02402 Statistik (Polyteknisk grundlag)

Uge 3: Stokastiske variable og kontinuerte fordelinger

Nicolai Siim Larsen
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable

Opsummering: Uge 1 og 2

Vi ønsker at undersøge en population.

Populationen kan beskrives ved bl.a. positionsmål og spredningsmål, f.eks. gennemsnittet og variansen.

Man kan opstille et eksperiment og udtage en stikprøve for at estimere populationens parametre.

Men hvis vi vil regne på eksperimentet før det udføres, har vi brug for en matematisk model. Her kan vi definere en stokastisk variabel og tilknytte den en fordeling, der beskriver, hvorledes sandsynlighedsmassen fordeler sig over de reelle tal.

Opsummering: Uge 1 og 2

En sandsynlighedsfordeling er en teoretisk model, der beskriver, hvor sandsynligt det er, at den stokastiske variable antager forskellige værdier.

Vi skelner mellem diskrete og kontinuerte stokastiske variable (og fordelinger) baseret på tælleligheden af variablens værdimængde.

Man kan beskrive en sandsynlighedsfordeling igennem dens fordelingsfunktion eller dens tæthedsfunktion. Fordelingensfunktionen er defineret ens for både diskrete og kontinuerte fordelinger, hvilket ikke er tilfældet for tæthedsfunktionen.

I den teoretiske model repræsenterer middelværdien af den stokastiske variabel $\mathbb{E}[X]$ det sande populationsgennemsnit og variansen af den stokastiske variabel $\mathbb{V}[X]$ den sande populationsvarians.

Opsummering: Uge 1 og 2

For en diskret stokastisk variabel X med værdimængde A, defineres tæthedsfunktionen som

$$f(x) = \mathbb{P}(X = x)$$

og fordelingsfunktionen som

$$F(x) = \mathbb{P}(X \leq x) = \sum_{\{y \in A : y \leq x\}} \mathbb{P}(X = y) = \sum_{\{y \in A : y \leq x\}} f(y).$$

På engelsk kaldes tæthedsfunktionen ofte for *the probability mass function (pmf)* i det diskrete tilfælde.

Forventningsværdien (middelværdien) af X beregnes som

$$\mathbb{E}[X] = \sum_{x \in A} x \mathbb{P}(X = x) = \sum_{x \in A} x f(x).$$

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable

Tæthedsfunktionen, Definition 2.32

- Tæthedsfunktionen (density function/probability density function, pdf) for en stokastisk variabel betegnes med $f(x)$.

Tæthedsfunktionen, Definition 2.32

- Tæthedsfunktionen (density function/probability density function, pdf) for en stokastisk variabel betegnes med $f(x)$.
- Tæthedsfunktionen $f(x)$ siger noget om hyppigheden af udfaldsværdien x for den stokastiske variabel X.

Tæthedsfunktionen, Definition 2.32

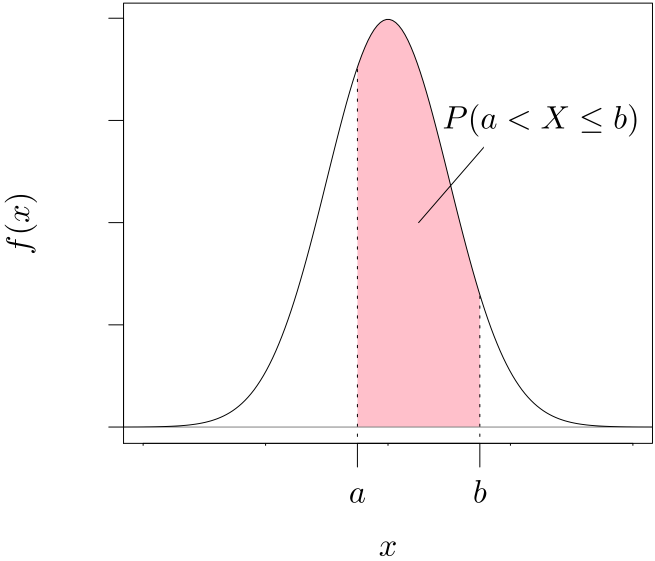
- Tæthedsfunktionen (density function/probability density function, pdf) for en stokastisk variabel betegnes med $f(x)$.
- Tæthedsfunktionen $f(x)$ siger noget om hyppigheden af udfaldsværdien x for den stokastiske variabel X.
- Tæthedsfunktionen for en kontinuert stokastisk variabel svarer *ikke* til sandsynligheden for at observere værdien x. Der gælder faktisk, $P(X = x) = 0$ for alle x.

