Kursus 02403
Introduktion til Statistik

Forelæsning 4: Kapitel 5: Kontinuerte fordelinger

Jan Klopenborg Møller

DTU Informatik
Bygning 305 - rum 210
Danmarks Tekniske Universitet
2800 Lyngby – Danmark

e-mail: jkm@imm.dtu.dk

DTU Informatik
Institut for Informatik og Matematisk Modellering

Oversigt

Oversigt

DTU Informatik
Institut for Informatik og Matematisk Modellering

Nogle anvendte fordelinger

(Både kontinuerte og diskerete)

- Diskrete:
  - Binomial fordelingen
  - Den hypergeometriske fordeling
  - Poisson fordelingen
- Kontinuerte:
  - Normal fordelingen
  - Log-Normal fordelingen
  - Uniform fordelingen
  - **Eksponential fordelingen**

Oversigt

Eksponential fordelingen

$X \sim Exp(\beta)$

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

Middelværdi:

$\mu = \beta$

Varians:

$\sigma^2 = \beta^2$

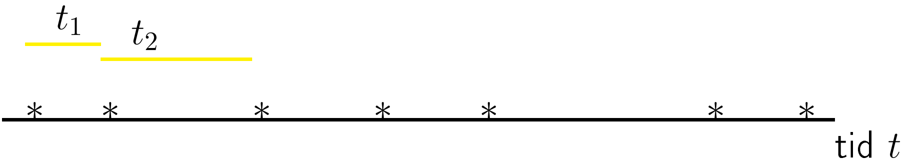
Eksponentialfordelingen

- Eksponential fordelingen er et special tilfælde af Gamma fordelingen
- Eksponential fordelingen anvendes f.eks. til at beskrive levetider og ventetider
- Eksponential fordelingen kan bruges til at beskrive (vente)tiden mellem hændelser i poisson fordelingen
- Middelværdi $\mu = \beta$
- Varians $\sigma^2 = \beta^2$

Sammenhæng mellem Eksponential og Poisson fordelingen

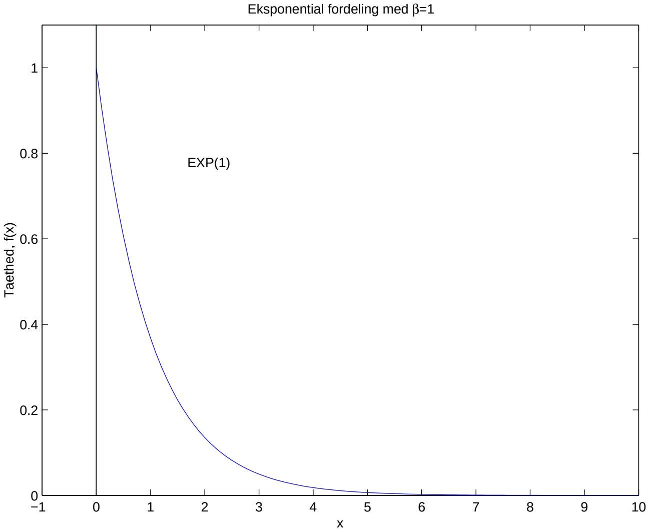
Poisson: Diskrete hændelser pr. enhed

Eksponential: Kontinuert afstand mellem hændelser



DTU Informatik
Institut for Informatik og Matematisk Modellering

Eksponential fordelingen



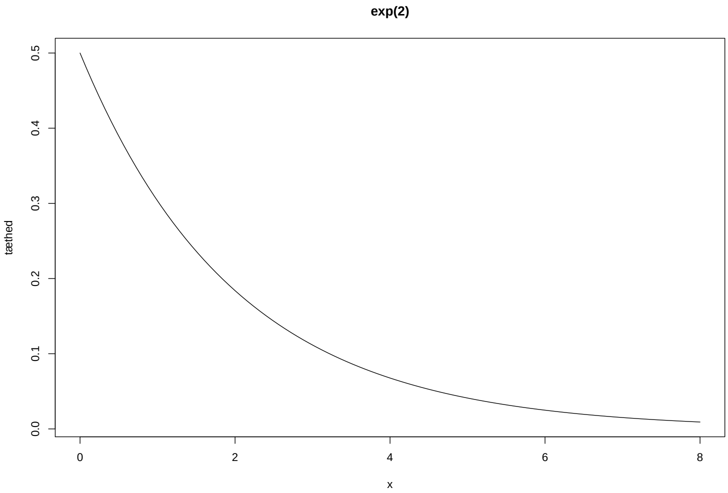
Eksempel 1

Tiden mellem kundeankomster på et posthus er eksponential fordelt med middelværdi $\mu = 2$ minutter.

