STK1100 våren 2018

Simultane, marginale og betingede fordelinger

·Repetisjon av avsnitt 5.1 og deler av avsnitt 5.3 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

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

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

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

$$p_X(x) = P(X = x) \; = \sum_y p(x,y)$$

Den marginale punktsannsynligheten for Y er

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

Eksempel (forts)

Den marginale punktsannsynligheten til X er gitt 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

Betingete fordeling for diskrete stokastiske variabler

Den betingete punktsannsynligheten for Y gitt at er $X = x$ er gitt ved

$$p_{Y|X}(y \mid x) = P(Y = y \mid X = x)$$

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

Den betingete punktsannsynligheten for X gitt at er $Y = y$ er gitt ved

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{p(x,y)}{p_Y(y)}$$

Eksempel (forts)

Simultan og marginal punktsannsynlighet for X = karakter i norsk og Y = karakter i matematikk er gitt i 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

Gitt at $X = 5$, har vi at

$$p_{Y|X}(4 \mid 5) = \frac{p(5,4)}{p_X(5)} = \frac{4/100}{13/100} = 0.31 \quad (31\%)$$

Betinget forventning og varians for diskrete stokastiske variabler

Den betingete forventningen for Y gitt at $X = x$ er gitt ved

$$\mu_{Y|X=x} = E(Y \mid X = x) = \sum_y y \cdot p_{Y|X}(y \mid x)$$

Den betingete variansen for Y gitt at $X = x$ er gitt ved

$$\underbrace{\sigma^2_{Y|X=x}} = V(Y \mid X = x) = \sum_y \underbrace{(y - \mu_{Y|X=x})^2}_{\text{kvadratavvik}} \cdot \underbrace{p_{Y|X}(y \mid x)}_{\text{sannsynligheter}}$$

Tilsvarende gjelder for betinget forventning og varians for X gitt at $Y = y$

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

Betinget fordeling for Y gitt $x=5$, dvs.

$$P(y|x) = P(Y=y \mid X=5) = \frac{P(\ )}{P_x(5)}$$

$P_{Y|X}(1|5) = 0$ $\quad P_{Y|X}(3|5) = \frac{2}{13}$ $\quad P_{Y|X}(5|5) = \frac{5}{13}$

$P_{Y|X}(2|5) = 0$ $\quad P_{Y|X}(4|5) = \frac{4}{13}$ $\quad P_{Y|X}(6|5) = \frac{2}{13}$

Hva er betinget forventning for Y gitt $x=5$

$$\mu_{Y|x=5} = \sum_{y=1}^{6} y\, P_{Y|X}(y|5) = 0 + 0 + 3\frac{3}{13} + 4\frac{4}{13} + 5\frac{5}{13} + 6\frac{2}{13}$$

$$= \frac{59}{13} = \underline{4{,}54}$$

Eksempel (forts)

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

$p_{Y|X}(y|5)$ er en sannsynlighetsfordeling

Gitt at $X = 5$, er forventet verdi av Y

$$\mu_{Y|X=5} = \sum_{y=1}^{6} y \cdot p_{Y|X}(y\,|\,5) = 4.54$$

Gitt at $X = 3$, er forventet verdi av Y

$$\mu_{Y|X=3} = \sum_{y=1}^{6} y \cdot p_{Y|X}(y\,|\,3) = 2.86$$

(MATLAB kommandoer på neste slide)

MATLAB kommandoer for å beregne de betingete forventningene på forrige slide

```
[x,y]=meshgrid(1:6,1:6);
pxy=[1 4  5  3 0 0;
       1 4 11 6 0 0;
       1 4  8  8 2 0;
       0 3  7  6 4 0;
       0 1  3  6 5 1;
       0 0  1  2 2 1]/100;
px=sum(pxy)
sum((1:6)'.*pxy(x==5)/px(5))
sum((1:6)'.*pxy(x==3)/px(3))
```

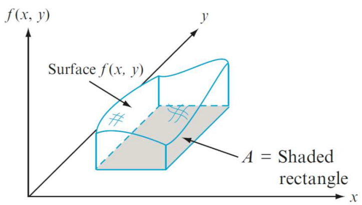
10

Simultanfordeling for kontinuerlige stokastiske variabler

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

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

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



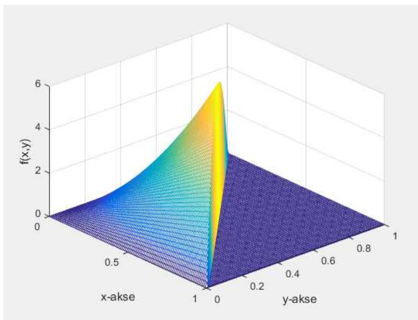
Sannsynligheten er volumet under den simultane tettheten over området A

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

Eksempel (fra læreboka)

X og Y har simultan tetthet

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

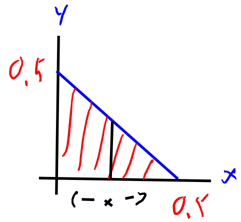


MATLAB:

```
x=0:0.01:1;
y=0:0.01:1;
[x,y]=meshgrid(x,y);
fxy=24*x.*y.*(x+y <=1);
mesh(x,y,fxy)
xlabel('x-akse')
ylabel('y-akse')
zlabel('f(x,y)')
```