Tæthedsfunktionen, Definition 2.32

- Tæthedsfunktionen (density function/probability density function, pdf) for en stokastisk variabel betegnes med $f(x)$.
- Tæthedsfunktionen $f(x)$ siger noget om hyppigheden af udfaldsværdien x for den stokastiske variabel X.
- Tæthedsfunktionen for en kontinuert stokastisk variabel svarer *ikke* til sandsynligheden for at observere værdien x. Der gælder faktisk, $P(X = x) = 0$ for alle x.
- Tæthedsfunktionen $f(x)$ hørende til fordelingen af en kontinuert stokastisk variabel opfylder at:

$$f(x) \geq 0 \text{ for alle } x \quad \text{og} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1\,.$$

Tæthedsfunktionen



Fordelingsfunktionen, Definition 2.33

- Fordelingsfunktionen (distribution function/cumulative distribution function, cdf) hørende til en kontinuert stokastisk variabel benævnes med $F(x)$.

Fordelingsfunktionen, Definition 2.33

- Fordelingsfunktionen (distribution function/cumulative distribution function, cdf) hørende til en kontinuert stokastisk variabel benævnes med $F(x)$.
- Fordelingsfunktionen er defineret ved

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

Fordelingsfunktionen, Definition 2.33

- Fordelingsfunktionen (distribution function/cumulative distribution function, cdf) hørende til en kontinuert stokastisk variabel benævnes med $F(x)$.
- Fordelingsfunktionen er defineret ved

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

- Som følge af denne definition gælder, at

$$f(x) = F^{'}(x)\,,$$

hvor fordelingsfunktionen er differentiabel.

Fordelingsfunktionen, Definition 2.33

- Fordelingsfunktionen (distribution function/cumulative distribution function, cdf) hørende til en kontinuert stokastisk variabel benævnes med $F(x)$.
- Fordelingsfunktionen er defineret ved

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

- Som følge af denne definition gælder, at

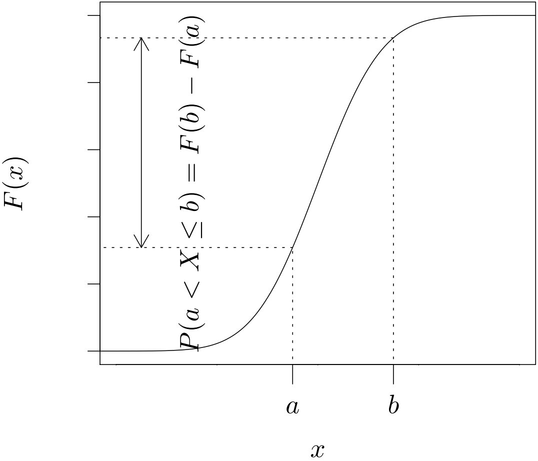
$$f(x) = F^{'}(x)\,,$$

hvor fordelingsfunktionen er differentiabel.

- Bemærk også, at:

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)\,dx\,.$$

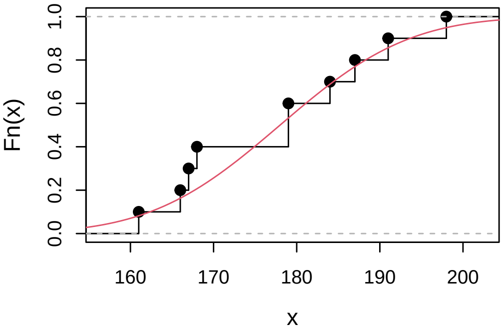
Fordelingsfunktionen



Empirisk fordelingsfunktion (ecdf)

```r
# Empirical cdf for sample of height data from Chapter 1
x <- c(168, 161, 167, 179, 184, 166, 198, 187, 191, 179)
plot(ecdf(x), verticals = TRUE, main = "")

# 'True cdf' for normal distribution (with sample mean and variance)
xp <- seq(0.9*min(x), 1.1*max(x), length = 100)
lines(xp, pnorm(xp, mean(x), sd(x)), col = 2)
```



Middelværdi (mean) af en kontinuert stokastisk variabel , Definition 2.34

Middelværdien af en kontinuert stokastisk variabel:

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

Middelværdi (mean) af en kontinuert stokastisk variabel , Definition 2.34

Middelværdien af en kontinuert stokastisk variabel:

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

Sammenlign med definitionen for en diskret stokastisk variabel:

$$\mu = \sum_{\text{alle } x} x f(x)$$

Varians af en kontinuert stokastisk variabel, Definition 2.34

Variansen af en kontinuert stokastisk variabel:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Varians af en kontinuert stokastisk variabel, Definition 2.34

Variansen af en kontinuert stokastisk variabel:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Sammenlign med definitionen for en diskret stokastisk variabel:

$$\sigma^2 = \sum_{\text{alle} x} (x-\mu)^2 f(x)$$

Kovarians, Definition 2.58

Kovariansen af to stokastisk variable:

Lad X og Y være to stokastiske variable. Kovariansen mellem X og Y er defineret som

