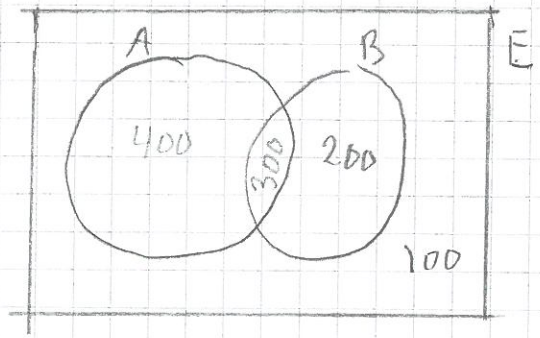
TRON FOSS

Litt repetisjon av betingede sannsynligheter.
Anta at du spør $N = 1000$ personer om de har lest Aftenposten eller Buskerud Blad.
La $n(A)$ være antall personer som svarte at de hadde lest Aftenposten
La $n(B)$ være antall personer som svarte at de hadde lest Buskerud Bla. Osv.

Svarene var $n(A) = 700$, $n(B) = 500$.
$n(A \cap B) = 300$. Se fig.



Fra fig ser vi

$$P(A) = \frac{n(A)}{n(E)} = \frac{700}{1000} = 0.7$$

$$P(B) = \frac{500}{1000} = 0.5$$

Anta at vi trekker ut en tilfeldig person og at han svarer at han har lest B.
Hva er da sannsynligheten for at han har lest A. $P(A \mid B)$.

Vi vet at vi har trukket ut en av de 500 personene som har lest B. Av disse er det 300 som også har lest A

Sannsynligheten for at han har lest A gitt at han har lest B er derfor.

$$P(A|B) = \frac{300}{500} = 0.6$$

Dette kunne vi også beregnet ved å benytte formelen $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{300}{1000}}{\frac{500}{1000}} = 0.6$.

A og B er avhengige begivenheter da

$$P(A \cap B) = 0.3 \neq P(A) \cdot P(B) = 0.7 \cdot 0.5 = 0.35.$$

La oss innføre følgende stokastiske variable.

La $X = 1$ hvis personen leser A, null ellers

La $Y = 1$ hvis personen leser B, null ellers.

La $p(x,y) = P(X=x, Y=y)$ vi får da.

$p(0,0) = 0.1$ $\quad$ $p(0,1) = 0.2$

$p(1,0) = 0.4$ $\quad$ $p(1,1) = 0.3$.

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

Nå er $p_Y(0) = p(0,0) + p(1,0) = 0.1 + 0.4 = 0.5$.

$p_Y(1) = p(0,1) + p(1,1) = 0.2 + 0.3 = 0.5$

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

$$p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{0.2}{0.5} = 0.4$$

$$p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{0.3}{0.5} = 0.6.$$

Dette tilsvarer $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.5} = 0.6.$

Dersom den simultane fordelingen er kontinuerlig vil den betingede fordelingen være

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

Eksempel.

$$f(x,y) = 6xy(2-x-y) \quad \begin{matrix} 0<x<1 \\ 0<y<1 \end{matrix}$$

$f(x,y) = 0$ ellers.

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

Beregner først

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

$$\int_0^1 6xy(2-x-y)\,dx = \int_0^1 (12xy - 6x^2y - 6xy^2)\,dx$$

$$= \left[12\cdot\frac{x^2}{2}y - 6\cdot\frac{x^3}{3}y - 6\frac{x^2}{2}\cdot y^2\right]_0^1 =$$

$$6y - 2y - 3y^2$$

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

$$\frac{6xy(2-x-y)}{4y-3y^2} =$$

$$\frac{6xy(2-x-y)}{y(4-3y)} = \frac{6x(2-x-y)}{4-3y}.$$

$$E[X \mid Y=y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\,dx =$$

$$\int_0^1 x \frac{6x(2-x-y)}{4-3y}\,dx = \frac{1}{4-3y}\int_0^1 x(12x-6x^2-6xy)\,dx =$$

$$\frac{1}{4-3y}\int_0^1 (12x^2-6x^3-6x^2y)\,dx =$$

$$\frac{1}{4-3y}\left[12\frac{x^3}{3} - 6\cdot\frac{x^4}{4} - 6\cdot\frac{x^3}{3}y\right]_0^1$$

$$\frac{1}{4-3y}\left[4-\frac{3}{2}-2y\right] = \frac{5-4y}{8-6y}$$

Eksempel:

$$f(x,y) = 12\,\frac{x-x^2}{(y+1)^2} \qquad \begin{matrix} 0<x<1 \\ 0<y<1 \end{matrix}$$

$$f_Y(y) = \int_0^1 12\,\frac{x-x^2}{(y+1)^2}\,dx =$$

$$\frac{12}{(y+1)^2}\int_0^1 (x-x^2)\,dx =$$

$$\frac{12}{(y+1)^2}\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1 = \frac{12}{(y+1)^2}\left[\frac{1}{2}-\frac{1}{3}\right]$$

$$\frac{12}{(y+1)^2}\left[\frac{1}{6}\right] = \frac{2}{(y+1)^2}$$

Jeg vil vise at $f_Y(y)$ er en sannsynlighets tetthet.

