

# Normalfordelingen

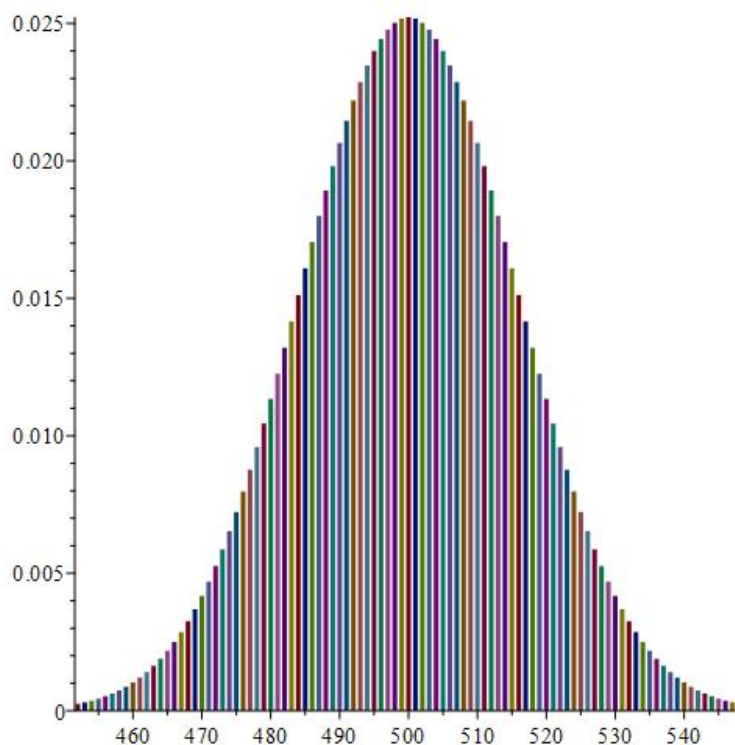
## Approksimation af binomialfordelingen

Vi starter vores normalfordelingsforløb op med at betragte binomialfordelingen fra sidste år.

**Eksempel 1.1.** En mønt kastes 1000 gange. Vi betragter udfaldet krone som en succes. Antallet af kroner kan så modelleres som en binomialfordelt stokastisk variabel med sandsynlighedsparameter  $p = 0.5$  og antalsparameter  $n = 1000$  skrevet  $X \sim B(1000, 0.5)$ . Vi husker på, at sandsynlighedsfunktionen for denne binomialfordelte stokastiske variabel er

$$P(X = k) = \binom{n}{k} (p)^k (1 - p)^{n-k} = \binom{1000}{k} (0.5)^k (0.5)^{1000-k}.$$

Vi tegner sandsynlighedsfunktionen som et pindediagram. Dette kan ses på Fig. 1



Figur 1: Plot af sandsynlighedsfunktion for  $X \sim B(n, p)$ .

Af Fig. 1 kan det tydeligt ses, at det mest sandsynlige udfald er 500. Vi husker på, at middelværdien  $\mu$  for  $X \sim B(n, p)$  er givet ved

$$\mu = np = 1000 \cdot 0.5 = 500$$

og spredningen  $\sigma$  er givet ved

$$\sigma = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0.5 \cdot 0.5} \approx 15.8$$

Den første henseende vi skal betragte normalfordelingen i er til at approksimere binomialfordelingen når  $n$  bliver stor. Lad os derfor betragte den binomialfordelte stokastiske variabel fra Eksempel 1.1

**Eksempel 1.2.** Vi betragter fire forskellige binomialfordelte variable. Én hvor vi kaster 10 gange med en mønt, én hvor vi kaster 100 gange med en mønt, én hvor vi kaster 300 gange med en mønt og én hvor vi kaster 1000 gange med en mønt. Disse lyder henholdsvis

$$\begin{aligned} X_1 &\sim B(10, 0.5), & X_2 &\sim B(100, 0.5), \\ X_3 &\sim B(300, 0.5), & X_4 &\sim B(1000, 0.5). \end{aligned}$$