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Kovariansen:

Hvis to stokastiske variable er *uafhængige*, så er kovariansen 0. *Det modsatte er ikke nødvendigvis tilfældet!*

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable

Vigtige kontinuerte fordelinger

Der findes en række statistiske fordelinger (både kontinuerte og diskrete), som kan bruges til at beskrive og analysere forskellige problemstillinger med

I dag ser vi nærmere på følgende kontinuerte fordelinger:

- Den uniforme fordeling
- Normalfordelingen
- Log-normalfordelingen
- Eksponentialfordelingen

Kontinuerte fordelinger in R

R	Fordeling
`norm`	Normalfordelingen
`unif`	Uniform fordeling
`lnorm`	Log-normalfordelingen
`exp`	Eksponentialfordelingen

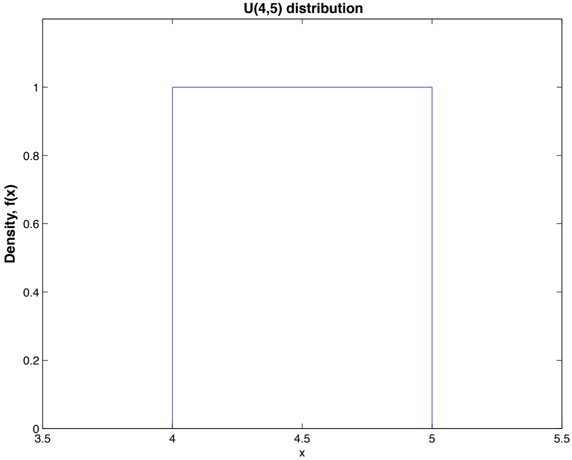
`d` Tæthedsfunktion (density)

`p` Fordelingsfunktion (probability)

`q` Fraktilfunktion (quantile)

`r` Tilfældighedsgenerator (random).

Tæthed for en uniform fordeling (eksempel)



Den uniforme fordeling, Def. 2.35 & sæt. 2.36

Notation:

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

Den uniforme fordeling, Def. 2.35 & sæt. 2.36

Notation:

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

Tæthedsfunktion:

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$ og ellers nul

Den uniforme fordeling, Def. 2.35 & sæt. 2.36

Notation:

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

Tæthedsfunktion:

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$ og ellers nul

Middelværdi:

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

Den uniforme fordeling, Def. 2.35 & sæt. 2.36

Notation:

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

Tæthedsfunktion:

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$ og ellers nul

Middelværdi:

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

Varians:

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

Eksempel 1

Studerende på et statistikkursus ankommer til en forelæsning mellem 8.00 og 8.30. Det antages, at ankomsttiden kan beskrives ved en uniform fordeling.

Spørgsmål:

Hvad er sandsynligheden for, at en tilfældigt udvalgt studerende ankommer mellem 8:20 og 8:30?

Eksempel 1

Studerende på et statistikkursus ankommer til en forelæsning mellem 8.00 og 8.30. Det antages, at ankomsttiden kan beskrives ved en uniform fordeling.

Spørgsmål:

Hvad er sandsynligheden for, at en tilfældigt udvalgt studerende ankommer mellem 8:20 og 8:30?

Svar:

$10/30 = 1/3$

Lad $X \sim U(0, 30)$ repræsentere "ankomsttiden" for en tilfældigt udvalgt studerende:
$P(20 \leq X \leq 30) = P(X \leq 30) - P(X \leq 20) = 1 - 2/3 = 1/3$

```
punif(30, 0, 30) - punif(20, 0, 30)
```

```
[1] 0.3333
```

Eksempel 1 (fortsat)

Spørgsmål:

Hvad er sandsynligheden for, at en tilfældigt udvalgt studerende ankommer efter 8:30?

Eksempel 1 (fortsat)

Spørgsmål:

Hvad er sandsynligheden for, at en tilfældigt udvalgt studerende ankommer efter 8:30?

Svar:

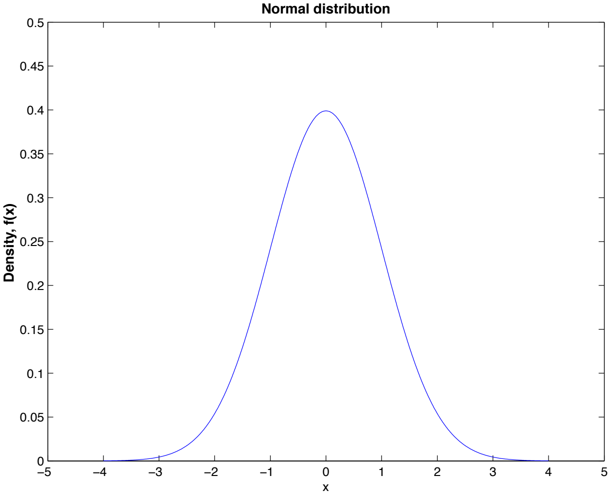
0

Lad $X \sim U(0,30)$ repræsentere "ankomsttiden" for en tilfældigt udvalgt studerende:
$P(X > 30) = 1 - P(X \leq 30) = 1 - 1 = 0$

```
1 - punif(30, 0, 30)
```

```
[1] 0
```