En kunde er netop ankommet. Hvad er sandsynligheden for at der ikke kommer flere kunder indenfor en periode på 2 minutter?

DTU Informatik
Institut for Informatik og Matematisk Modellering

Eksempel 1



Eksempel 2

Tiden mellem kundeankomster på et posthus er eksponential fordelt med middelværdi $\mu = 2$ minutter.

En kunde er netop ankommet. Beregn sandsynligheden for at der ikke kommer flere kunder indefor en periode på 2 minutter vha. Poissonfordelingen

Eksempel 3

Tiden mellem kundeankomster på et posthus er eksponential fordelt med middelværdi $\mu = 2$ minutter. Vi betragter nu en periode på 10 minutter

Beregn sandynligheden for at der ikke kommer nogen kunder i perioden vha. Poissonfordelingen

Beregn sandynligheden for at der ikke kommer nogen kunder i perioden vha. Binomialfordelingen

Oversigt

Regneregler for stokastiske variable

(Gælder BÅDE kontinuert og diskret)
Vi antager at a og b er konstanter og X er en stokastisk variabel. Da gælder

$$E(aX + b) = aE(X) + b$$

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

Eksempel 4

En stokastisk variabel X har middelværdi 4 og varians 6.

Beregn middelværdi og varians for $Y = -3X + 2$

DTU Informatik
Institut for Informatik og Matematisk Modellering

Regneregler for stokastiske variable

Følgende linear kombinationer gælder:

$$E(a_1X_1 + a_2X_2 + .. + a_nX_n)$$
$$= a_1E(X_1) + a_2E(X_2) + .. + a_nE(X_n)$$

Hvis X_i og X_j $(i \neq j)$ er uafhængige

$$Var(a_1X_1 + a_2X_2 + .. + a_nX_n)$$
$$= a_1^2Var(X_1) + .. + a_n^2Var(X_n)$$

Eksempel 5

Vægten af passagerer på en flystrækning antages normalfordelt $X \sim N(70, 10^2)$.

Et fly, der kan tage 55 passagerer, må max. lastes med 4000 kg (kun passageres vægt betragtes som last).

Beregn sandsynligheden for at flyet bliver overlastet

DTU Informatik
Institut for Informatik og Matematisk Modellering

Eksempel 6

En færge medtager X passagerer. Det oplyses, at $E[X] = 400$ og $Var[X] = 40^2$ og at X er normalfordelt.

Færgeselskabet har en fortjeneste på 120 kr. pr. passager

Beregn middelværdi og varians for færgeselskabets fortjeneste, og beregn sandsynligheden for at fortjenesten vil overstige 50.000 kr.

Eksempel 7

Antag, at du vil købe noget sand, da du vil lave en ny indkørsel til dit hus. Det er specielt vigtigt for dig, at der ikke er sten blandet i sandet. Producenten lover, at der i gennemsnit kun er 0.5 sten pr. m^3.

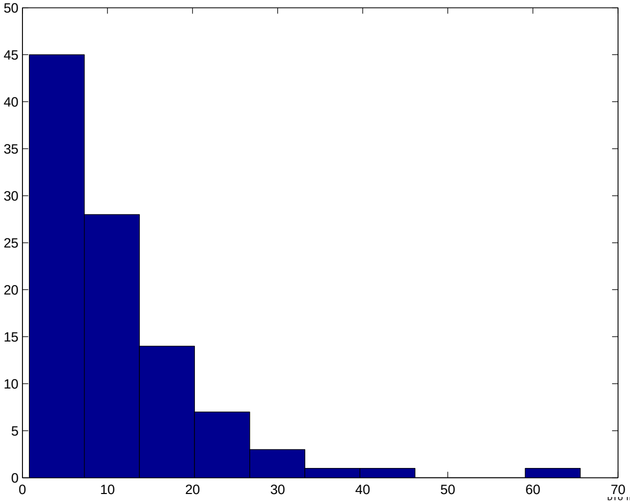
Du køber nu 10 m^3 *sand og erfarer, at der er 11 sten i blandingen. Føler du dig snydt af producentens løfte om at der kun er 0.5 sten pr.* m^3 *?*

