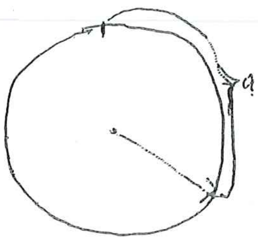
3.3 Kontinuerlege sannsynsfordelingar

Eks. X = høgda av tilfeldig valgt mann

Y = levetida til ei lyspære

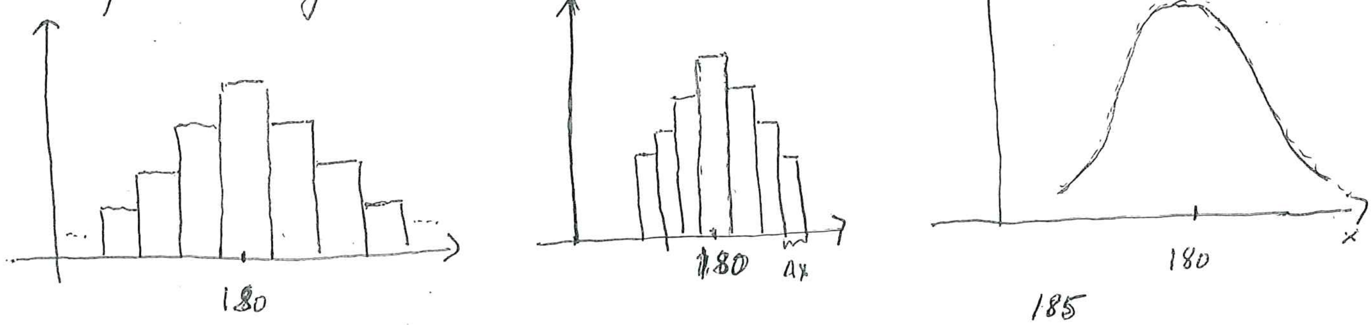
Z = avstanden, a, målt i radianar frå eit visst startpunkt til der ei pil stansar $0 \le a \le 2\pi$



For kontinuerleg fordelte variable er alle punktsannsyn 0

Men naturleg at $P(0 \le Z \le a) = \dfrac{a}{2\pi}$

histogram over høgde $\rightarrow$ mange observasjonar — uendeleg mange



$$P(175 \le X \le 185) \approx \sum_{175 \le X \le 185} f(x)\,\Delta x \longrightarrow \int_{175}^{185} f(x)\,dx$$

Legg merke til at vi har:

$$P(a \le X \le b) = P(X=a) + P(a < X < b) + P(X=b)$$

$$= P(a < X < b) = \int_a^b f(x)\,dx$$

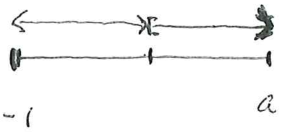
<u>Definisjon 3.6</u>. $f(x)$ er ein sannsynstettleik for ein kontinuerleg variabel X dersom

1. $f(x) \geq 0$
2. $\int f(x)\,dx = 1$
3. $P(a \leq X < b) = \int_a^b f(x)\,dx$

<u>Definisjon 3.7</u>. Kumulativ fordelingsfunksjon til ein variabel X med sannsynstettleik $f(x)$ er:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \quad \text{for} \quad -\infty < x < \infty$$

<u>Eks</u>. Avrundingsfeil ved målingar. Avrunding til næraste heiltal. La X vere avrundingsfeil.



$$f(x) = \begin{cases} 1 & -0.5 < x \leq 0.5 \\ 0 & \text{elles.} \end{cases}$$

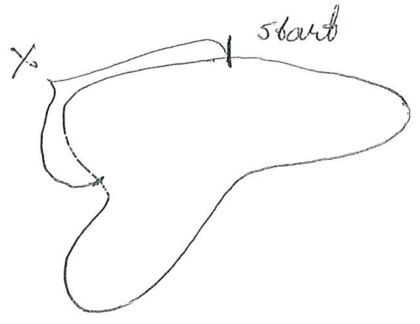
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x \leq 0.5 \\ \int_{-\frac{1}{2}}^{x} 1\,dt, & -0.5 < x \leq 0.5 \\ 1, & x > 0.5 \end{cases} = \begin{cases} 0, & x \leq 0.5 \\ x + \frac{1}{2}, & -0.5 < x \leq 0.5 \\ 1, & x > 0.5 \end{cases}$$

Eks. Pilspel: $\cancel{f(x)}\ P(X \leq x) = \frac{x}{2\pi}, \quad 0 \leq x \leq 2\pi$, $\quad f(x) = \frac{1}{2\pi}, \quad 0 < x \leq 2\pi$

<u>NB!</u> $f(x) = \frac{dF(x)}{dx}$, i dei punkt der den deriverte eksisterer.