Tætheden for en normalfordeling (eksempel)



Normalfordelingen, Def. 2.37 & sæt. 2.38

Notation:

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

Normalfordelingen, Def. 2.37 & sæt. 2.38

Notation:

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

Tæthedsfunktion:

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

Normalfordelingen, Def. 2.37 & sæt. 2.38

Notation:

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

Tæthedsfunktion:

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

Middelværdi:

μ

Normalfordelingen, Def. 2.37 & sæt. 2.38

Notation:

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

Tæthedsfunktion:

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

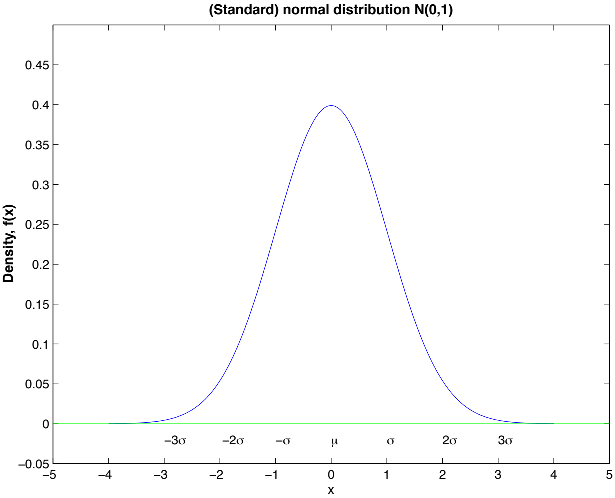
Middelværdi:

μ

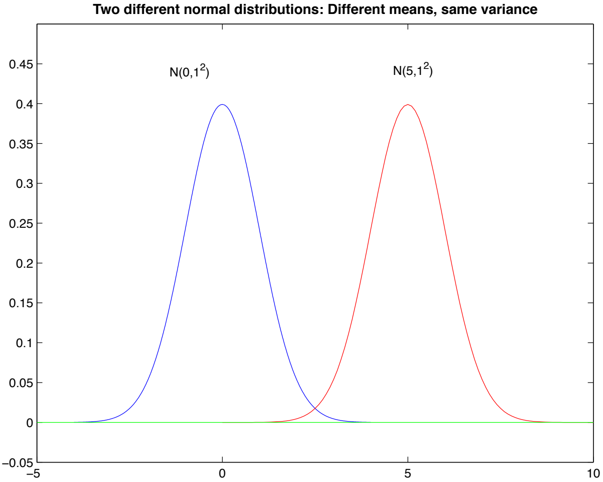
Varians:

σ^2

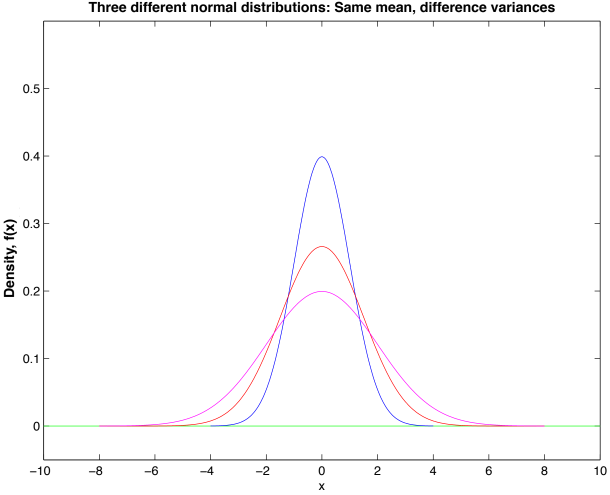
Tætheden for en standardnormalfordeling



Tætheder for to normalfordelinger (eksempel)



Tætheder for tre normalfordelinger (eksempel)



Standardnormalfordelingen

Standardnormalfordelingen:

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

Normalfordelingen med middelværdi 0 og varians 1.

Standardnormalfordelingen

Standardnormalfordelingen:

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

Normalfordelingen med middelværdi 0 og varians 1.

Standardisering:

En vilkårlig normalfordelt variabel $X \sim N(\mu, \sigma^2)$ kan *standardiseres* ved

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

Eksempel 2

Målefejl:

En given vægt har en målefejl (målt i gram), Z, som kan beskrives med en standardnormalfordeling,

$$Z \sim N(0, 1^2).$$

Dvs. at den gennemsnitlige målefejl er $\mu = 0$ gram og standardafgivelsen er $\sigma = 1$ gram. Antag, at vægten bruges til at veje et produkt.

