Sammenhængen mellem den generelle normalfordeling og standardnormalfordelingen

Det er de færreste observationer i den virkelige verden, som følger en standardnormalfordeling - det vil sige en normalfordeling med middelværdi 0 og spredning 1. Alligevel spiller standardnormal-fordelingen en central rolle. Det skyldes i særdeleshed nedenstående sætning, som beskriver sammenhængen mellem fordelingsfunktionen $F(x) = P(X \leq x)$ for en vilkårlig normalfordelt stokastisk variabel og fordelingsfunktionen for standardnormalfordelingen $\Phi(x)$.

Sætning 1

Lad X være normalfordelt med middelværdi μ og spredning σ. Da gælder

$$P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right),$$

hvor Φ er fordelingsfunktionen for standardnormalfordelingen.

Bevis

Vi skal vise, at

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

Vi vil derfor opskrive udtryk for $P(X \leq a)$ og $\Phi\left(\frac{a-\mu}{\sigma}\right)$ og vise, at de er ens.

Da

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

er tæthedsfunktionen for X:

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

Og derfor er

$$P(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

Dette udtryk regner vi videre på om lidt. Først ser vi på

$$\Phi\left(\frac{a-\mu}{\sigma}\right)$$

For at opstille et udtryk for denne sandsynlighed får vi brug for tæthedsfunktionen for standardnormalfordelingen:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot x^2}$$

Og derfor er

$$\Phi\left(\frac{a-\mu}{\sigma}\right) = \int_{-\infty}^{\frac{a-\mu}{\sigma}} \phi(x)\,dx = \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot x^2} dx$$

Vi vil nu vise, at

$$P(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}\,dx$$

er identisk med udtrykket for $\Phi\left(\frac{a-\mu}{\sigma}\right)$.

Vi udfører bestemt integration ved substitution og sætter

$$t = \frac{x-\mu}{\sigma}$$

Da bliver

$$dt = \frac{1}{\sigma}\,dx$$

Nu mangler vi blot at bestemme nye grænser:

Når $x \to -\infty$ så vil også $t \to -\infty$ og den nedre grænse forbliver derfor uændret.

Når $x = a$ er $t = \frac{a-\mu}{\sigma}$ og den nye øvre grænser bliver således $\frac{a-\mu}{\sigma}$.

Ved at lave denne substitution får vi derfor

$$P(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}\,dx \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot t^2} dt$$

Hvilket er det samme som udtrykket for $\Phi\left(\frac{a-\mu}{\sigma}\right)$, som vi opskrev ovenfor. Altså er

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

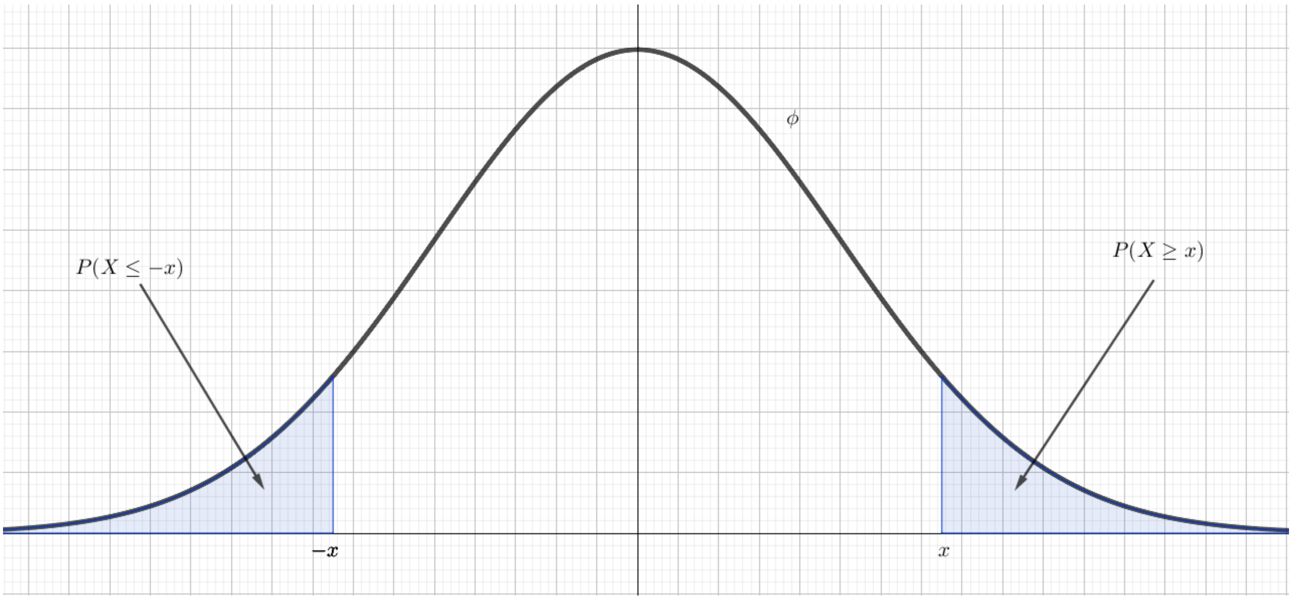
hvilket netop var det, vi skulle vise.

