Landmålingens fejlteori
Kontinuerte stokastiske variable
Lektion 3

Torben Tvedebrink - tvede@math.aau.dk
http://www.math.aau.dk/∼tvede/teaching/L4

Institut for Matematiske Fag
Aalborg Universitet

30. april 2009

Definitioner

Definition 1

Fordelingsfunktion for en stokastisk variabel X er

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

Definitioner

Definition 1

Fordelingsfunktion for en stokastisk variabel X er

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

Definition 2

Lad X være en stokastisk variabel med fordelingsfunktion F. Da siges X at være **kontinuert**, hvis F er kontinuert, og hvis der findes en ikke-negativ reel funktion f, således at

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

Funktionen f kaldes **tæthedsfunktion** for X

Egenskaber

Tæthedsfunktion for kontinuerte stokastiske variable svarer sandsynlighedsfunktion for diskrete stokastiske variable.

Kontinuert variabel	Diskret variabel
$f(x) = F'(x)$	$f(x) = P(X = x)$
$F(x) = \int_{-\infty}^{x} f(y)dy$	$F(x) = \sum_{y \leq x} f(y)$

Bemærk at $P(X = a) = \int_a^a f(y)dy = 0$ for alle værdier $a \in \mathbb{R}$. Dvs sandsynligheden for at en kontinuert variabel antager en bestemt værdi er 0.

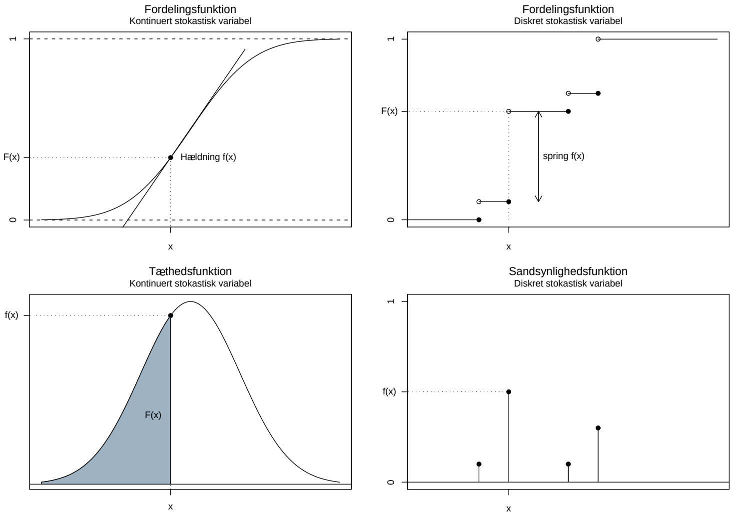
Egenskaber

Fordelingsfunktionen springer ikke, dvs. $P(X = x) = 0$ og

$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a \leq X \leq b) \\ &= F(b) - F(a) \\ &= \int_{-\infty}^{b} f(y)dy - \int_{-\infty}^{a} f(y)dy \\ &= \int_{a}^{b} f(y)dy. \end{aligned}$$

Tæthed- og fordelingsfunktion

Tæthed-, sandsynlighed- og fordelingsfunktion



Middelværdi og varians

Definition 3

Lad X være en kontinuert stokastisk variabel med tæthedsfunktion f.
Middelværdien af X defineres som

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx < \infty$$

- også kaldet forventet værdi.

Middelværdi og varians

Definition 3

Lad X være en kontinuert stokastisk variabel med tæthedsfunktion f. **Middelværdien** af X defineres som

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx < \infty$$

- også kaldet forventet værdi.

Lad X være en kontinuert stokastisk variabel med tæthedsfunktion f. **Variansen** af X defineres som

$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx < \infty,$$

hvor σ kaldes spredningen (eller standard afvigelse) - mål for forventet variabilitet i X.

Variansformel

Definition 3

Lad X være en kontinuert stokastisk variabel med tæthedsfunktion f. **Variansen** af X defineres som

$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx < \infty,$$

hvor σ kaldes spredningen (eller standard afvigelse) - mål for forventet variabilitet i X.