Oversigt

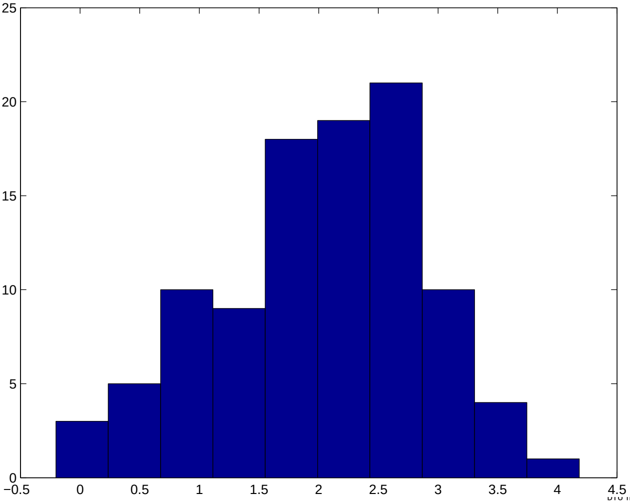
Transformationer

Såfremt data afviger fra at være normal fordelt, kan man ofte med fordel transformere data, således at de transformerede data kan antages at være normal fordelt.

Transformationer - Før



Transformationer - Efter



Hyppigt anvendte transformationer

- Gør store værdier mindre:
  - $-\frac{1}{x}$
  - $\ln x$
  - $x^{1/4}$
  - $\sqrt{x}$
- Gør store værdier større:
  - x^2
  - x^3

DTU Informatik
Institut for Informatik og Matematisk Modellering

Transformationer

Ofte vil man udføre statistiske analyser på de transformerede data, såfremt det viser sig, at de transformerede data er 'pænere', f.eks. mere symmetriske.

Oversigt

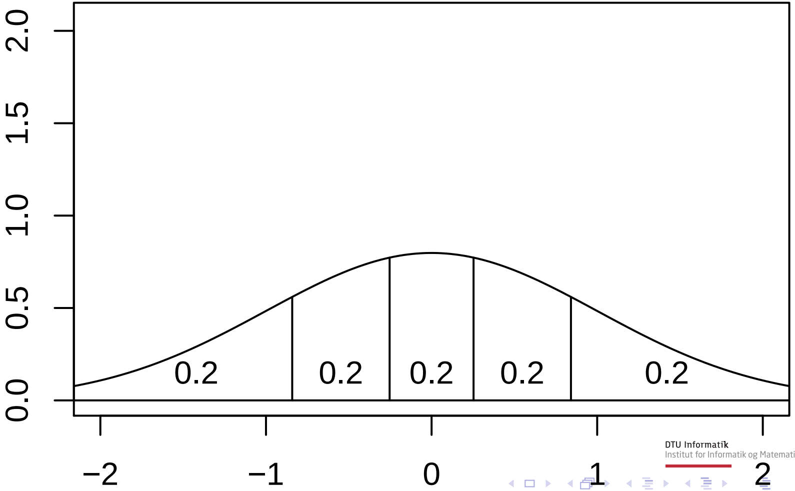
DTU Informatik
Institut for Informatik og Matematisk Modellering

