

Kursus 02403

Introduktion til Statistik

Forelæsning 3: Kapitel 5, Kontinuerte fordelinger

Jan Kloppenborg Møller

DTU Informatik

Bygning 305 - rum 219

Danmarks Tekniske Universitet

2800 Lyngby – Danmark

e-mail: jkm@imm.dtu.dk

Oversigt

- 1 Kontinuerte Stokastiske variable og fordelinger
 - Tæthedsfunktion
 - Fordelingsfunktion
 - Middelværdi af en kontinuert stokastisk variabel
 - Varians af en kontinuert stokastisk variabel
 - Uniform fordelingen
 - Eksempel 1
- 2 Konkrete Statistiske fordelinger
 - Normalfordelingen
 - Eksempel 2
 - Eksempel 3
 - Eksempel 4
 - Eksempel 5
 - Eksempel 6: Approximation af binomialfordeling
 - Log-Normal fordelingen
 - Eksempel 7
- 3 R (R Note afsnit 4)

Oversigt

- 1 Kontinuerte Stokastiske variable og fordelinger
 - Tæthedsfunktion
 - Fordelingsfunktion
 - Middelværdi af en kontinuert stokastisk variabel
 - Varians af en kontinuert stokastisk variabel
 - Uniform fordelingen
 - Eksempel 1
- 2 Konkrete Statistiske fordelinger
 - Normalfordelingen
 - Eksempel 2
 - Eksempel 3
 - Eksempel 4
 - Eksempel 5
 - Eksempel 6: Approximation af binomialfordeling
 - Log-Normal fordelingen
 - Eksempel 7
- 3 R (R Note afsnit 4)

Tæthedsfunktion

- Tæthedsfunktionen for en stokastisk variabel betegnes ved $f(x)$
- $f(x)$ siger noget om hyppigheden af udfaldet x for den stokastiske variabel X
- For kontinuerte variable svarer tætheden ikke til sandsynligheden, dvs. $f(x) \neq P(X = x)$
- Et godt plot af $f(x)$ er et histogram (kontinuert)

Tæthedsfunktion for en kontinuert variabel

For en kontinuert stokastisk variabel skrives tæthedsfunktionen som:

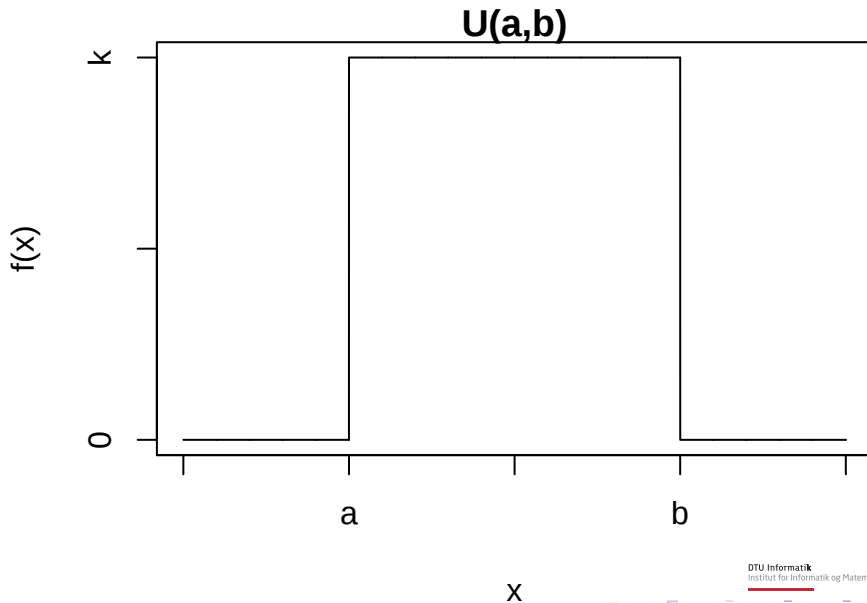
$$f(x)$$

Der gælder:

$$f(x) > 0 \quad \text{for } x \in S$$

$$f(x) = 0 \quad \text{for } x \notin S$$

$$\int_{-\infty}^{\infty} f(x) dx = 1$$



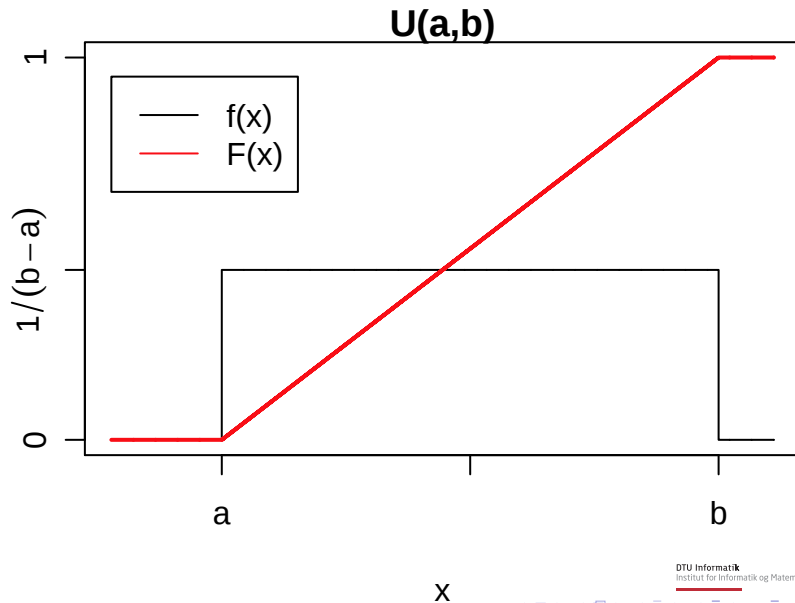
Fordelingsfunktion

- Fordelingsfunktion for en kontinuert stokastisk variabel betegnes ved $F(x)$.
Fordelingsfunktionen svarer til den kumulerede tæthedsfunktion:

$$F(x) = P(X \leq x)$$

$$F(x) = \int_{t=-\infty}^x f(t)dt$$

- Et godt plot for fordelingsfunktionen er den kumulative fordeling



Middelværdi af en kontinuert stokastisk variabel

Middelværdien af en kontinuert stokastisk variabel beregnes ved:

$$\mu = \int_S x \cdot f(x) dx$$

hvor S er udfaldsrummet for X

Varians af en kontinuert stokastisk variabel

Variansen af en kontinuert stokastisk variabel beregnes ved:

$$\sigma^2 = \int_S (x - \mu)^2 \cdot f(x) dx$$

hvor S er udfaldsrummet for X

Uniform fordelingen

$$X \sim U(\alpha, \beta)$$

tæthedsfunktion:

$$f(x) = \frac{1}{\beta - \alpha}$$

Middelværdi:

$$\mu = \frac{\alpha + \beta}{2}$$

Varians:

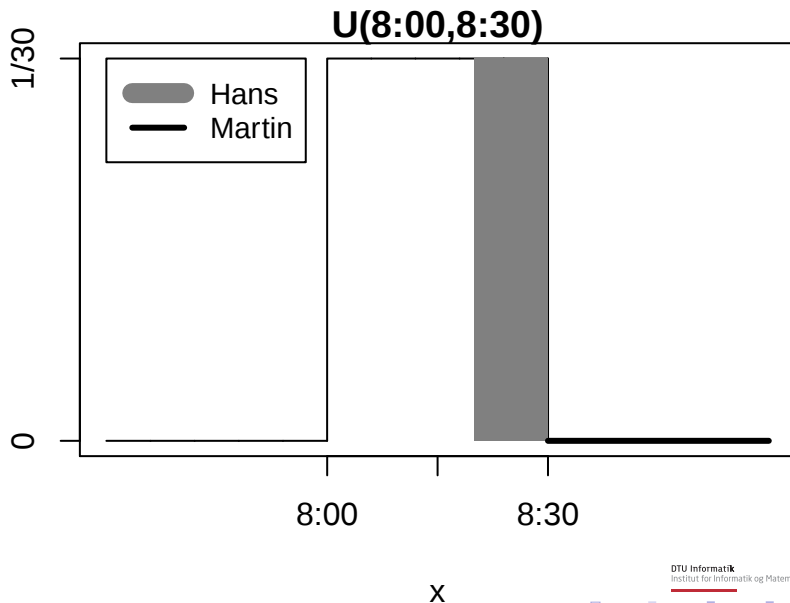
$$\sigma^2 = \frac{1}{12}(\beta - \alpha)^2$$

Eksempel 1

Medarbejdere på en arbejdsplads ankommer mellem klokken 8.00 og 8.30. Det antages, at ankomsttiden kan beskrives ved en uniform fordeling.

Hvad er sandsynligheden for at en tilfældig udvalgt medarbejder (Hans) ankommer mellem 8.20 og 8.30?