□

Inden vi går videre til de næste spændende sætninger, minder vi lige om en vigtig egenskab ved standardnormalfordelingen (det gælder som regel <u>ikke</u> for en vilkårlig normalfordeling). Den vigtige egenskab er:

$$\Phi(-x) = 1 - \Phi(x).$$

Argumentet for denne påstand ses nemt ved at betragte nedenstående figur. Antag her, at $X \sim N(0,1)$.



Da grafen for tæthedsfunktionen ϕ er symmetrisk omkring y-aksen, så må der gælde, at

$$P(X \leq -x) = P(X \geq x)$$

Se figuren (det er de to blåt skraverede områder, som er lige store). Samtidig gælder der helt generelt, at

$$P(X \geq x) = 1 - P(X \leq x)$$

Indsættes dette i ovenstående udtryk fås

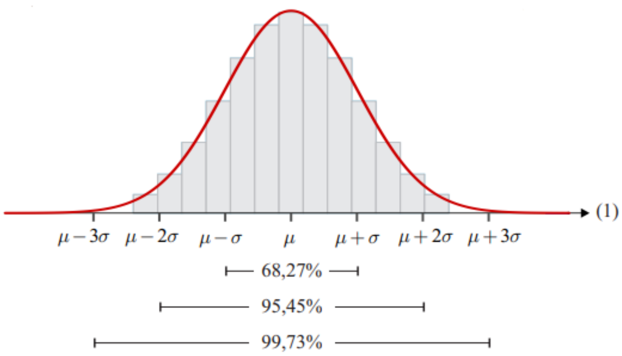
$$P(X \leq -x) = 1 - P(X \leq x)$$

Og da der er tale om standardnormalfordelingen er $\Phi(x) = P(X \leq x)$ og derfor ender vi med

$$\Phi(-x) = 1 - \Phi(x)$$

Du kan jo selv overveje hvorfor noget tilsvarende ikke gælder for en generel normalfordelt stokastisk variabel (og hvad der evt. skal til for, at det gælder).

I formelsamlingen kan man se følgende figur:



Budskabet i denne figur svarer til påstanden i sætning 3, men inden vi beviser denne sætning, vil vi først vise noget lidt mere generelt:

Sætning 2

Lad X være normalfordelt med middelværdi μ og spredning σ. Da gælder

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) = 2 \cdot \Phi(k) - 1,$$

hvor Φ er fordelingsfunktionen for standardnormalfordelingen.

Bemærk, at sandsynligheden hverken afhænger af μ eller af σ, men kun af k!

Bevis

Vi bruger først, at vi tidligere har vist, at

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

Derfor kan vi omskrive omskrive:

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) = P(X \leq \mu + k \cdot \sigma) - P(X \leq \mu - k \cdot \sigma)$$

Nu kan vi bruge sætning 1 og få:

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) = P(X \leq \mu + k \cdot \sigma) - P(X \leq \mu - k \cdot \sigma)$$

$$= \Phi\left(\frac{\mu + k \cdot \sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k \cdot \sigma - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{k \cdot \sigma}{\sigma}\right) - \Phi\left(\frac{-k \cdot \sigma}{\sigma}\right)$$

$$= \Phi(k) - \Phi(-k)$$

Vi har netop redegjort for, at

$$\Phi(-k) = 1 - \Phi(k)$$

Indsættes dette i ovenstående får vi altså, at

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) = \Phi(k) - \big(1 - \Phi(k)\big)$$