Eksempel 2

Målefejl:

En given vægt har en målefejl (målt i gram), Z, som kan beskrives med en standardnormalfordeling,

$$Z \sim N(0, 1^2) .$$

Dvs. at den gennemsnitlige målefejl er $\mu = 0$ gram og standardafgivelsen er $\sigma = 1$ gram. Antag, at vægten bruges til at veje et produkt.

Spørgsmål a):

Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram mindre end den sande vægt af produktet?

Eksempel 2

Målefejl:

En given vægt har en målefejl (målt i gram), Z, som kan beskrives med en standardnormalfordeling,

$$Z \sim N(0, 1^2).$$

Dvs. at den gennemsnitlige målefejl er $\mu = 0$ gram og standardafgivelsen er $\sigma = 1$ gram. Antag, at vægten bruges til at veje et produkt.

Spørgsmål a):

Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram mindre end den sande vægt af produktet?

Svar:

$P(Z \leq -2) = 0.02275$

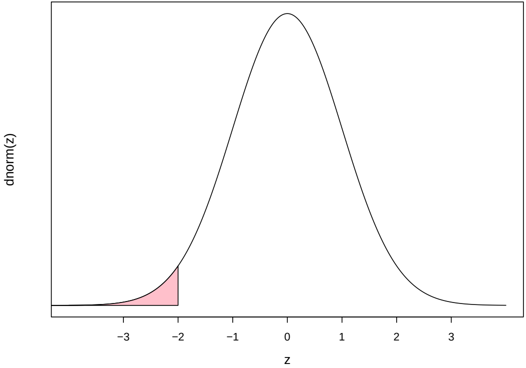
```
pnorm(-2)
```

Eksempel 2

Svar:

```
pnorm(-2)
```

[1] 0.02275



Eksempel 2

Spørgsmål b):

Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram større end den sande vægt af produktet?

Eksempel 2

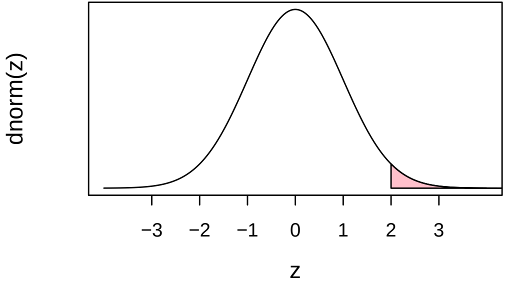
Spørgsmål b):

Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram større end den sande vægt af produktet?

Svar:

$P(Z \geq 2) = 0.02275$

```
1 - pnorm(2)
```



Eksempel 2

Spørgsmål c):

Hvad er sandsynligheden for, at vægten har en afvigelse på højst ± 1 *gram?*

Eksempel 2

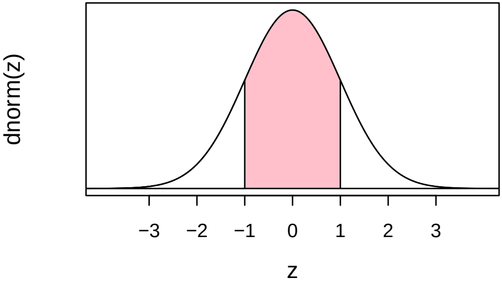
Spørgsmål c):

Hvad er sandsynligheden for, at vægten har en afvigelse på højst ± 1 *gram?*

Svar:

$$P(|Z| \leq 1) = P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.683$$

```
pnorm(1) - pnorm(-1)
```



Eksempel 3

Indkomstfordeling:

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

Eksempel 3

Indkomstfordeling:

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

Spørgsmål a):

Hvad er sandsynligheden for, at en tilfældigt udvalgt lærer tjener mere end 300.000 kr.?

Eksempel 3

Spørgsmål a):

Hvad er sandsynligheden for, at en tilfældigt udvalgt lærer tjener mere end 300.000 kr?

Eksempel 3

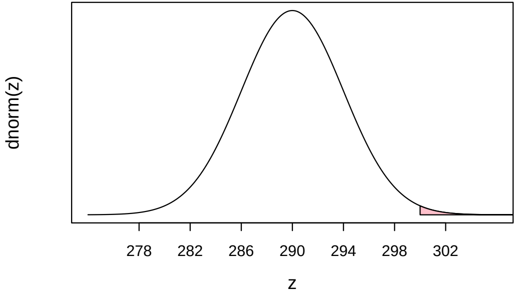
Spørgsmål a):

Hvad er sandsynligheden for, at en tilfældigt udvalgt lærer tjener mere end 300.000 kr?