La $A = \{(x,y): 0 \le x \le 1,\ 0 \le y \le 1,\ x + y \le 0.5\}$

Da er



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

$y \le 0.5 - x$

$$= \int_0^{0.5} \int_0^{0.5-x} 24xy\,dy\,dx$$

$$= 0.063$$

Marginal sannsynlighetstetthet for kontinuerlige stokastiske variabler

De marginale sannsynlighetstetthetene for X og Y er

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

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

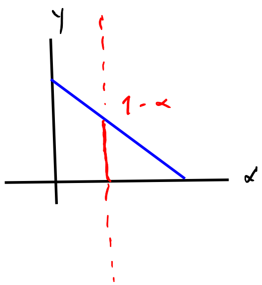
Eksempel (forts)

X og Y har simultantetthet

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

Den marginale tettheten for X er (for $0 \le x \le 1$)

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



$$= \int_0^{1-x} 24xy\,dy$$

$$= 12x(1-x)^2$$

Betingete sannsynlighetstettheter for kontinuerlige stokastiske variabler

Den betingede sannsynlighetstettheten for Y gitt at $X = x$ er gitt ved

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

Den betingede sannsynlighetstettheten for X gitt at $Y = y$ er gitt ved

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

Betingede tettheter

Vi har at

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f_X(\,)\dots \approx f_X(x)\Delta x$$

Tilsvarende

$$P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y) \approx f(x,y)\,\Delta x \Delta y$$

Har dermed at

$$P(y \leq Y \leq y + \Delta y \mid x \leq X \leq x + \Delta x) = \frac{P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y)}{P(x \leq X \leq x + \Delta x)}$$

$$\approx \frac{f(x,y)\,\Delta x \Delta y}{f_X(x)\,\Delta x} = \underbrace{\frac{f(x,y)}{f_X(x)}}_{f_{Y|X}(y|x)} \Delta y$$

Eksempel (forts)

X og Y har simultan tetthet

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

og X har marginal tetthet (for $0\le x\le 1$)

$$f_X(x)=12x(1-x)^2$$

Den betingete tettheten for Y gitt $X = x$ er (for $0\le y\le 1-x$)

gitt i [0,1]

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}=\frac{24xy}{12x(1-x)^2}=\frac{2y}{(1-x)^2}$$

Betinget forventning og varians for kontinuerlige stokastiske variabler

Den betingete forventningen for Y gitt at $X=x$ er gitt ved

Tetthet

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

Den betingete variansen for Y gitt at $X=x$ er gitt ved

$$\sigma^2_{Y|X=x}=V(Y\mid X=x)=\int_{-\infty}^{\infty}(y-\mu_{Y|X=x})^2\cdot f_{Y|X}(y\mid x)dy$$

Tilsvarende gjelder for betinget forventning og varians for X gitt at $Y=y$

Eksempel (forts)

Den betingete tettheten for Y gitt $X = x$ er (for $0\le y\le 1-x$)

$$f_{Y|X}(y\mid x)=\frac{2y}{(1-x)^2}$$

Den betingete forventningen til Y gitt $X = x$ er

E(Y | X = x) er en funksjon av x

$$\mu_{Y|X=x}=\int_{-\infty}^{\infty}y\cdot f_{Y|X}(y\mid x)dy$$

$$=\int_{0}^{1-x}y\cdot\frac{2y}{(1-x)^2}dy=\frac{2}{3}(1-x)$$

Uavhengige stokastiske variabler

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

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

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

X og Y er uavhengige hvis og bare hvis de betingete punktsannsynlighetene / tetthetene lik de marginale punktsannsynlighetene / tetthetene

Uavhengighet og betingede fordelinger

Anta X og Y uavhengige

Da kan vi

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

⋮

Så er

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

og det gir

$$f(x,y) = f_X(x) f_Y(y)$$

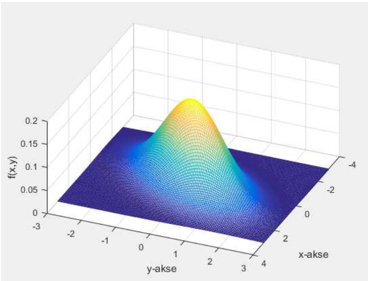
dvs. X og Y ...

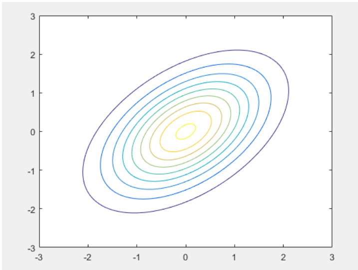
Binormalfordeling

X og *Y* er binormalt fordelt med simultantetthet

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right] \Big/ \left[2(1-\rho^2)\right] \right\}$$

Illustrasjon for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ og $\rho = 0.50$
(MATLAB kommandoer på neste slide)





MATLAB:

```
rho=0.5;
x=-3:0.05:3;
y=-3:0.05:3;
[x,y]=meshgrid(x,y);
fxy= (2*pi*sqrt(1-rho^2))^(-1)*exp(-(x.^2-2*rho*x.*y+y.^2)/(2*(1-rho^2)));
mesh(x,y,fxy);
xlabel('x-akse');
ylabel('y-akse');
zlabel('f(x,y)');

contour(x,y,fxy);
```

Den marginale tettheten for *X* er (krever en del regning)

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma_1^2}(x-\mu_1)^2 \right\}$$

Altså er $X \sim N(\mu_1, \sigma_1^2)$

Tilsvarende er $Y \sim N(\mu_2, \sigma_2^2)$

Videre har vi at $\rho = \text{corr}(X,Y)$

Hvis $\rho = 0$ er *X* og *Y* uavhengige