Bevis

Vi vil gerne vise, at hvis et datasæt $x_1, x_2, \ldots, x_n$ er approksimativt normalfordelt, så vil deres z-værdier

$$z_i = \frac{x_i - \mu}{\sigma}$$

være standardnormalfordelte.

Sætning 1.1. *Lad $X \sim \mathrm{N}(\mu, \sigma)$ være en normalfordelt stokastisk variabel. Så gælder der, at*

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

altså at Z er standardnormalfordelt.

Bevis. Strategien i beviset er at vise, at $P(Z < a) = \Phi(a)$, altså at fordelingsfunktionen for Z er den samme som fordelingsfunktionen for den standardnormalfordelte stokastiske variabel. Vi betragter følgende.

$$\begin{aligned} P(Z < a) &= P\left(\frac{X - \mu}{\sigma} < a\right) \\ &= P(X - \mu < a\sigma) \\ &= P(X < a\sigma + \mu). \end{aligned}$$

Dette er netop fordelingsfunktion F for X, som bekendt er givet ved

$$\begin{aligned} P(X < a\sigma + \mu) &= F(a\sigma + \mu) \\ &= \int_{-\infty}^{a\sigma+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \end{aligned}$$

Vi ønsker at lave substitutionen $z = \frac{x-\mu}{\sigma}$, og vi får derfor

$$\frac{dz}{dx} = \frac{1}{\sigma}$$

eller

$$dz\sigma = dx.$$

Vi skal også ændre grænserne. For den nedre grænse får vi

$$\frac{-\infty - \mu}{\sigma} = -\infty$$

og for den øvre grænse får vi

$$\frac{a\sigma + \mu - \mu}{\sigma} = \frac{a\sigma}{\sigma} = a.$$

Vi substituerer og får

$$\begin{aligned}
\int_{-\infty}^{a\sigma+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx &= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(z)^2} dz\sigma \\
&= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= \Phi(a).
\end{aligned}$$

Vi har nu fundet ud af, at

$$P(Z < a) = \Phi(a),$$

så derfor må Z være standardnormalfordelt. ■