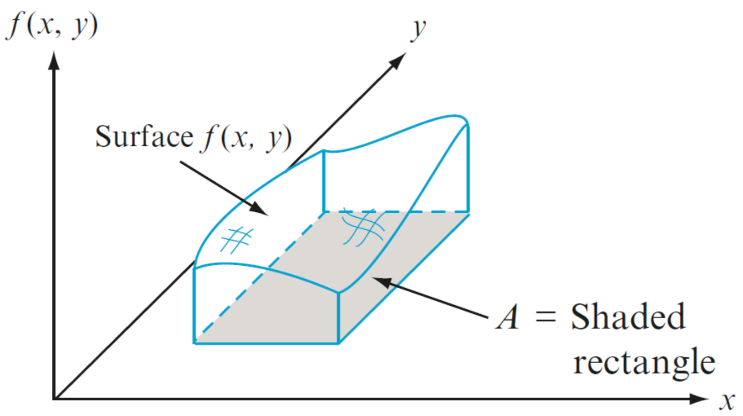
Vi vil se hvordan vi kan lage modeller for den simultane fordelingen av to kontinuerlige stokastiske variabler

Simultanfordeling for kontinuerlige stokastiske variabler

For to kontinuerlige stokastiske variabler X og Y har vi en simultan synlighetstetthet $f(x,y)$

La $A \subset R^2$, dvs. at A er et område i planet. Da er

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



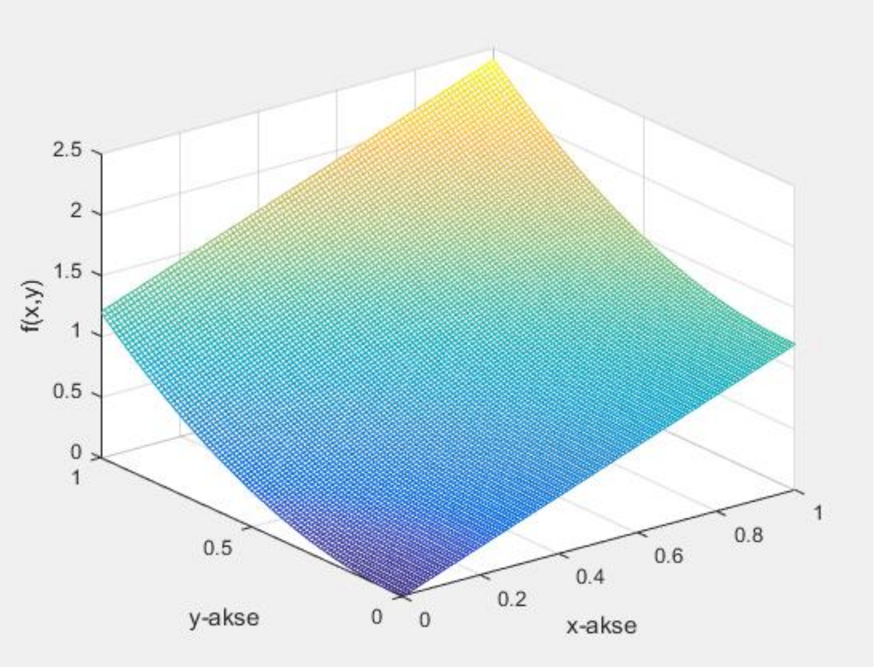
Sannsynligheten er volumet under den simultane tettheten over området A

Merk at $f(x,y) \geq 0$ og $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Eksempel 5.3 i Devore & Berk

X og Y har simultantetthet

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{ellers} \end{cases}$$



Vi har at

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 \frac{6}{5}(x+y^2)\,dx\,dy = 1$$

slik det skal være for en simultantetthet

Vi har videre at

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right)$$

$$= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)\,dx\,dy$$

$$= \frac{7}{640} = 0.011$$

Marginal sannsynlighetstetthet for kontinuerlige stokastiske variabler

De marginale sannsynlighetstetthetene for X og Y er

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

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

Eksempel 5.4 i Devore & Berk

X og Y har simultantetthet

$$f(x,y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{ellers} \end{cases}$$

De marginale tetthetene er (for $0 \le x \le 1, \quad 0 \le y \le 1$)