<u>Eks</u>. Joggetur i rundløype (9 km). Mista husnøkkelen.

$Y \sim$ posisjonen der husnøkkelen er mista (målt i km frå starten av løypa).

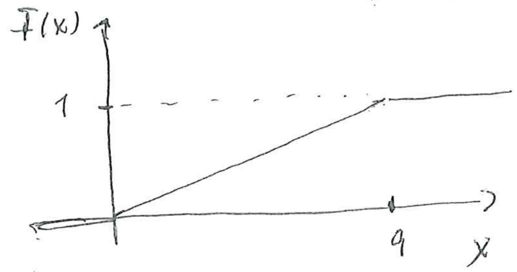


Samsyns modell

$$f(x) = \begin{cases} k, & 0 \leq x \leq 9 \\ 0, & \text{ellers} \end{cases}$$

$$\int_0^9 k \, dt = 1 \Leftrightarrow k \cdot 9 = 1 \Rightarrow k = \frac{1}{9}$$

$$F(x) = \int_{-\infty}^{x} k \, dt = \begin{cases} 0, & x < 0 \\ \int_0^x k \, dt, & 0 \leq x \leq 9 \\ 1, & x \geq 9 \end{cases} = \begin{cases} 0, & x < 0 \\ \frac{x}{9}, & 0 \leq x \leq 9 \\ 1, & x \geq 9 \end{cases}$$



$$P(4.6 \leq X \leq 6.3) = \int_{4.6}^{6.3} \frac{1}{9} \, dx$$

$$= P(X \leq 6.3) - P(X \leq 4.6)$$

$$= F(6.3) - F(4.6) = \frac{6.3 - 4.6}{9} = \frac{1.9}{9} = \underline{0.21}$$

3.4 Simultane Sannsynsfordelingar

Tef. valgt familie

X = talet på born

Y = talet på rom i tillegg til kjøkken

$x \backslash y$	1	2	3	4	$P(X=x)$
0	0.11	0.09	0.07	0.01	0.28
1	0.07	0.12	0.12	0.02	0.33
2	0.02	0.05	0.17	0.05	0.29
3	0.00	0.02	0.04	0.02	0.08
4	0.00	0.00	0.01	0.01	0.02
$P(Y=y)$	0.20	0.28	0.41	0.11	

$P(X=2, Y=3) = 0.17$

Definisjon 3.8

Funksjonen $f(x,y)$ blir kalla simultant punktsannsyn for dei diskrete variablane X og Y dersom

1. $f(x,y) \geq 0, \quad \forall (x,y)$

2. $\sum_x \sum_y f(x,y) = 1$

3. $P(X=x, Y=y) = f(x,y)$

Ein familie er trangbudd dersom $\frac{\# \text{ personar}}{\# \text{ rom}} > 2$

La Z = talet på personar. Gå ut frå at $Z = X+2$

$P(\text{tilfeldig valgt familie er trangbudd}) = P\left(\frac{X+2}{Y} > 2\right)$

$\Leftrightarrow P(X+2 > 2Y) \Leftrightarrow P(2Y < X+2) = \sum\sum_{\substack{2y<x+2 \\ y<\frac{x+2}{2}}} f(x,y)$

$= f(1,1) + f(2,1) + f(3,1) + f(4,1) + f(3,2) + f(4,2)$

$= 0.07 + 0.02 + 0 + 0 + 0.02 + 0 = 0.11$

Marginalfordelingar

$P(X = x)$?

$P(X=0) = f(0,1) + f(0,2) + f(0,3) + f(0,4) = 0.11 + 0.09 + 0.07 + 0.01 = 0.28$

$P(Y=1) = f(0,1) + f(1,1) + f(2,1) + f(3,1) + f(4,1)$

$= 0.11 + 0.07 + 0.02 + 0 + 0 = 0.20$

Definisjon 3.10

Marginalfordelinga til X er gitt ved: $P(X=x) = g(x) = \sum_y f(x,y)$

– " – " – til Y – " – $P(Y=y) = h(y) = \sum_x f(x,y)$

Betinga fordelingar

Gitt at $Y=1$. Kva er sannsynet for ingen born?