Der gælder (som vi så ved diskrete stokastiske variable)

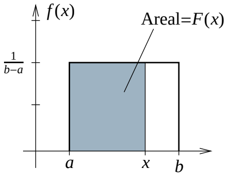
$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x)}_{=1} - 2\mu \underbrace{\int_{-\infty}^{\infty} x f(x) dx}_{=\mu} \\ &= \mathbb{E}(X^2) - \mu^2.\end{aligned}$$

Ligefordelingen

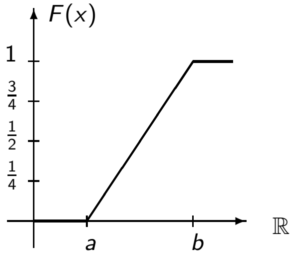
En stokastisk variabel X er **ligefordelt** på intervallet $]a;b[$ (eller $[a;b]$), hvis **fordelingsfunktionen** er givet som

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{ellers} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{a}^{x} \frac{1}{b-a}dy = \left[\frac{y}{b-a}\right]_a^x = \frac{x}{b-a} - \frac{a}{b-a}$$



$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$



Middelværdi - ligefordelingen

Vi kan bestemme middelværdien for ligefordelingen ved ovenstående udtryk for tæthedsfunktionen og middelværdi:

$$
\begin{aligned}
\mu = \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_a^b x \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \left[\frac{1}{2} x^2 \right]_a^b \\
&= \frac{1}{b-a} \frac{1}{2} (b^2 - a^2) \\
&= \frac{1}{b-a} \frac{1}{2} (b-a)(b+a) \\
&= \frac{b+a}{2}
\end{aligned}
$$

Som forventet er middelværdien midtpunktet i intervallet $[a; b]$.

Varians - ligefordelingen

Tilsvarende bestemmes variansen vha. $\sigma^2 = \mathbb{E}(X^2) - \mu^2$:

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b = \frac{1}{b-a}\frac{b^3-a^3}{3}$$

$$\mu^2 = \left(\frac{b+a}{2}\right)^2 = \frac{(b+a)^2}{4}$$

Dermed bliver variansen for ligefordelingen:

$$\sigma^2 = \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \cdots = \frac{(b-a)^2}{12}.$$

Normalfordelingen

Definition 4

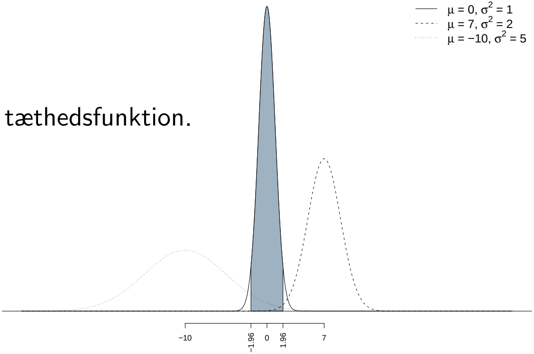
En stokastisk variabel X med **tæthedsfunktion**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},$$

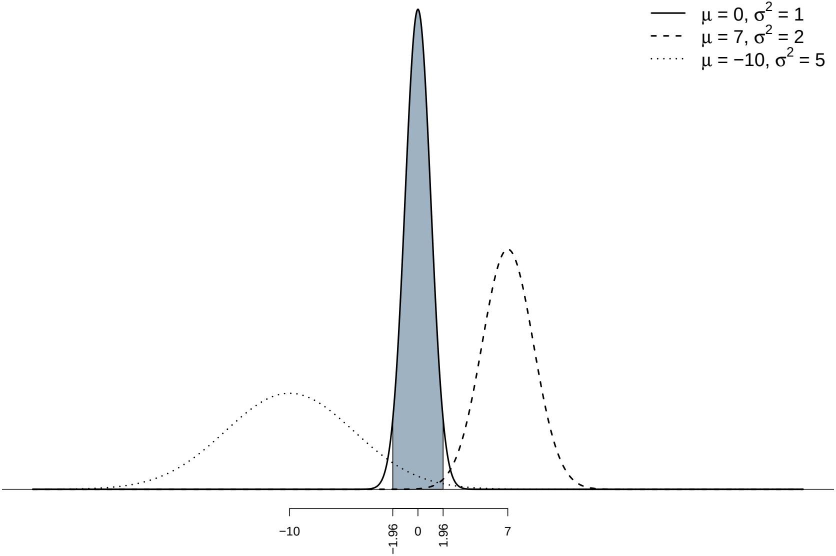
siges at være **normalfordelt** med parameterne μ og σ^2, hvor μ og σ er reelle tal og $\sigma > 0$.

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$.