Hvad er sandsynligheden for at en tilfældig udvalgt medarbejder (Martin) ankommer efter 8.30?

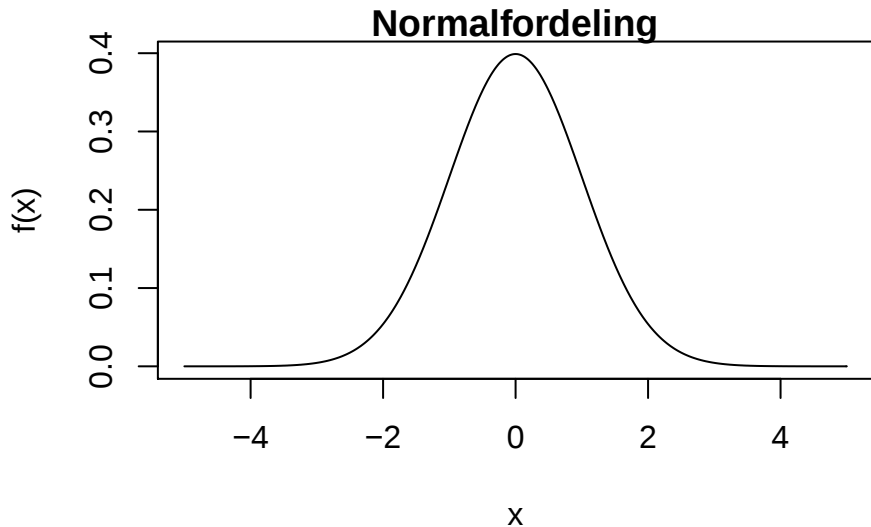


Oversigt

- 1 Kontinuerte Stokastiske variable og fordelinger
 - Tæthedsfunktion
 - Fordelingsfunktion
 - Middelværdi af en kontinuert stokastisk variabel
 - Varians af en kontinuert stokastisk variabel
 - Uniform fordelingen
 - Eksempel 1
- 2 Konkrete Statistiske fordelinger
 - Normalfordelingen
 - Eksempel 2
 - Eksempel 3
 - Eksempel 4
 - Eksempel 5
 - Eksempel 6: Approximation af binomialfordeling
 - Log-Normal fordelingen
 - Eksempel 7
- 3 R (R Note afsnit 4)

Konkrete statistiske fordelinger

- Der findes en række statistiske fordelinger, som kan bruges til at beskrive og analysere forskellige problemstillinger med
- Vi betragter nu kontinuerte fordelinger
- Uniform fordelingen
- Normal fordelingen
- Log-Normal fordelingen



Normal fordelingen

$$X \sim N(\mu, \sigma^2)$$

tæthedsfunktion:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

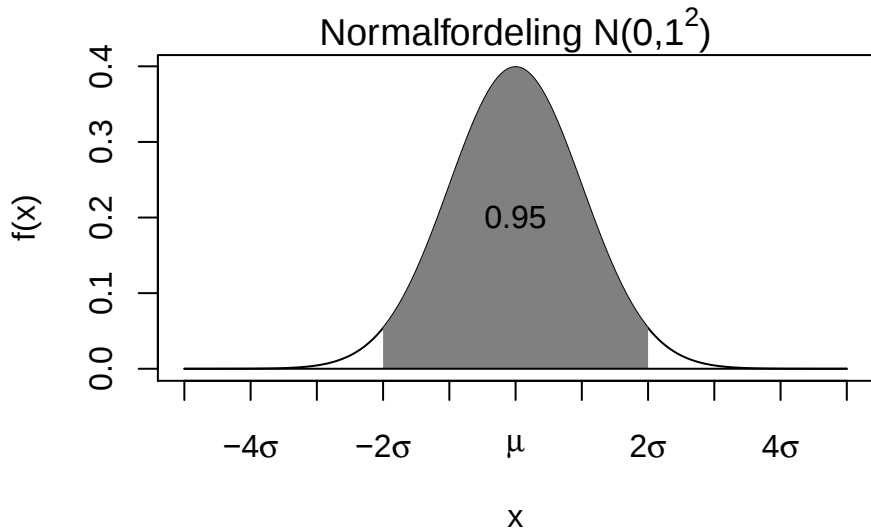
Middelværdi:

$$\mu = \mu$$

Varians:

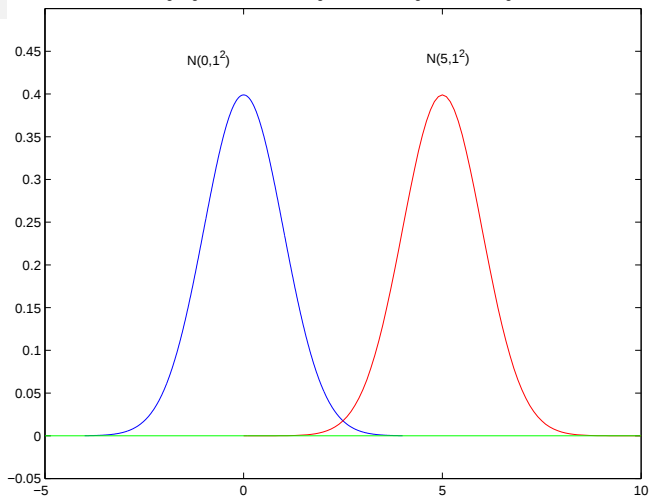
$$\sigma^2 = \sigma^2$$

Tabel 3 for $F(x)$

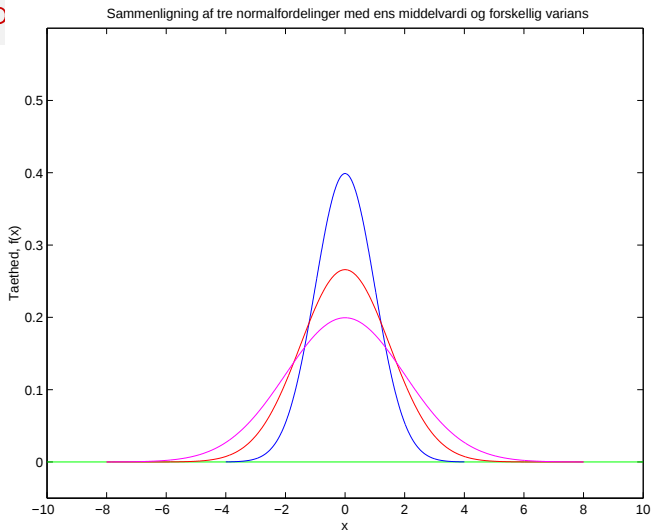


Normalfo

Sammenligning af to normalfordelinger med forskellig middelværdi og ens varians



Normalfordelingen



Normal fordelingen

En normal fordeling med middelværdi 0 og varians 1, dvs

$$X \sim N(0, 1^2)$$

kaldes en standard normal fordeling

En vilkårlig normal fordelt variabel $Y \sim N(\mu, \sigma^2)$ kan standardiseres ved at beregne

$$X = \frac{Y - \mu}{\sigma}$$

Eksempel 2

En vægt har en målefejl, E , der kan beskrives ved en standard normalfordeling, dvs

$$E \sim N(0, 1^2)$$

dvs. middelværdi $\mu = 0$ og spredning $\sigma = 1$ gram.

Vi måler nu vægten af ét emne

a) hvad er sandsynligheden for at vægten måler mindst 2 gram for lidt?

Eksempel 2

b) hvad er sandsynligheden for at vægten måler mindst 2 gram for meget?

c) hvad er sandsynligheden for at vægten måler højst ± 1 gram forkert?

Eksempel 3

Det antages, at blandt en gruppe lærere i folkeskolen, at lønnen kan beskrives ved en normalfordeling med middelværdi $\mu = 280.000$ og spredning $\sigma = 10.000$.

a) hvad er sandsynligheden for at en tilfældig udvalgt lærer tjener mere end 300.000?

Eksempel 4

Det antages, at blandt en gruppe lærere i folkeskolen, at lønnen kan beskrives ved en normalfordeling med middelværdi $\mu = 290.000$ og spredning $\sigma = 4.000$.

a) hvad er sandsynligheden for at en tilfældig udvalgt lærer tjener mere end 300.000?

Eksempel 5

Det antages, at blandt en gruppe lærere i folkeskolen, at lønnen kan beskrives ved en normalfordeling med middelværdi $\mu = 290.000$ og spredning $\sigma = 4.000$.

a) angiv det interval, der dækker over 95% af læreres løn

Eksempel 6: Approximation af binomialfordeling

I et dosis-respons forsøg med 80 rotter antages at sandsynligheden for at en rotte overlever forsøget er $p = 0.5$.

a) hvad er sandsynligheden for at højst 30 rotter dør i forsøget?

b) hvad er sandsynligheden for at mellem 38 og 42 rotter dør i forsøget?

