Plan

1 Kontinuerlige stokastiske variabler
2 Forventning og varians

Repetisjon: $f(\underline{x}) = \underline{x}^T A \underline{x} + B\underline{x} + C$

Stasjonære pkt: $f'(\underline{x}) = 2A\underline{x} + B^T = \underline{0}$

$H(f) = 2A$: i) f konveks $\Leftrightarrow$ A pos. semidefin. — Stasj. pkt er: min.

ii) f konkav $\Leftrightarrow$ A neg. semidefin. — max.

iii) A indefinit — sadelpkt

Eks: Lin. regresjon $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$

$$\min \underset{\underline{\varepsilon}^T \cdot \underline{\varepsilon}}{\underset{\parallel}{f(\underline{\beta})}} = \underline{\beta}^T \underbrace{(X^T X)}_{A} \underline{\beta} \underbrace{- 2(\underline{y}^T X)}_{B} \underline{\beta} + \underbrace{\underline{y}^T \underline{y}}_{C}$$

$$f'(\underline{\beta}) = 2(X^T X)\beta - 2(y^T X)^T = \underline{0}$$

$$(X^T X)\underline{\beta} = X^T \underline{y}$$

$$f''(\underline{\beta}) = 2(X^T X) \longleftarrow X^T X \text{ pos. semidefin.} \quad (\text{Oppg } 5.8)$$

Konkl: f konveks $\Rightarrow$ alle stasjonære pkt er min.

$|X^T X| = 0$: $|X^T X| = 0 \Leftrightarrow$ kolonnene i X er lineært avhengige

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{n1} \\ 1 & x_{12} & \dots & x_{n2} \\ 1 & x_{13} & \dots & x_{n3} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1N} & \dots & x_{nN} \end{pmatrix}$$

$X^T X$ har egenverdier $\lambda_1, \lambda_2, \dots, \lambda_n \geq 0$

$|X^T X| = 0 \Leftrightarrow$ minst en $\lambda_i = 0$

$\Uparrow$

$(X \cdot \underline{v})^T (X \underline{v}) = \|X \cdot \underline{v}\|$ $X \cdot \underline{v} = \underline{0}$ for en $\underline{v} \neq \underline{0}$.

$= \underline{v}^T X^T X \underline{v} = \underline{0}$

Eks: $y = \beta_0 + \beta_1 x_1 + \beta_2 + \varepsilon$

x_1	x_2	y
4	10	6
3	9	4
4	11	7
5	15	8

$$X = \begin{pmatrix} 1 & 4 & 10 \\ 1 & 3 & 9 \\ 1 & 4 & 11 \\ 1 & 5 & 15 \end{pmatrix} \qquad y = \begin{pmatrix} 6 \\ 4 \\ 7 \\ 8 \end{pmatrix}$$

$$X^T \cdot X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 3 & 4 & 5 \\ 10 & 9 & 11 & 15 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 & 10 \\ 1 & 3 & 9 \\ 1 & 4 & 11 \\ 1 & 5 & 15 \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 16 & 45 \\ 16 & 66 & 186 \\ 45 & 186 & 527 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 25 \\ 104 \\ 293 \end{pmatrix}$$

$$\boxed{\underline{\beta} = (X^T X)^{-1} \cdot (X^T \underline{y})}$$

$$(X^T X)\,\underline{\beta} = X^T \underline{y}$$

$$\begin{pmatrix} 4 & 16 & 45 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 25 \\ \cdot \\ \cdot \end{pmatrix}$$

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 4 & 16 & 45 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}^{-1} \cdot \begin{pmatrix} 25 \\ \cdot \\ \cdot \end{pmatrix}$$

$$= \begin{pmatrix} -20/11 \\ 25/11 \\ -1/11 \end{pmatrix}$$

$$\boxed{Y = -\frac{20}{11} + \frac{25}{11} x_1 - \frac{1}{11} x_2}$$

regresjons-plant

Har hoppet over noen av mellomregningene...

1 Kontinuerlige stokastiske variabler

Defn: Stokastisk forsøk (tilfeldig)
- mulige utfall er kjent
- faktisk utfall ukjent

Stokastiske variabel X: variabel med tallverdier, som avhenger av utfallet

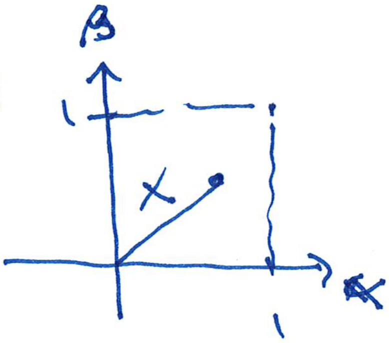
Eks: Vi kaster to terninger

X = summen av terningene

$$S = \{(1,1),(1,2),\dots,(6,6)\}$$

$$P(X=10) = 3 \cdot 1/36 = \underline{\underline{1/12}}$$

Eks:



Vi velger to tilfeldige tall (α,β) med $0 \le \alpha,\beta \le 1$.