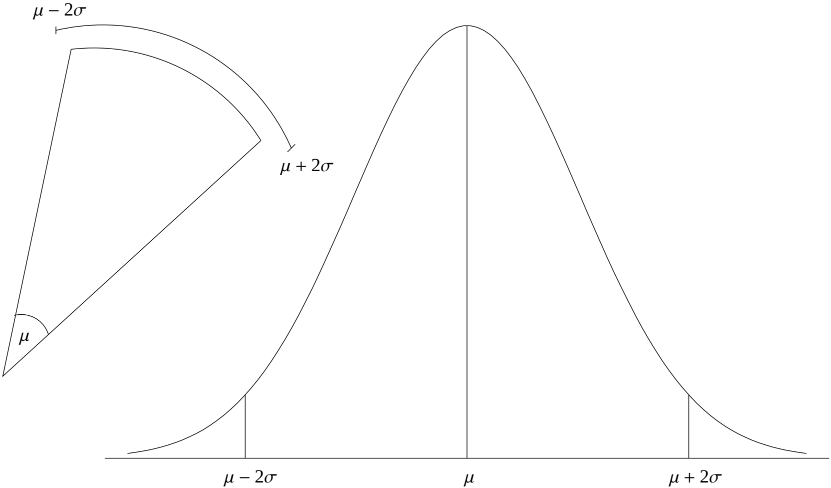
Klokkeformet symmetrisk tæthedsfunktion.



Normalfordelingen



Normalfordelingen



Sætning 2

Hvis $X \sim \mathcal{N}(\mu, \sigma^2)$, så er **middelværdi** og **varians** for X

$$\begin{aligned} \mathbb{E}(X) &= \mu, \\ \mathbb{V}(X) &= \sigma^2. \end{aligned}$$

Fordelingsfunktion for $X \sim \mathcal{N}(\mu, \sigma^2)$ er givet som

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(y-\mu)^2}{2\sigma^2} \right) dy.$$

Der findes ikke et pænt lukket udtryk.

Standard Normalfordelingen

Fordelingen $\mathcal{N}(0,1)$ kaldes **standard normalfordeling**. Typisk noteres standard normal fordelte variable Z.

Tæthedsfunktionen for $\mathcal{N}(0,1)$ er givet ved:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$$

Den tilhørende fordelingsfunktion betegnes $\Phi(z)$:

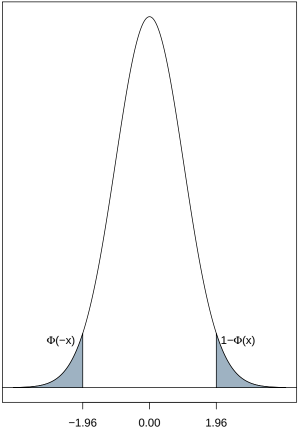
$$\text{Dvs } F(z) = \Phi(z) \text{ når } \mu = 0 \text{ og } \sigma = 1.$$

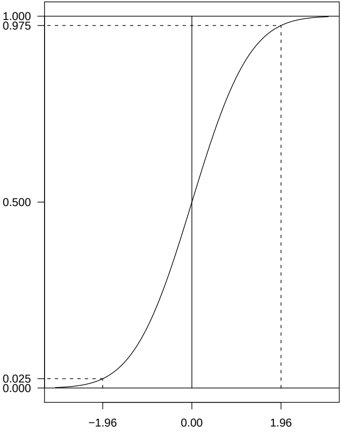
Der findes ikke et 'lukket' udtryk for nedenstående sandsynlighed:

$$P(Z \leq a) = F(a) = \Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)dz$$

Normalfordelingen

Tæthed- og fordelingsfunktion for $\mathcal{N}(0,1)$.





Normalfordelingen

Der gælder

$$\Phi(-a) = \int_{-\infty}^{-a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz = \int_{a}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz$$

Det sidste udtryk bestemmer $P(Z > a)$. Vi kan derfor beregne den komplementære hændelse:

$$\begin{aligned} P(Z > a) &= 1 - P(Z < a) \\ &= 1 - \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz \\ &= 1 - \Phi(a). \end{aligned}$$

Dvs at $\Phi(-a) = 1 - \Phi(a)$.

Transformationer

Sætning **Lineær transformation**
Hvis $X \sim \mathcal{N}(\mu, \sigma^2)$, og $a, b \in \mathbb{R}$ og $a \neq 0$, så er

$$Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

Dvs. at Y er normalfordelt når X er normalfordelt.

$$\begin{aligned} Y \text{ har middelværdi: } & \mathbb{E}(Y) = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = a\mu + b \\ \text{og varians: } & \mathbb{V}(Y) = \mathbb{E}(aX + b) = a^2\mathbb{V}(X) = a^2\sigma^2. \end{aligned}$$

Dermed kan vi transformere en standard normalfordelt variabel Z til at have den middelværdi og varians vi ønsker.