Beregner $\int \frac{2}{(y+1)^2}\,dy$. Jeg setter $u = y+1$

$du = dy$. $\int \frac{2}{(y+1)^2}\,dy = \int \frac{2}{u^2}\,du = \int 2u^{-2}\,du =$

$$2\cdot\frac{u^{-1}}{-1} + C = -\frac{2}{u} + C = -\frac{2}{y+1} + C$$

$$\int_0^1 \frac{2}{(y+1)^2}\,dy = \left[-\frac{2}{y+1}\right]_0^1 = -\frac{2}{1+1} + \frac{2}{1} = 1$$

Dessuten er $f_Y(y) = \frac{2}{(y+1)^2} \geq 0$ for alle $0 < y < 1$.

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

$$\frac{\frac{12(x-x^2)}{(y+1)^2}}{\frac{2}{(y+1)^2}} = 6(x-x^2)$$

$$E[X \mid Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx = \int_0^1 x \cdot 6(x-x^2)\,dx =$$

$$\int_0^1 (6x^2 - 6x^3)\,dx = \left[6 \cdot \frac{x^3}{3} - 6 \cdot \frac{x^4}{4}\right]_0^1 = 2 - \frac{3}{2} = \frac{1}{2}$$

Her legger vi merke til at $E[X \mid Y=y]$ er uavhengig av y.

Dette kommer av at X og Y er uavhengige stokastiske variable.

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

$$\int_0^1 12\,\frac{x-x^2}{(y+1)^2}\,dy = 6(x-x^2)\int_0^1 \frac{2}{(y+1)^2}\,dy =$$

$$\cancel{12(x-x^2)} \qquad 6(x-x^2)\cdot 1 = 6(x-x^2).$$

Vi får derfor

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

$$6(x-x^2)\cdot\frac{2}{(y+1)^2} = \frac{12(x-x^2}{(y+1)^2}$$

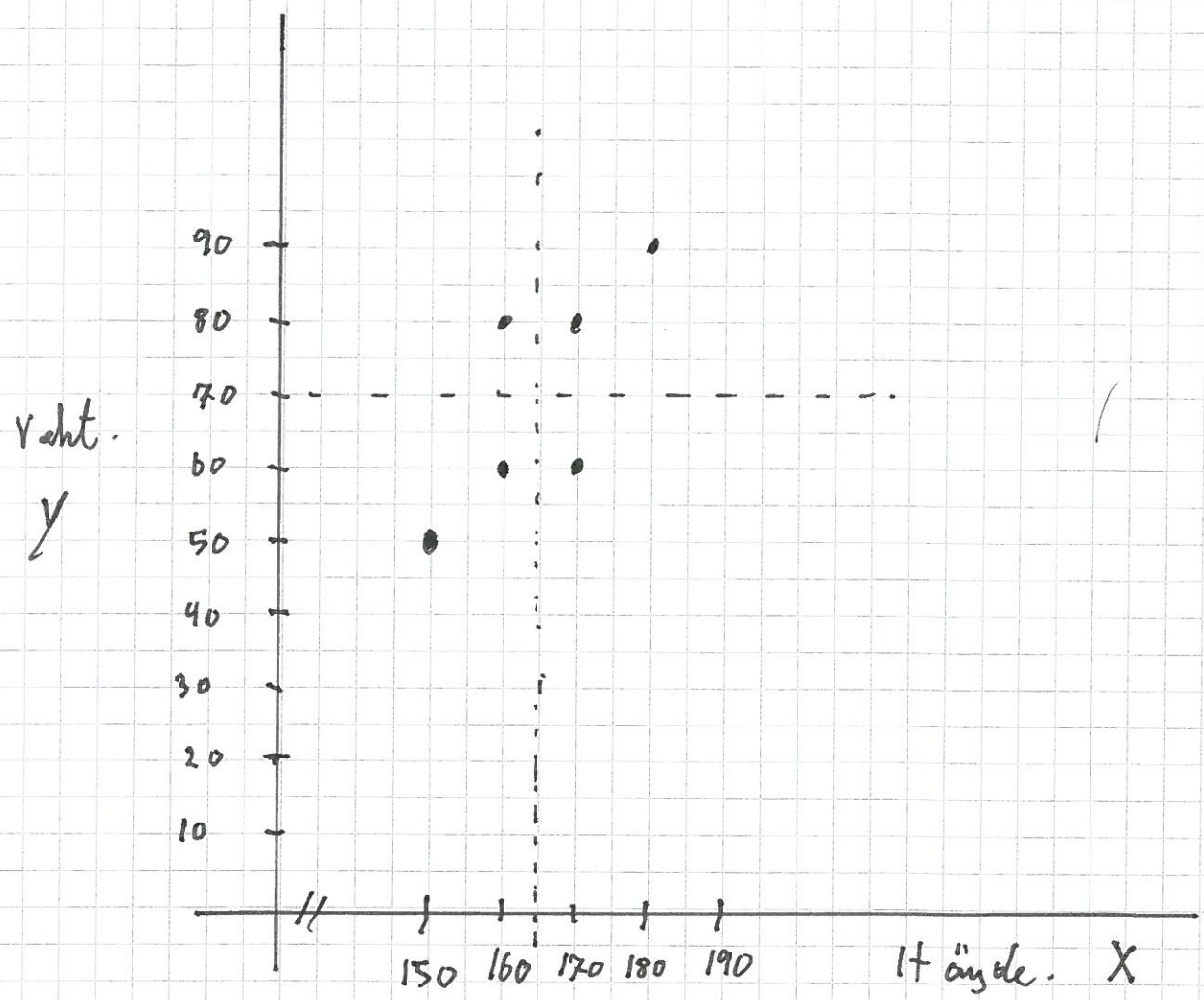
Som viser at X og Y er uavhengige.