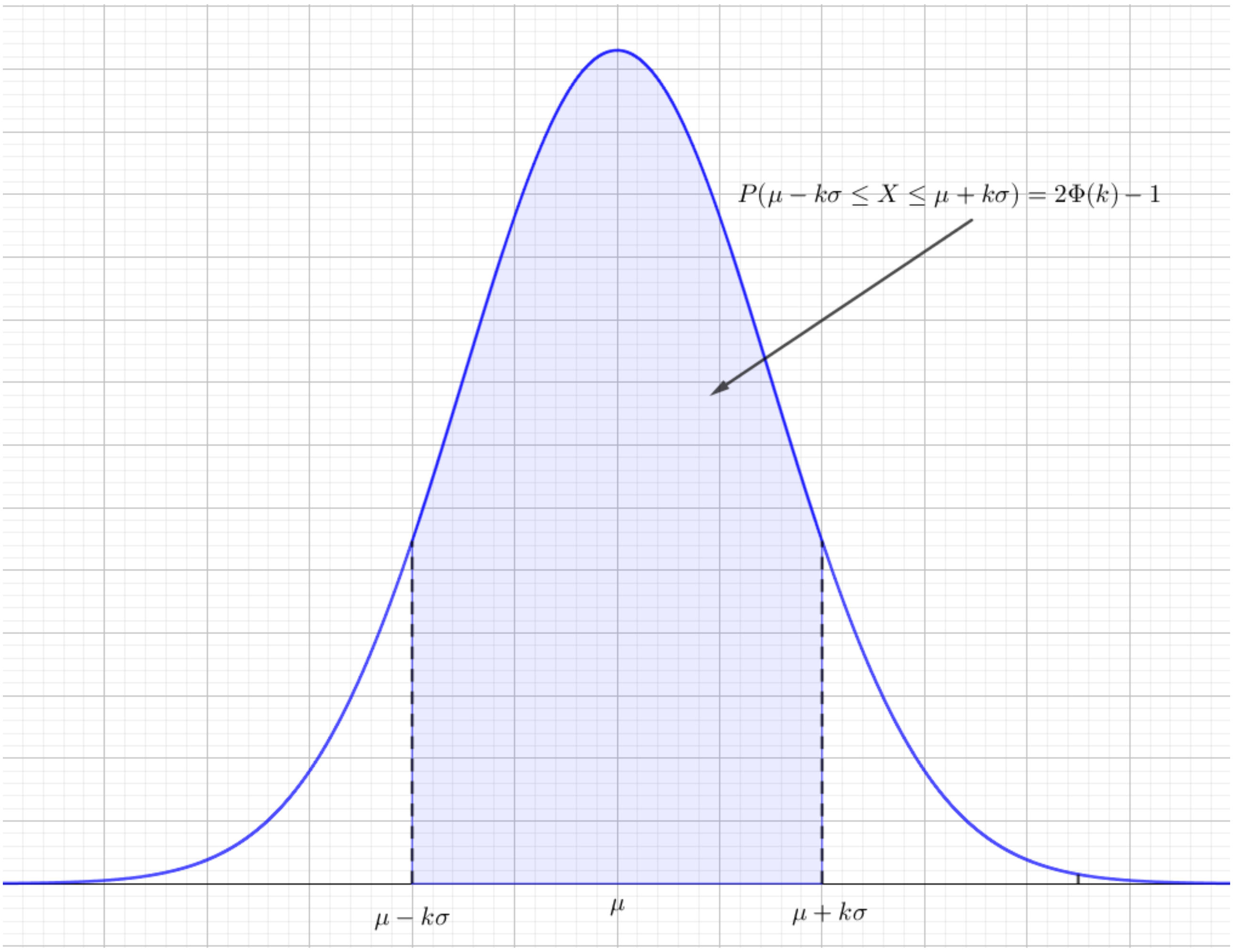
$$= \Phi(k) - 1 + \Phi(k)$$

$$= 2 \cdot \Phi(k) - 1$$

$\square$

Sætning 2 siger, at hvis en stokastisk variabel er normalfordelt, så er sandsynligheden for, at en tilfældig observation ligger k spredninger væk fra middelværdien, på $2 \cdot \Phi(k) - 1$. Denne sandsynlighed afhænger *kun* af k og altså ikke af hverken middelværdien eller spredningen. Det er da interessant! Resultatet er illustreret på nedenstående figur:

Bevis



Så mangler vi kun den sidste sætning:

Sætning 3

Lad X være normalfordelt med middelværdi μ og spredning σ. Da gælder

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 68{,}27\%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 95{,}45\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99{,}73\%$$

Bemærk, at sandsynlighederne hverken afhænger af μ eller af σ!

Bevis

Beviset for denne sætning følger nemt af sætning 2. Der er nemlig blot tale om tre special tilfælde med $k = 1$, $k = 2$ og $k = 3$.

Når $k = 1$ får vi

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 2 \cdot \Phi(1) - 1$$

$$= 2 \cdot 0{,}8413 - 1 = 68{,}26\%$$

For $k = 2$ fås

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2 \cdot \Phi(2) - 1$$

$$= 2 \cdot 0{,}9772 - 1 = 95{,}44\%$$

Og endelig for $k = 3$ har vi

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2 \cdot \Phi(3) - 1$$

$$= 2 \cdot 0{,}9987 - 1 = 99{,}74\%$$

I alle tre tilfælde kan $\Phi(1)$, $\Phi(2)$ og $\Phi(3)$ slås op i en tabel over fordelingsfunktionen for standardnormalfordelingen. Bemærk, at decimalerne ikke stemmer helt. Det skyldes afrunding i tabellen.