Svar:

```
1 - pnorm(300, m = 290, s = 4)
```

[1] 0.00621



Eksempel 4

(Samme indkomstfordeling):

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

Eksempel 4

(Samme indkomstfordeling):

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

"Modsat" spørgsmål:

Specificér et løninterval (som er symmetrisk omkring middelværdien), som dækker 95% af lærernes lønninger.

Eksempel 4

(Samme indkomstfordeling):

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

"Modsat" spørgsmål:

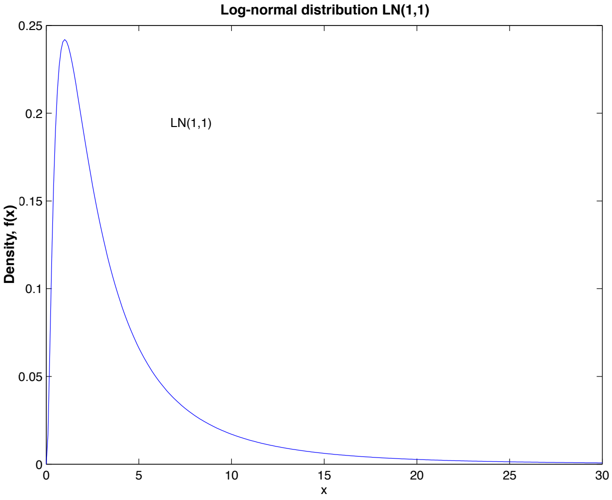
Specificér et løninterval (som er symmetrisk omkring middelværdien), som dækker 95% af lærernes lønninger.

Svar:

```r
qnorm(c(0.025, 0.975), m = 290, s = 4)
```

[1] 282.2 297.8

Log-normalfordelingen



Log-normalfordelingen, Def. 2.46 & sæt. 2.47

Notation:

$X \sim LN(\alpha, \beta^2)$ (hvor $\beta > 0$)

Log-normalfordelingen, Def. 2.46 & sæt. 2.47

Notation:

$X \sim LN(\alpha, \beta^2)$ (hvor $\beta > 0$)

Tæthedsfunktion:

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

Log-normalfordelingen, Def. 2.46 & sæt. 2.47

Notation:

$X \sim LN(\alpha, \beta^2)$ (hvor $\beta > 0$)

Tæthedsfunktion:

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

Middelværdi:

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

Log-normalfordelingen, Def. 2.46 & sæt. 2.47

Notation:

$X \sim LN(\alpha, \beta^2)$ (hvor $\beta > 0$)

Tæthedsfunktion:

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

Middelværdi:

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

Varians:

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

Log-normalfordelingen

Log-normal- og normalfordelingen:

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

$$X = \ln(Y).$$

Her er X normalfordelt med middelværdi α og varians β^2, dvs. $X \sim N(\alpha, \beta^2)$.

Log-normalfordelingen

Log-normal- og normalfordelingen:

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

$$X = \ln(Y).$$

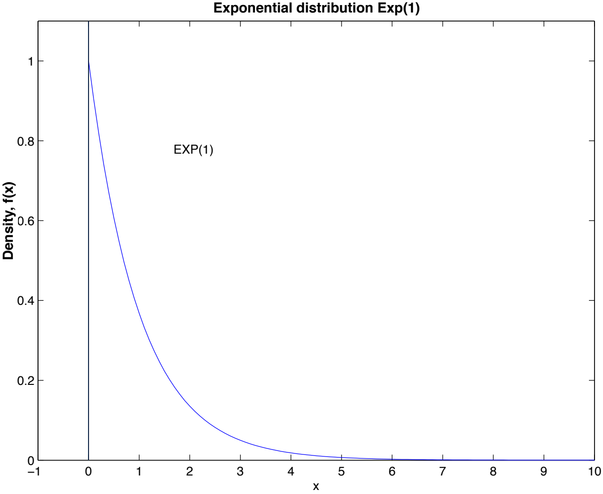
Her er X normalfordelt med middelværdi α og varians β^2, dvs. $X \sim N(\alpha, \beta^2)$.

Ved at standardisere X igennem

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}} = \frac{\ln(Y) - \alpha}{\beta}$$

opnås en standardnormalfordelt variabel $Z \sim N(0, 1)$.

Eksponentialfordelingen



Eksponentialfordelingen, Def. 2.48 & sæt. 2.49

Notation:

$X \sim \text{Exp}(\lambda)$, hvor $\lambda > 0$.