Normalplot

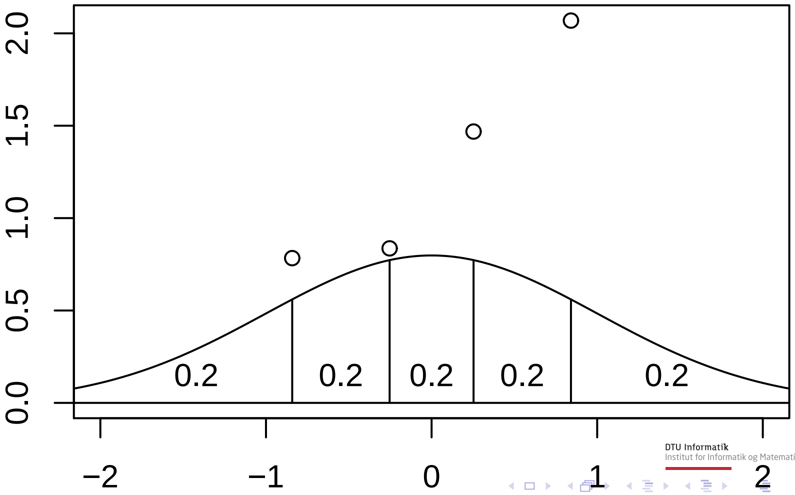
Man kan kontrollere/undersøg om data ser ud til at være normalfordelt:

- Plotte histogram
- Lave box plots
- Lave et normalplot (R: `qqnorm(x)`):
  - En direkte sammenligning med normalfordelingen