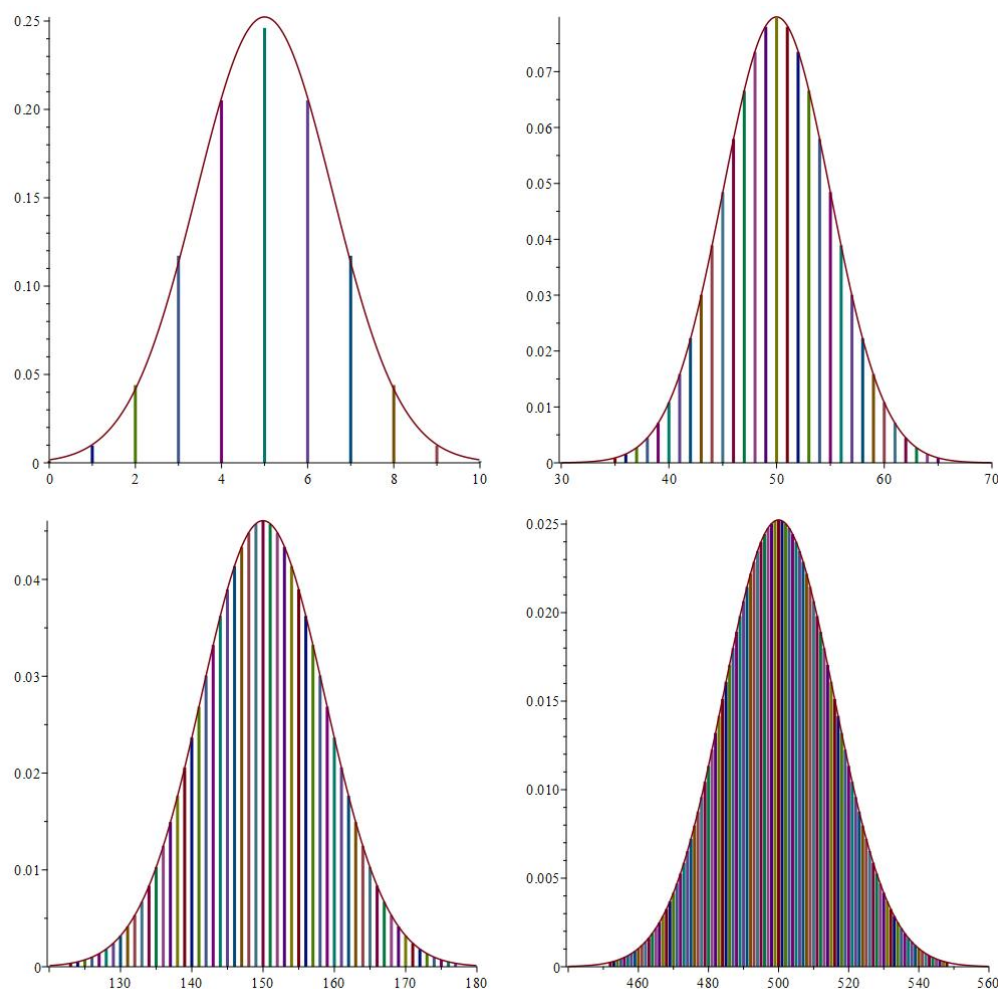
Disse har henholdsvis middelværdierne

$$\begin{aligned} \mu_1 &= 5, & \mu_2 &= 50, \\ \mu_3 &= 150, & \mu_4 &= 500. \end{aligned}$$

De har desuden spredningerne

$$\begin{aligned} \sigma_1 &\approx 1.58, & \sigma_2 &= 5 \\ \sigma_3 &\approx 8.66, & \sigma_4 &\approx 15.81. \end{aligned}$$

Vi vil nu for hver af disse stokastiske variable approksimere dem med en normalfordelt stokastisk variabel der har samme middelværdi og spredning som de binomialfordelte stokastiske variable henholdsvis. Hvad dette rent faktisk betyder vil vi præcisere senere. Approximationerne fremgår af Fig. 2



Figur 2: Normalfordelingsapproximation af binomialfordeling.

Af Fig 2 kan vi se, at approksimationen bliver bedre og bedre jo større  $n$  er. Det skal også bemærkes, at approksimationen bliver dårlig, når  $p$  er tæt på 0 eller 1, men den er god, hvis  $p = 0.5$  som i dette tilfælde.

Kurverne der approksimerer binomialfordelingerne kaldes for *Gauss-kurver* og de udgør det, vi kalder for *tæthedsfunktionen* for normalfordelingen. Vi definerer nu den normalfordelte stokastiske variabel.

**Definition 1.3** (Normalfordelt stokastisk variabel). Den normalfordelte stokastiske variabel  $X \sim N(\mu, \sigma)$  er den kontinuerte stokastiske variabel, der har tæthedsfunktionen

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

og fordelingsfunktionen

$$P(X \leq a) = F(a) = \int_{-\infty}^a f(x)dx.$$

Der er en vigtig pointe her! For en kontinuert stokastisk variabel er sandsynligheden for et diskret udfald  $P(X = k) = 0$ , da det dækker over integralet på et interval med længde 0.