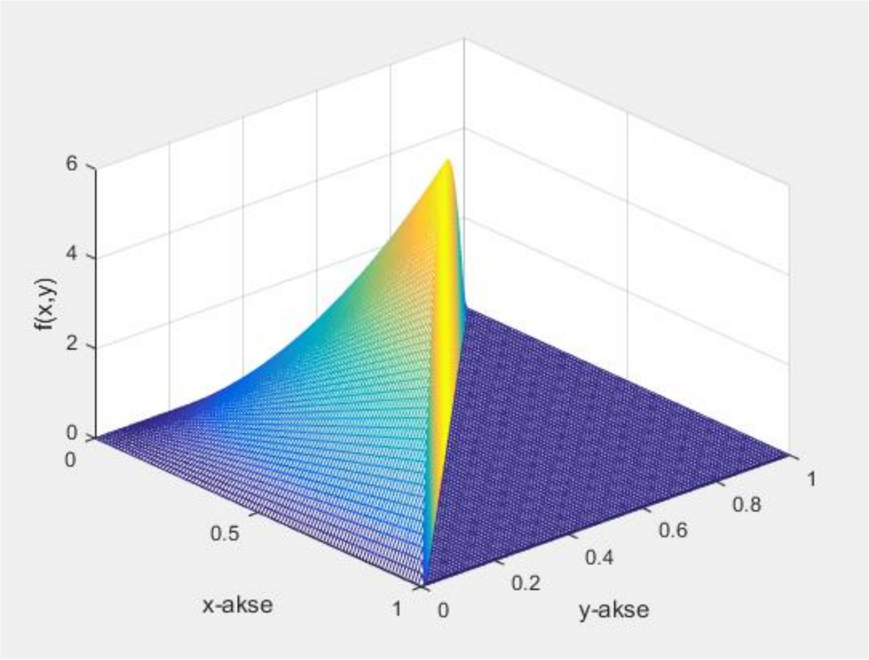
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{6}{5}(x + y^2)dy = \frac{6}{5}x + \frac{2}{5}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 \frac{6}{5}(x + y^2)dx = \frac{6}{5}y^2 + \frac{3}{5}$$

Eksempel 5.5 i Devore & Berk

X og Y har simultantetthet

$$f(x,y) = \begin{cases} 24xy & 0 \le x \le 1, \quad 0 \le y \le 1, \quad x + y \le 1 \\ 0 & \text{ellers} \end{cases}$$



Merk at

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{0}^{1}\int_{0}^{1-x} 24xy\,dy\,dx = 1$$

La $A = \left\{(x,y): 0 \le x \le 1, 0 \le y \le 1, x + y \le 1, y \le \frac{x}{2}\right\}$

Da er

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

$$= \int_{0}^{1/3}\int_{2y}^{1-y} 24xy\;dx\;dy$$

$$= \frac{7}{27}$$

Uavhengige stokastiske variabler

To begivenheter *A* og *B* er uavhengige så sant

$$P(A \cap B) = P(A) \cdot P(B)$$

To stokastiske variabler *X* og *Y* er uavhengige
hvis vi for alle (x, y) har at

$$p(x,y) = p_X(x) \cdot p_Y(y) \qquad \text{(diskret)}$$

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \text{(kontinuerlig)}$$

Hvis det finnes par (x, y) som likheten ikke holder for, er *X* og *Y* avhengige

Hvis *X* og *Y* er uavhengige, så har vi at

$$P(a \le X \le b,\ c \le Y \le d) = P(a \le X \le b) \cdot P(c \le Y \le d)$$

Eksempel – Fordeling av karakterer

X = karakteren i norsk

Y = karakteren i matematikk

y \ x	1	2	3	4	5	6	$P(Y=y)$
1	1	4	5	3	-	-	13
2	1	4	11	6	-	-	22
3	1	4	8	8	2	-	23
4	-	3	7	6	4	-	20
5	-	1	3	6	5	1	16
6	-	-	1	2	2	1	6
$P(X=x)$	3	16	35	31	13	2	100

Vi har (f. eks.) at

$$P(X=5, Y=2) = 0 \neq 0.13 \cdot 0.22 = P(X=5) \cdot P(Y=2)$$

så X og Y er avhengige

Eksempel 5.4 i Devore & Berk (forts)

X og Y har simultan tetthet

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{ellers} \end{cases}$$

De marginale tetthetene er (for $0 \le x \le 1, \quad 0 \le y \le 1$)

$$f_X(x) = \frac{6}{5}x + \frac{2}{5} \qquad \text{og} \qquad f_Y(y) = \frac{6}{5}y^2 + \frac{3}{5}$$

Vi har at

$$f(x,y) \neq f_X(x) \cdot f_Y(y)$$

så X og Y er avhengige

Eksempel

X og Y har simultantetthet

$$f(x,y) = \begin{cases} 6e^{-2x-3y} & x > 0,\ y > 0 \\ 0 & \text{ellers} \end{cases}$$

De marginale tetthetene er (for $x,y > 0$)

$$f_X(x) = 2e^{-2x} \qquad \text{og} \qquad f_Y(y) = 3e^{-3y}$$

For alle $x,y > 0$ har vi at

$$f(x,y) = 6e^{-2x-3y} = 2e^{-2x} \cdot 3e^{-3y} = f_X(x) \cdot f_Y(y)$$

Også hvis $x \leq 0$ og/eller $y \leq 0$ har vi at

$f(x,y) = f_X(x) \cdot f_Y(y)$, så X og Y er uavhengige

Flere enn to stokastiske variabler

Hvis $X_1, X_2, \ldots, X_n$ er diskrete stokastiske variabler, så er den simultane punktsannsynligheten gitt ved

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Hvis $X_1, X_2, \ldots, X_n$ er kontinuerlige stokastiske variabler med simultan sannsynlighetstetthet $f(x_1, x_2, \ldots, x_n)$, så har vi at

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n$$

$X_1, X_2, \ldots, X_n$ er uavhengige stokastiske variabler hvis simultantettheten (-punktsannsynligheten) er produktet av de marginale tetthetene (punktsannsynlighetene)