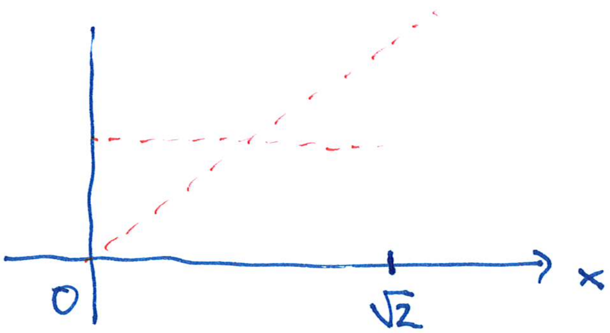
X = avstand til $(0,0)$

$= \sqrt{\alpha^2+\beta^2}$

Mulige verdier for X:

$[0, \sqrt{2}]$ kontinuerlig mengde

$$P(X \le 1/2) \qquad P(1/2 \le X \le 1)$$



Defn: Sannsynlighetstetthet $f_X(x)$

En funksjon $f(x)$ slik at

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

Har følgende egenskaper:

- $f(x) \ge 0$ for alle x
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Defn: Kumulativ sannsynlighetsfordeling $F(x)$ ← $F_X(x)$
er slik at

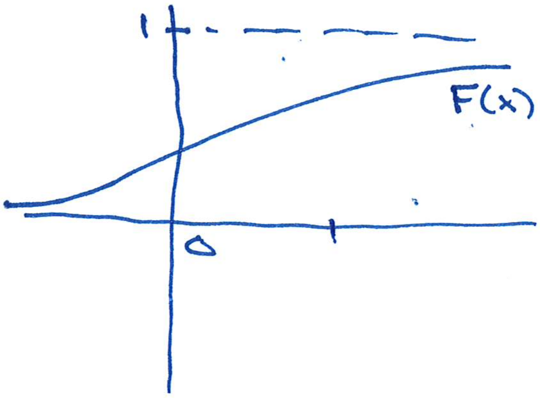
$$F(b) = \int_{-\infty}^{b} f(x)\,dx = p(X \le b)$$

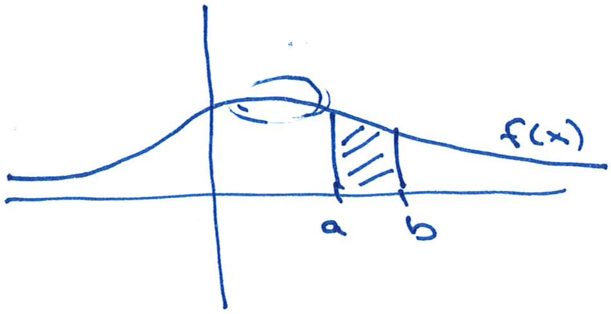
Den har egenskapene:

i) $F'(x) = f(x)$ analysens fundamentalteorem

ii) $F(x)$ er en voksende funksjon med

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{og} \quad \lim_{x\to\infty} F(x) = 1$$





$p(a \le X \le b) = F(b) - F(a)$

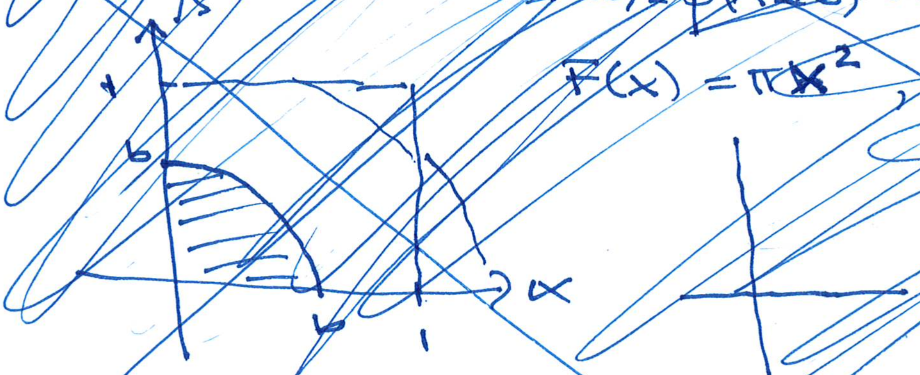
$p(X \le b) = F(b)$

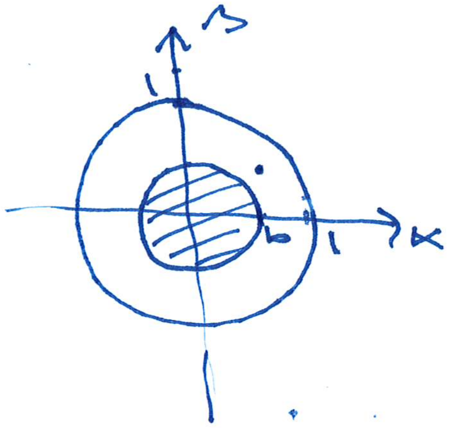
$p(a \le X \le b) = \int_{b}^{a} f(x)\,dx$

