Normalfordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Normalfordeling

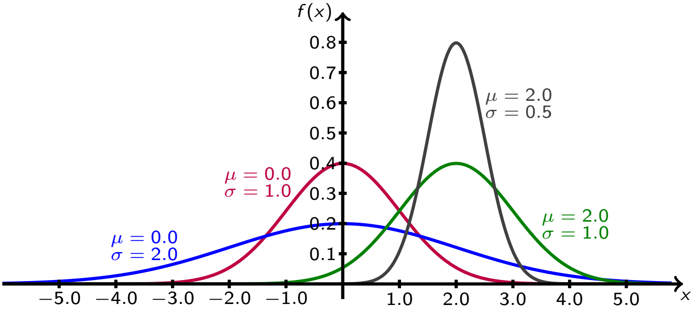
Definisjon (Normalfordeling)

En kontinuerlig fordelt stokastisk variabel X sies å være normalfordelt med forventningsverdi μ og varians $\sigma^2 > 0$ dersom sannsynlighetstettheten er gitt ved

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{ -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right\} \quad \textit{for} \quad -\infty < x < \infty.$$

Hvis $\mu = 0$ og $\sigma^2 = 1$ sier vi at X er standard normalfordelt.

- ⋆ Normalfordelingen er den klart viktigste fordelingen
  - erfaring tilsier at mye er normalfordelt (målefeil, naturlige variasjoner, osv)
  - enkel å regne på
  - mange fordelinger kan tilnærmes med normalfordeling
  - «mye blir normalfordelt når noe blir stort» — mer om dette senere



Forventningsverdi og varians i normalfordeling

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

Varians: $\mathrm{Var}[X] = E[(X-\mu)^2]$

Delvis integrasjon: $\int_a^b u dv = [uv]_a^b - \int_a^b v du$

- Merk: $f(x)$ er symmetrisk om μ

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

- Siden $f(x)$ er symmetrisk om μ har vi at

$$E[X] = \mu$$

- Definisjonen av varians gir

Delvis integrasjon:

$$u = \frac{\sigma}{\sqrt{2\pi}}(x-\mu), \quad dv = \frac{x-\mu}{\sigma^2} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$$

$$du = \frac{\sigma}{\sqrt{2\pi}} dx, \quad v = -e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$\begin{aligned}
\mathrm{Var}[X] &= E\left[(X-\mu)^2\right] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\
&= \left[\frac{\sigma}{\sqrt{2\pi}}(x-\mu) \cdot \left(-e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\right) \cdot \frac{\sigma}{\sqrt{2\pi}} dx \\
&= 0 - 0 + \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\
&= \sigma^2
\end{aligned}$$

Kumulativ fordeling og standardisering

- Sannsynlighetstetthet

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

- Kumulativ fordelingsfunksjon

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{ -\frac{1}{2} \frac{(t-\mu)^2}{\sigma^2} \right\} dt$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{ -\frac{1}{2} v^2 \right\} \sigma dv$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(v) dv$$

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

Subst:
$v = \frac{t-\mu}{\sigma}$
$t = \mu + \sigma v$
$dt = \sigma dv$

$\varphi(x)$: sannsynlighetstetthet for standard normalfordeling
$\Phi(x)$: kumulativ fordelingsfunksjon for standard normalfordeling

- Tilstrekkelig med tabell over sannsynligheter i standard normalfordeling
- Eksempel: Anta $X \sim N(\mu, \sigma^2)$ med $\mu = 1.2$, $\sigma = 0.45$

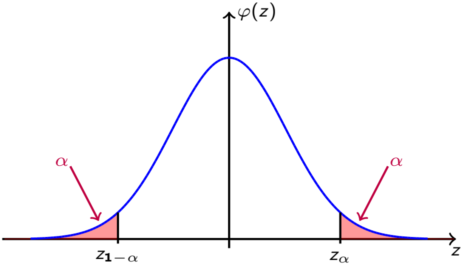
$$P(X \le 0.67) = P\left(Z \le \frac{0.67 - 1.2}{0.45} \right) = \Phi(-1.18) = 0.1190$$

  - eventuelt finne $P(X \le x)$ ved å kalle en funksjon

Kvantiler i standard normalfordeling

- Sannsynlighetstetthet i standard normalfordeling

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}, -\infty < z < \infty$$

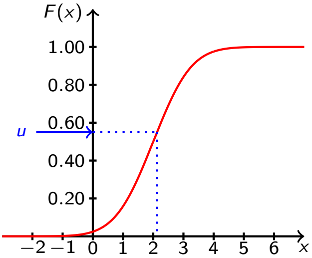


- Ønsker nå å finne en verdi z_α slik at $P(Z > z_\alpha) = \alpha$
  - verdien z_α kan finnes i tabell
  - verdien z_α kan finnes ved å kalle en funksjon
- Merk: Siden $\varphi(z)$ er symmetrisk om $z = 0$ får vi at

$$z_{1-\alpha} = -z_\alpha$$

Simulering fra normalfordeling

- Har tidligere diskutert generell algoritme for å simulere fra en kontinuerlig fordeling



- For normalfordeling
  - har ikke formel for $F(x)$
  - kan ikke finne formel for $F^{-1}(u)$

- Kan vises: Dersom $U_1 \sim \text{Unif}[0,1]$ og $U_2 \sim \text{Unif}[0,1]$ uavhengig av hverandre, har vi at

$$Z_1 = \sqrt{-2\ln(U_1)} \cdot \cos(2\pi U_2) \sim N(0,1),$$
$$Z_2 = \sqrt{-2\ln(U_1)} \cdot \sin(2\pi U_2) \sim N(0,1),$$

  og Z_1 og Z_2 er uavhengige.

- Hvordan simulere fra $N(\mu, \sigma^2)$ når $\mu \neq 0$ og/eller $\sigma \neq 1$?
  - kommer senere!

Oppsummering

- ⋆ Har definert normalfordeling
  - – ved å angi sannsynlighetstettheten, $f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{ -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right\}, -\infty < x < \infty$
- ⋆ Har funnet
  - – forventningsverdi, $E[X] = \mu$
  - – varians, $\text{Var}[X] = \sigma^2$
- ⋆ Har diskutert hvordan man kan
  - – finne sannsynligheter $P(X \leq x)$ ved hjelp av tabell over standard normalfordeling
  - – finne kvantiler z_α i standard normalfordeling fra tabell

- ⋆ Normalfordeling kalles også gaussfordeling, oppkalt etter Carl Friedrich Gauß (1777-1855): tysk matematiker, astronom, geodet og fysiker