Transformationer

Omvendt kan vi ligeledes standardisere en hvilken som helst normalfordelt variabel X til en standard normalfordelt variabel:

Sætning **Standardisere**
Hvis $X \sim \mathcal{N}(\mu, \sigma^2)$, så gælder

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Beregner middelværdi og varians vi vha. foregående sætning får vi:

$$\mathbb{E}\left(\frac{X - \mu}{\sigma}\right) = \mathbb{E}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{\mathbb{E}(X)}{\sigma} - \frac{\mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

$$\mathbb{V}\left(\frac{X - \mu}{\sigma}\right) = \mathbb{V}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{\mathbb{V}(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

Standard Normalfordelingen

Jf. de foregående slides ved vi:

- $\Phi(-a) = 1 - \Phi(a)$
- $\mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = 0$ når $\mathbb{E}(X) = \mu$.
- $\mathbb{V}\left(\frac{X-\mu}{\sigma}\right) = 1$ når $\mathbb{V}(X) = \sigma$.

Dvs.

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Sandsynligheder vha. tabel for standard normalfordelingen.

Eksempel

Lad $X \sim \mathcal{N}(2,4)$, så er

$$Z = \frac{X-2}{2} \sim \mathcal{N}(0,1).$$

Udsnit fra `normaltab.pdf`:

	0.00	0.01	0.0
0.0	0.5000	0.5040	0.508
0.1	0.5398	0.5438	0.547
0.2	0.5793	0.5832	0.587
0.3	0.6179	0.6217	0.625
0.4	0.6554	0.6591	0.662
0.5	0.6915	0.6950	0.698
0.6	0.7257	0.7291	0.732
0.7	0.7580	0.7611	0.764
0.8	0.7881	0.7910	0.793
0.9	0.8159	0.8186	0.821
1.0	0.8413	0.8438	0.846
1.1	0.8643	0.8665	0.868
1.2	0.8849	0.8869	0.888

Dvs.

$$\begin{aligned} P(X \leq 4) &= P\left(\frac{X-2}{2} \leq \frac{4-2}{2}\right) \\ &= P(Z \leq 1) \\ &= \Phi(1) \\ &= 0.8413 \end{aligned}$$

Tæthedsfunktion - fraktiler

Lad $X \sim \mathcal{N}(\mu, \sigma^2)$, dvs. $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Betragt sandsynligheden

$$
\begin{aligned}
p = P(\mu - z\sigma < X < \mu + z\sigma) &= P\left(-z < \frac{X-\mu}{\sigma} < z\right) \\
&= P(-z < Z < z) \\
&= P(Z < z) - P(Z \leq -z) \\
&= \Phi(z) - \Phi(-z) \\
&= \Phi(z) - (1 - \Phi(z)) \\
&= 2\Phi(z) - 1.
\end{aligned}
$$

Tæthedsfunktion - fraktiler

Lad $X \sim \mathcal{N}(\mu, \sigma^2)$, dvs. $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Betragt sandsynligheden

$$p = P(\mu - z\sigma < X < \mu + z\sigma) = 2\Phi(z) - 1.$$

For $p = 0.95$ gælder

$$2\Phi(z) - 1 = 0.95 \quad \Leftrightarrow \quad \Phi(z) = 0.975$$

Ved opslag i tabellen findes det z som svarer til $\Phi(z) = 0.975$: $z = 1.96$.

Udsnit fra `normaltab.pdf`:

		0.06	0.07
		0.5239	0.5279
		0.5636	0.56
1.7	0.955	0.9608	0.9616
1.8	0.9641	0.9686	0.9693
1.9	0.9713	0.9750	0.9756

Dvs.

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95.$$

Tæthedfuntion - fraktiler

Fra resultatet

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95.$$

har vi altså middelværdien $\pm 2\times$standard afvigelse dækker ca. 95% af sandsynlighedsmassen.

Tilsvarende for $p = 0.99$ og $p = 0.999$

$$\begin{aligned} P(\mu - 2.58\sigma < X < \mu + 2.58\sigma) &= 0.99, \\ P(\mu - 3.29\sigma < X < \mu + 3.29\sigma) &= 0.999. \end{aligned}$$

Tæthedsfunktion