Log-Normal fordelingen

$$X \sim LN(\alpha, \beta^2)$$

tæthedsfunktion:

$$f(x) = \begin{cases} \frac{1}{\beta\sqrt{2\pi}} x^{-1} e^{-(\ln(x)-\alpha)^2/2\beta^2} & x > 0, \beta > 0 \\ 0 & \text{ellers} \end{cases}$$

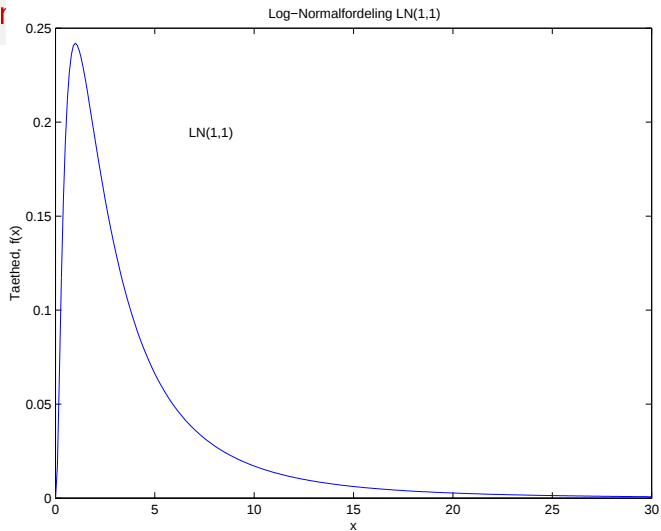
Middelværdi:

$$\mu = e^{\alpha+\beta^2/2}$$

Varians:

$$\sigma^2 = e^{2\alpha+\beta^2}(e^{\beta^2} - 1)$$

Log-Normal



Log-Normal fordelingen

En log-normal fordelt variabel $X \sim LN(\alpha, \beta^2)$, kan transformeres til en standard normal fordelt variabel Y ved

$$Y = \frac{\ln(X) - \alpha}{\beta}$$

dvs.

$$Y \sim N(0, 1^2)$$

Eksempel 7

Partikelstørrelsen (μm) i et stof kan antages at være Log-Normal fordelt. Vi har observationerne (y_i)

2.2 3.4 1.6 0.8 2.7 3.3 1.6 2.8 1.9

Vi tager logaritmen af data og får (x_i):

0.8 1.2 0.5 -0.2 1.0 1.2 0.5 1.0 0.6

Heraf beregnes $\bar{x} = 0.732$ og $s = 0.46$.

Hvad er den forventede andel af partikler med en størrelse i intervallet $[2; 3]$

Oversigt

- 1 Kontinuerte Stokastiske variable og fordelinger
 - Tæthedsfunktion
 - Fordelingsfunktion
 - Middelværdi af en kontinuert stokastisk variabel
 - Varians af en kontinuert stokastisk variabel
 - Uniform fordelingen
 - Eksempel 1
- 2 Konkrete Statistiske fordelinger
 - Normalfordelingen
 - Eksempel 2
 - Eksempel 3
 - Eksempel 4
 - Eksempel 5
 - Eksempel 6: Approximation af binomialfordeling
 - Log-Normal fordelingen
 - Eksempel 7
- 3 R (R Note afsnit 4)

R (R note afsnit 4)

R	Betegnelse
norm	Normalfordelingen
unif	Den uniforme fordeling
lnorm	Log-normalfordelingen
exp	Exponentialfordelingen

- d Tæthedsfunktion $f(x)$ (probability density function).
- p Fordelingsfunktion $F(x)$ (cumulative distribution function).
- q Fraktil (quantile) i fordeling.

Eksempel:

$$P(Z \leq 2) \qquad \text{pnorm}(2)$$

Oversigt

- 1 Kontinuerte Stokastiske variable og fordelinger
 - Tæthedsfunktion
 - Fordelingsfunktion
 - Middelværdi af en kontinuert stokastisk variabel
 - Varians af en kontinuert stokastisk variabel
 - Uniform fordelingen
 - Eksempel 1
- 2 Konkrete Statistiske fordelinger
 - Normalfordelingen
 - Eksempel 2
 - Eksempel 3
 - Eksempel 4
 - Eksempel 5
 - Eksempel 6: Approximation af binomialfordeling
 - Log-Normal fordelingen
 - Eksempel 7
- 3 R (R Note afsnit 4)