Tæthedsfunktion:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

Middelværdi:

$\mu = \frac{1}{\lambda}$

Varians:

$\sigma^2 = \frac{1}{\lambda^2}$

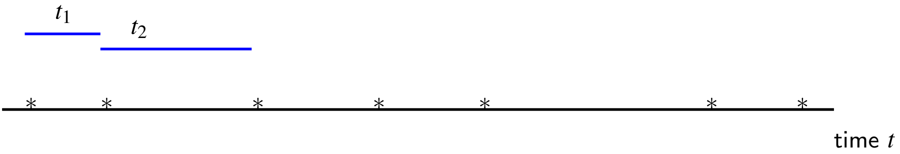
Eksponentialfordelingen

- Eksponentialfordelingen er et specialtilfælde af *gammafordelingen*.
- Eksponentialfordelingen anvendes f.eks. til at beskrive levetider og ventetider.
- Eksponentialfordelingen kan bruges til at beskrive (vente)tiden mellem hændelser i en poissonproces.

Sammenhæng mellem eksponential- og poissonfordelingen

Poisson: Diskrete hændelser pr. enhed

Eksponential: Kontinuert afstand mellem hændelser



Eksempel 5

Kø-model: Poissonproces

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

Eksempel 5

Kø-model: Poissonproces

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

Spørgsmål:

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

Eksempel 5

Kø-model: Poissonproces

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

Spørgsmål:

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

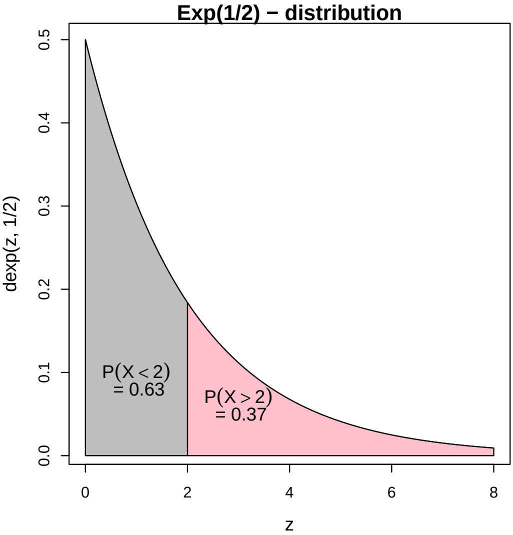
Svar:

Lad $X \sim \text{Exp}(1/2)$ repræsentere ventetiden indtil ankomsten af den næste kunde.
$P(X > 2) = 1 - P(X \leq 2)$

```
1 - pexp(2, rate = 1/2)
```

```
[1] 0.3679
```

Eksempel 5


Exp(1/2) − distribution
dexp(z, 1/2)
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
z
P(X < 2)
= 0.63
P(X > 2)
= 0.37

Eksempel 6

Spørgsmål:

En kunde er netop ankommet.
Brug poissonfordelingen til at beregne sandsynligheden for, at der ikke kommer flere kunder inden for de næste to minutter.

Eksempel 6

Spørgsmål:

En kunde er netop ankommet.
Brug poissonfordelingen til at beregne sandsynligheden for, at der ikke kommer flere kunder inden for de næste to minutter.

Svar:

$\lambda_{2min} = 1$, $P(X = 0) = \frac{e^{-1}}{1!} 1^0 = e^{-1}$

```
dpois(0,1)
```

[1] 0.3679

```
exp(-1)
```

[1] 0.3679

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable

Regneregler for stokastiske variable

Disse regneregler gælder både for kontinuerte og diskrete stokastiske variable!

Lad X være en stokastisk variabel, medens a og b er konstanter.

Regneregler for stokastiske variable

Disse regneregler gælder både for kontinuerte og diskrete stokastiske variable!

Lad X være en stokastisk variabel, medens a og b er konstanter.

Middelværdi-regel:

$$\mathsf{E}(aX+b)=a\mathsf{E}(X)+b$$

Regneregler for stokastiske variable

Disse regneregler gælder både for kontinuerte og diskrete stokastiske variable!

Lad X være en stokastisk variabel, medens a og b er konstanter.

Middelværdi-regel:

$$\mathsf{E}(aX+b) = a\mathsf{E}(X)+b$$

Varians-regel:

$$\mathsf{Var}(aX+b) = a^2\mathsf{Var}(X)$$

Eksempel 7

Lad X være en stokastisk variabel med middelværdi 4 og varians 6.

Spørgsmål:

Beregn middelværdien og variansen af $Y = -3X + 2$.

Eksempel 7

Lad X være en stokastisk variabel med middelværdi 4 og varians 6.

Spørgsmål:

Beregn middelværdien og variansen af $Y = -3X + 2$.

Svar:

$$\mathsf{E}(Y) = -3\mathsf{E}(X) + 2 = -3 \cdot 4 + 2 = -10$$

$$\mathsf{Var}(Y) = (-3)^2 \mathsf{Var}(X) = 9 \cdot 6 = 54$$

Regneregler for stokastiske variable

Lad $X_1, \ldots, X_n$ være *uafhængige* stokastiske variable.

Regneregler for stokastiske variable

Lad $X_1, \ldots, X_n$ være *uafhængige* stokastiske variable.