~~Eks:~~

$\cancel{F(b) = p(X \le b) = \frac{\pi \cdot b^2}{1 \cdot 1} = \pi b^2}$

$\cancel{F(x) = \pi x^2, \quad 0 \le x \le 1}$





Forsøk: Vi velger et tilfeldig pkt innenfor enhetssirkel, dvs (α, β) s.a. $\alpha^2+\beta^2 \le 1$

X = avstand til origo = $\sqrt{\alpha^2+\beta^2}$

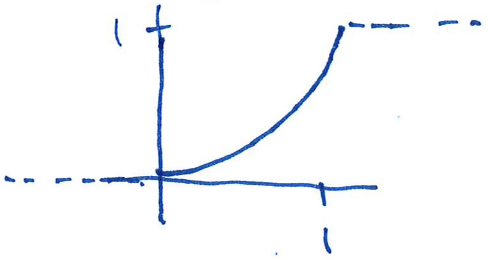
mulige verdier: $[0,1]$

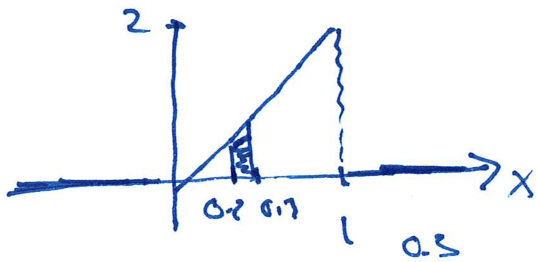
$b \le 1$:

$$F(b) = p(X \le b) = \frac{\cancel{\pi} b^2}{\cancel{\pi} \cdot 1^2} = b^2$$

$F(x) = x^2, 0 \le x \le 1$

$f(x) = 2x, 0 \le x \le 1$





$$F(0.3) - F(0.2) = p(0.2 \le X \le 0.3) = \int_{0.2}^{0.3} 2x\,dx$$

$$F(0.3) - F(0.2) = 0.3^2 - 0.2^2 = \underline{\underline{0.05}}$$

$$= \left[x^2\right]_{0.2}^{0.3} = 0.3^2 - 0.2^2 = \underline{\underline{0.05}}$$

② Forventning og varians: X kont. stok. variabel, $f(x)$ tetthetsfn. til X

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

forventningsverdien til X

Eks: $\int_0^1 x \cdot f(x)\,dx = \int_0^1 x \cdot 2x\,dx = \left[2/3 \cdot x^3\right]_0^1 = 2/3 - 0 = \underline{\underline{2/3}}$

Hvis $h(x)$ er en funksjon av en stok. variabel X, så er forventningsverdien til $h(x)$:

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) \cdot f(x)\,dx$$

Eks: $E[x^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_0^1 x^2 \cdot 2x\,dx = \left[\frac{2}{4} x^4 \right]_0^1$

$$= \tfrac{1}{2} - 0 = \underline{\underline{\tfrac{1}{2}}}$$

Varians: X stokastisk var. (kont.) med $E(X) = \mu$

$$Var(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = E[(X-\mu)^2]$$

Eks: $Var(x) = \int_{-\infty}^{\infty} (x - 2/3)^2 \cdot f(x)\,dx = \int_0^1 (x-2/3)^2 \cdot 2x\,dx$

$$= \int_0^1 2x\left(x^2 - \tfrac{4}{3}x + \tfrac{4}{9}\right)dx = \left[\tfrac{2}{4}x^4 - \tfrac{8}{3}\cdot\tfrac{1}{3}x^3 + \tfrac{8}{9}\tfrac{1}{2}x^2 \right]_0^1$$

$$= \left(\tfrac{1}{2} - 8/9 + 8/18\right) - 0 = \frac{9 - 8\cdot 2 + 8}{18} = \underline{\underline{\frac{1}{18}}}$$

Regneregler:

i) $E(aX+b) = a \cdot E(X) + b$

ii) $Var(X) \geq 0$ og $Var(X) = 0 \iff X = \text{konst.}$

iii) $Var(aX+b) = a^2 Var(X)$

iv) $Var(X) = E(X^2) - E(X)^2$

Om iv): $\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$

$\mu = E(X)$.

$$= E(X^2) - E(2\mu X) + E(\mu^2)$$

$$= E(X^2) - 2\mu \cdot E(X) + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

$$= E(X^2) - E(X)^2$$

Ex: $\mathrm{Var}(X) = E(X^2) - E(X)^2$

$$= \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9}{18} - \frac{8}{18} = 1/18$$

Defn: $\sigma_X = \sqrt{\mathrm{Var}(X)}$ std. avvik til X