

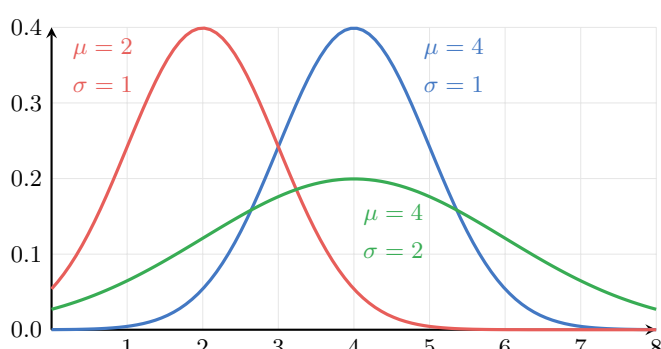
Normalfordelingen

Normalfordelingen er en sammenhængende kurve for *sandsynlighedstætheden*, som bestemmes ud fra en middelværdi μ og en spredning σ . Ud over at kunne give en god approksimation til binomialfordelingen under visse betingelser, forekommer normalfordelingen også hyppigt når man undersøger stokastiske variable, der har en symmetrisk variation omkring en middelværdi som f.eks. måleusikkerhed.

Funktionsforskriften for en normalfordelt *tæthedsfunktion* med middelværdi μ og spredning σ er givet ved:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma^2} \cdot \exp\left(-\frac{(x - \mu)^2}{2 \cdot \sigma^2}\right)$$

I figuren nedenfor er der vist hvilken betydning middelværdien og spredningen har på normalfordelingens form. Det ses at middelværdien μ afgør hvor på 1. akse at normalfordelingen har sit toppunkt og spredningen σ afgør hvor meget fordelingen er spredt fra midtpunktet.



Som det kan ses af figuren er tæthedsfunktionen for normalfordelingen symmetrisk omkring middelværdien og at spredningen virker som en strækningsfaktor.

For at kunne beregne sandsynligheder ud fra normalfordelingen skal man kunne beregne den kumulerede sandsynlighedstæthed. Dette kan gøres ud fra stamfunktionen til tæthedsfunktionen som er givet ved:

$$F(x) = \frac{1}{2} \cdot \left(1 + \operatorname{erf}\left(\frac{x - \mu}{\sqrt{2} \cdot \sigma}\right)\right)$$

Her er funktionen erf den såkaldte *error function*. Stamfunktionen $F(x)$ opfylder betingelsen at:

$$F(x) \rightarrow 0 \quad \text{når} \quad x \rightarrow -\infty$$

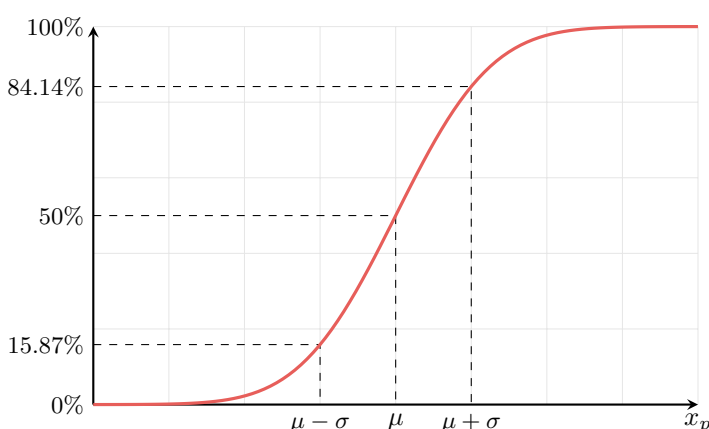
Sandsynligheder for, at en normalfordelt stokastisk variabel tilhører et givet interval beregnes ved:

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

$$P(a \leq x \leq b) = F(b) - F(a)$$

$$P(x \geq a) = 1 - F(a)$$

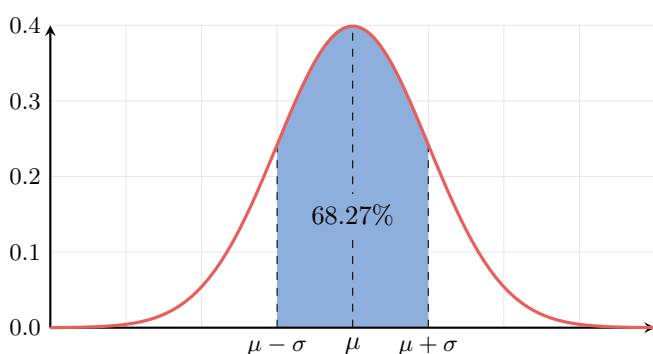
Sammenlignet med statistiske fordelinger udgør stamfunktionen $F(x)$ altså sumkurven for normalfordelingen, også kaldet *fordelingsfunktionen* og er vist i figuren nedenfor.



For intervallet der dækker 1 spredning på hver side af middelværdien beregnes sandsynligheden for at den stokastiske variabel ligger i dette interval til:

$$\begin{aligned} P(\mu - \sigma \leq x \leq \mu + \sigma) &= F(\mu + \sigma) - F(\mu - \sigma) \\ &\approx 0.8414 - 0.1587 \\ &\approx 68.27\% \end{aligned}$$

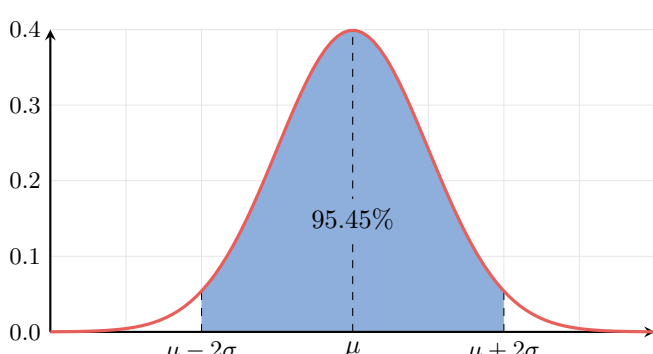
Intervallet er illustreret i figuren nedenfor hvor den kumulerede sandsynlighed er vist som arealet under tæthedsfunktionen på 68.27%.



Intervallet, der dækker 2 spredninger på hver side af middelværdien, dækker over det man kalder *normale udfald*, idet der er lidt over 95% sandsynlighed for at et udfald vil ligge i dette interval. Dette kan ses ved at beregne sandsynligheden for at den stokastiske variabel ligger i dette interval ved:

$$\begin{aligned} P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) &= F(\mu + 2\sigma) - F(\mu - 2\sigma) \\ &\approx 95.45\% \end{aligned}$$

Intervallet er illustreret i figuren nedenfor hvor den kumulerede sandsynlighed igen er vist som arealet under tæthedsfunktionen på 95.45%.



For det sidste interval, der dækker 3 spredninger på hver side af middelværdien, beregnes sandsynligheden for at den stokastiske variabel ligger i dette interval ved:

$$\begin{aligned} P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) &= F(\mu + 3\sigma) - F(\mu - 3\sigma) \\ &\approx 99.73\% \end{aligned}$$

Udfald der ligger uden for dette interval kaldes *ekceptionelle udfald* idet der kun er 0.27% sandsynlighed for at ligge uden for dette interval.