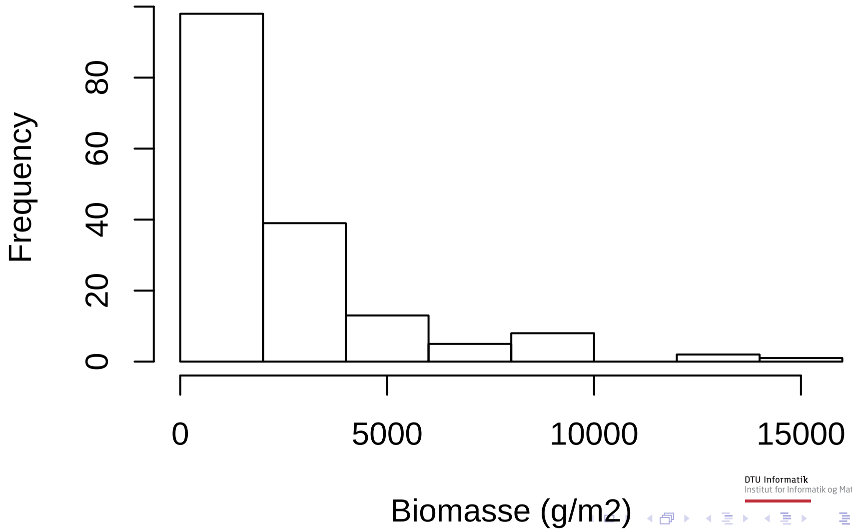
Normalplot



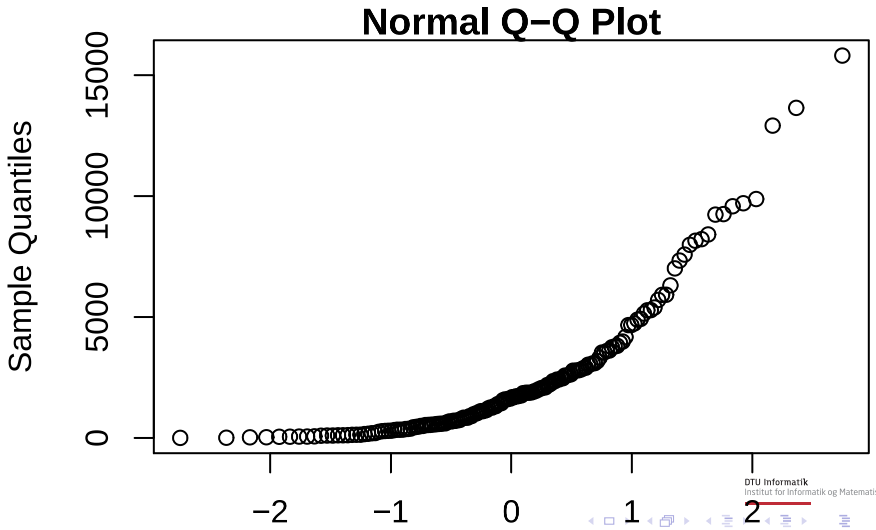
Normalplot



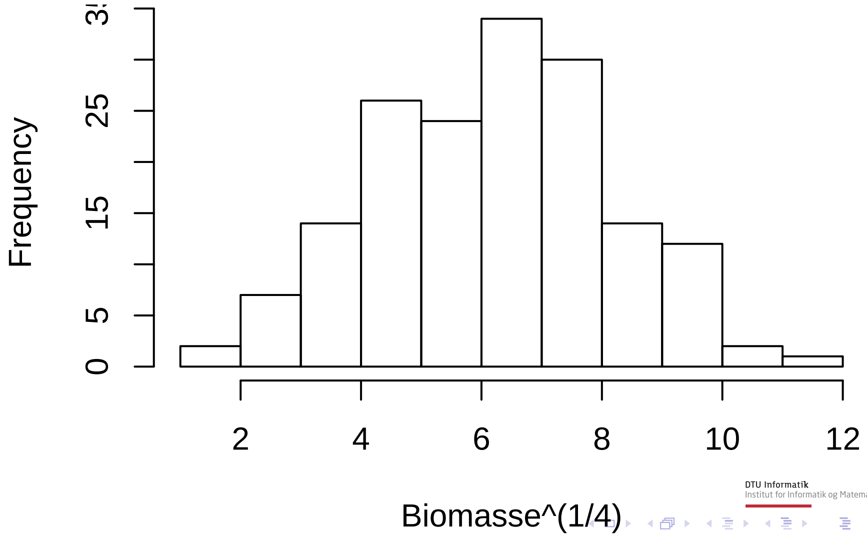
Eksempel muslinger i Skive fjord



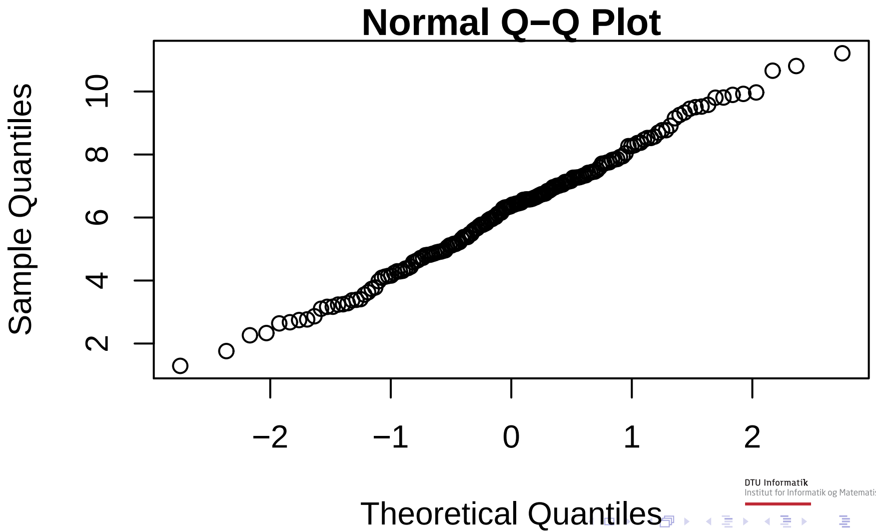
Eksempel muslinger i Skive fjord



Eksempel muslinger i Skive fjord



Eksempel muslinger i Skive fjord



DTU Informatik
Institut for Informatik og Matematisk Modellering

Eksempel muslinger i Skive fjord

Vi finder

$$X^{1/4} \sim N(\mu, \sigma^2)$$

med $\mu = 6.2$ og $\sigma = 2$

Oversigt

DTU Informatik
Institut for Informatik og Matematisk Modellering

R (R note 5)

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

- `d` Tæthedsfunktion $f(x)$ (probability distribution).
- `p` Fordelingsfunktion $F(x)$ (cumulative distribution function).
- `r` Tilfældige tal fra fordelingen.
- `q` Fraktil (quantile) i fordeling.

Eksempel: $P(X \leq 2), X \sim \text{Exp}(3)$: pexp(2,1/3)

Oversigt