Generelt har vi: (hvis X og Y er uavhengige)

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

$$\frac{f_X(x)\cdot f_Y(y)}{f_Y(y)} = f_X(x)$$

Kovarians



I populasjonen ovenfor er $\mu_x = 165$

$\mu_y = 70$.

$$Cov(X,Y) = \sum (X-\mu_x)(Y-\mu_y) =$$

$$\frac{(150-165)(50-70)+(160-165)(60-70)+\cdots+(180-165)(90-70)}{6} = 100$$

Trekker vi en person tilfeldig fra populasjonen og lar X være høyden, og Y vektens vi X, Y være simultanfordelt. der.

$$P(X=x, Y=y) = \frac{1}{6}.$$

$$Cov(X, Y) = E(X-\mu_x)(Y-\mu_y) =$$

$$E(XY) - \mu_x \mu_y = \sum xy\, P(x,y) - \mu_x \mu_y =$$

$$150 \cdot 50 \cdot \frac{1}{6} + 160 \cdot 60 \cdot \frac{1}{6} + 160 \cdot 80 \cdot \frac{1}{6} +$$

$$170 \cdot 60 \cdot \frac{1}{6} + 170 \cdot 80 \cdot \frac{1}{6} + 180 \cdot 90 \cdot \frac{1}{6} - 165 \cdot 70 =$$

$$\frac{69900}{6} - 165 \cdot 70 = 11650 - 165 \cdot 70 = 100$$

Hvis X er en diskret stokastisk variabel med sannsynlighetstetthet $p(x)$, da vil for alle reelle funksjoner g.

$$E[g(X)] = \sum_{x:p(X)>0} g(x)p(x)$$

Hvis X er en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f(x)$, da vil for alle reelle funksjoner g

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Hvis a og b er konstanter så er $E[aX+b] = aE[X]+b$
$E[aX+bY] = aE[X]+bE[Y]$

La $E[X] = \mu$. Da er $Var(X) = E\left[(X-\mu)^2\right] = E\left[X^2\right] - \mu^2$

La X og Y være simultanfordelte diskrete stokastiske variabler med sannsynlighetstetthet $p(x,y)$ og $g(x,y)$ en reel fukjon ,da er

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

La X og Y være simultanfordelte kontinuerlige stokastiske variabler med sannsynlighetstetthet $f(x,y)$ og $g(x,y)$ en reel fukjon ,da er

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \quad .$$

Er $g(x,y) = x+y$ vil

$$E[X+Y] = E[X]+E[Y]$$

Er X og Y uavhengige vil $E[g(X)h(Y)] = E[g(X)]\,E[h(Y)]$

La X og Y være to stokastiske variable, og la $E[X] = \mu_X$, $E[Y] = \mu_Y$

Kovariansen mellom to stokastiske variabler er definert ved
$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY]$ -$\mu_X \cdot \mu_Y$

Egenskaper ved kovariansen.

$Cov(X,X) = Var(X)$
$Cov(X,Y) = Cov(Y,X)$
$Cov(cX,Y) = cCov(X,Y)$
$Cov(X,Y+Z) = Cov(X,Y)+Cov(X,Z)$

$$Cov(\sum_{i=1}^{n} X_{i,} \sum_{j=1}^{n} Y_{i,}) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, Y_j)$$

Spesielt vil $Var(\sum_{i=1}^{n} X_{i,}) = Cov(\sum_{i=1}^{n} X_{i,} \sum_{j=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$

Er $n = 2$ vil $Var(X_1 + X_2) = Cov(X_1 + X_2, X_1 + X_2) =$
$Cov(X_{1,}X_1) + Cov(X_{1,}X_2) + Cov(X_{2,}X_1) + Cov(X_{2,}X_2) =$
$Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

Dersom $X_{1,}.....X_n$ er uavhengige og identisk fordelte med forventning μ og varians σ^2 så vil

$$Var(\sum_{i=1}^{n} X_{i,}) = \sum_{i=1}^{n} Var(X_i)$$
$$E\left[\overline{X}\right] = \mu$$
$$Var\left[\overline{X}\right] = \sigma^2/n$$
$$Cov(\overline{X}, X_i - \overline{X}) = 0$$