Middelværdi-regel:

$$\mathsf{E}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$
$$= a_1 \mathsf{E}(X_1) + a_2 \mathsf{E}(X_2) + \cdots + a_n \mathsf{E}(X_n)$$

Regneregler for stokastiske variable

Lad $X_1, \ldots, X_n$ være *uafhængige* stokastiske variable.

Middelværdi-regel:

$$\mathsf{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1\mathsf{E}(X_1) + a_2\mathsf{E}(X_2) + \cdots + a_n\mathsf{E}(X_n)$$

Varians-regel:

$$\mathsf{Var}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1^2\mathsf{Var}(X_1) + \cdots + a_n^2\mathsf{Var}(X_n)$$

Eksempel 8

Planlægning for flyselskab

Vægten af én passager på en flytur, X, antages at være normalfordelt så $X \sim N(70, 10^2)$.

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

Eksempel 8

Planlægning for flyselskab

Vægten af én passager på en flytur, X, antages at være normalfordelt så $X \sim N(70, 10^2)$.

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

Spørgsmål:

Beregn sandsynligheden for, at flyet bliver overlastet.

Eksempel 8

Planlægning for flyselskab

Vægten af én passager på en flytur, X, antages at være normalfordelt så $X \sim N(70, 10^2)$.

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

Spørgsmål:

Beregn sandsynligheden for, at flyet bliver overlastet.

Hvad er den samlede passagervægt Y på en afgang?

Eksempel 8

Planlægning for flyselskab

Vægten af én passager på en flytur, X, antages at være normalfordelt så $X \sim N(70, 10^2)$.

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

Spørgsmål:

Beregn sandsynligheden for, at flyet bliver overlastet.

Hvad er den samlede passagervægt Y på en afgang?

Hvad er Y?

IKKE: $Y = 55 \cdot X$

Eksempel 8

Hvad er den samlede passagervægt Y?

$Y = \sum_{i=1}^{55} X_i$, hvor $X_i \sim N(70, 10^2)$ (som antages at være uafhængige)

Eksempel 8

Hvad er den samlede passagervægt Y?

$Y = \sum_{i=1}^{55} X_i$, hvor $X_i \sim N(70, 10^2)$ (som antages at være uafhængige)

Middelværdi og varians af Y:

$$\mathsf{E}(Y) = \sum_{i=1}^{55} \mathsf{E}(X_i) = \sum_{i=1}^{55} 70 = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = \sum_{i=1}^{55} \mathsf{Var}(X_i) = \sum_{i=1}^{55} 100 = 55 \cdot 100 = 5500$$

Eksempel 8

Hvad er den samlede passagervægt Y?

$Y = \sum_{i=1}^{55} X_i$, hvor $X_i \sim N(70, 10^2)$ (som antages at være uafhængige)

Middelværdi og varians af Y:

$$\mathsf{E}(Y) = \sum_{i=1}^{55} \mathsf{E}(X_i) = \sum_{i=1}^{55} 70 = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = \sum_{i=1}^{55} \mathsf{Var}(X_i) = \sum_{i=1}^{55} 100 = 55 \cdot 100 = 5500$$

Y er normalfordelt, så vi kan finde $P(Y > 4000)$ ved:

```
1-pnorm(4000, mean = 3850, sd = sqrt(5500))
```

```
[1] 0.02156
```

Eksempel 8 - FORKERT analyse

Hvad er Y?

IKKE: $Y = 55 \cdot X$

Eksempel 8 - FORKERT analyse

Hvad er Y?

IKKE: $Y = 55 \cdot X$

Middelværdi og varians af FORKERT Y:

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$
$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Eksempel 8 - FORKERT analyse

Hvad er Y?

IKKE: $Y = 55 \cdot X$

Middelværdi og varians af FORKERT Y:

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$
$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Det FORKERTE Y er også normalfordelt. Her finder vi $P(Y > 4000)$ med FORKERT Y:

```
1 - pnorm(4000, mean = 3850, sd = 550)
```

[1] 0.3925

Eksempel 8 - FORKERT analyse

Hvad er Y?

IKKE: $Y = 55 \cdot X$

Middelværdi og varians af FORKERT Y:

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Det FORKERTE Y er også normalfordelt. Her finder vi $P(Y > 4000)$ med FORKERT Y:

```
1 - pnorm(4000, mean = 3850, sd = 550)
```

[1] 0.3925

Konsekvens af forkert beregning:

MANGE spildte penge for flyselskabet!!!

Dagsorden

1. Opsummering
2. Kontinuerte fordelinger
   - Tætheds- og fordelingsfunktioner
   - Middelværdi, varians og kovarians
3. Vigtige kontinuerte fordelinger
   - Den uniforme fordeling
   - Normalfordelingen
   - Log-normalfordelingen
   - Eksponentialfordelingen
4. Regneregler for stokastiske variable