$P(X=0 \mid Y=1) = P(\{e \mid e \in S \text{ og } X(e)=0\} \mid \{e \mid e \in S \text{ og } Y(e)=1\})$

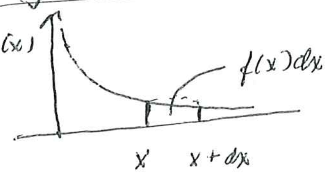
$= \frac{P(A \cap B)}{P(B)} = \frac{P(X=0, Y=1)}{P(Y=1)} = \frac{0.11}{0.2} = \underline{0.55}$

Def. 3.11

La X og Y vere to diskrete variable. Betinga fordeling for X gitt at $Y=y$, $f(x|y) \stackrel{\Delta}{=} \frac{P(X=x \cap Y=y)}{P(Y=y)}$

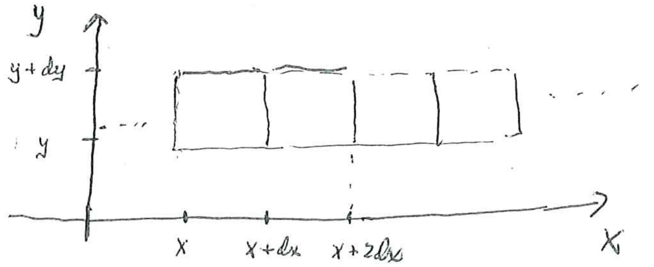
X	0	1	2	3	4
$P(X=x \mid Y=1)$	$\frac{0,11}{0,2} = 0,55$	$\frac{0,07}{0,2} = 0,35$	$\frac{0,02}{0,2} = 0,1$	$0 = 0$	$0 = 0$

3.4 Simultanfordeling for kontinuerlig fordelte variable



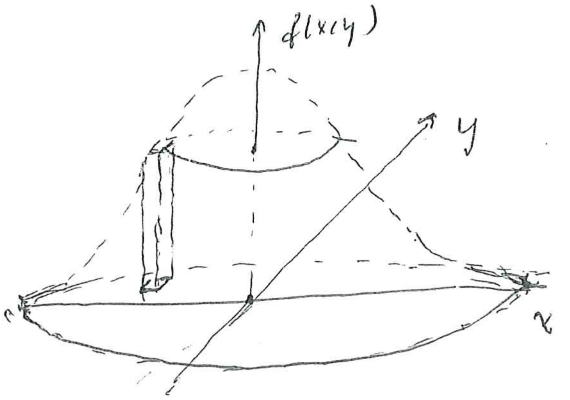
Ein variabel

$$P(x \le X \le x+dx) \approx f(x)\,dx$$



$$P(X \le x, Y \le y) = F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,ds\,dt$$

$$P(b \le X \le a,\ d \le Y \le c) = \int_{b}^{a} \int_{d}^{c} f(x,y)\,dx\,dy$$



Marginalfordeling

$$h(y)\,dy = P(y \le Y \le y+dy) = P(y \le Y \le y+dy,\ x_{min} \le X \le x_{min} + dx)$$

$$+ P(y \le Y \le y+dy,\ x_{min} + dx \le X \le x_{min} + 2dx) + \dots \approx \sum_x f(x,y)\,dx\,dy$$

slik at $\quad h(y) \approx \sum_x f(x,y)\,dx$

Definisjon 3.10

Definerer marginalfordelingane til X og Y, $g(x)$ og $h(y)$ som

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Betinga fordeling

$f(x \mid Y=y)$

$$f(x|y)\,dx \approx \lim_{dy\to 0} P(x \le X \le x+dx \mid y \le Y < y+dy) \approx \lim_{dy\to 0} \frac{f(x,y)\,dx\,dy}{h(y)\,dy}$$

$$= \frac{f(x,y)\,dx}{h(y)}, \quad \text{slik at} \quad f(x|y) = \frac{f(x,y)}{h(y)}$$

Definisjon 3.11

Definerer betinga sannsynstettleik for $X \mid Y=y$ som

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad h(y) > 0$$

Eks.

Ein brus-automat blir kun påfylt i starten av dagen.

$Y =$ påfylt brus i gallon

$X =$ tappa brus

$$f(x,y) = \begin{cases} \frac{1}{2}, & 0 \le x \le y, \; 0 \le y \le 2 \\ 0, & \text{elles} \end{cases}$$