En særlig normalfordelt stokastisk variabel fås, når  $\mu = 0$ , og  $\sigma = 1$ . Så siges  $X \sim N(0, 1)$  at være en *standardnormalfordelt stokastisk variabel*.

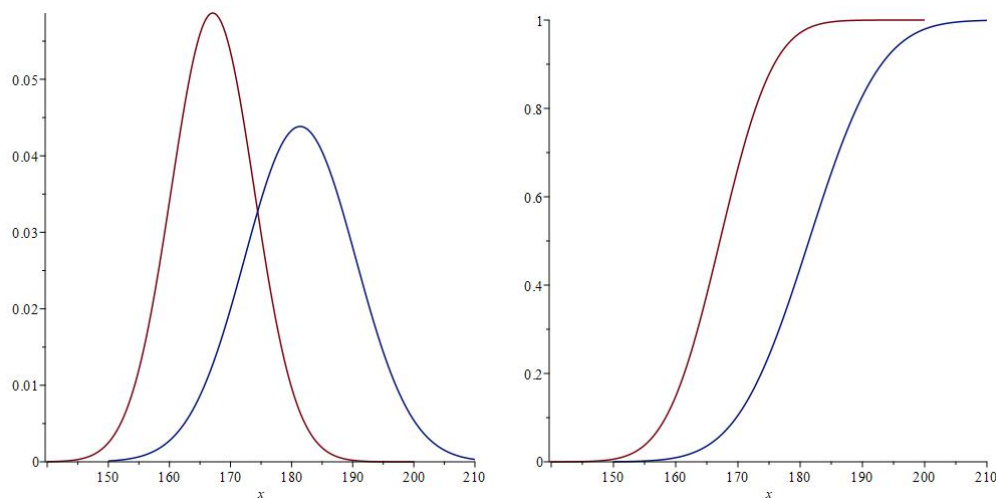
**Eksempel 1.4.** En stor undersøgelse af højden af danske kvinder har bestemt, at danske kvinders højde er approksimativt normalfordelt med middelværdi 167.1 og spredning 6.8, altså

$$X_k \sim N(167.1, 6.8)$$

Tilsvarende er højden af mænd normalfordelt med

$$X_m \sim N(181.4, 9.1).$$

Tæthedsfunktionen og fordelingsfunktioner henholdsvis for fordelingen af mænd og kvinders højde i Danmark.



Figur 3: Tæthedsfunktioner og fordelingsfunktioner henholdsvis for fordelingen af mænd og kvinders højde i Danmark.

For den normalfordelte stokastiske variabel gælder der, at

$$\begin{aligned}P(\mu - \sigma < X < \mu + \sigma) &\approx 0.683, \\P(\mu - 2\sigma < X < \mu + 2\sigma) &\approx 0.954, \\P(\mu - 3\sigma < X < \mu + 3\sigma) &\approx 0.997.\end{aligned}$$

Alle udfald med afstand to spredninger fra middelværdien kaldes for normale udfald. Alle udfald med større end tre spredningers afstand fra middelværdien kaldes for exceptionelle udfald. Tager vi udgangspunkt i højderne for kvinderne så vil vi altså have, at

$$\begin{aligned}P(160.3 < X < 173.9) &\approx 0.683, \\P(153.5 < X < 180.7) &\approx 0.954, \\P(146.7 < X < 187.5) &\approx 0.997.\end{aligned}$$

Et exceptionelt udfald kan i dette tilfælde være en person med højden 140 centimeter. Det skal dog bemærkes, at højden kun er approksimativt normalfordelt, og bliver dårlig særligt i halerne på fordelingen.

## Opgave 1

En person slår med en terning 115 gange, og han antager, at antallet af seksere er binomialfordelt.

- i) Bestem antalsparameteren  $n$  og sandsynlighedsparameteren  $p$  for den binomialfordelte stokastiske variabel  $X_b$ , der beskriver antallet af seksere.
- ii) Bestem middelværdien og spredningen for  $X_b$ .
- iii) Bestem en normalfordelt stokastisk variabel  $X_n$ , der kan bruges til at approksimere  $X_b$
- iv) Brug fordelingsfunktionen for  $X_n$  til at approksimere sandsynligheden for at få mindre end 14 seksere. (Sammenlign eventuelt med sandsynlighed fra fordelingsfunktionen for  $X_b$ )

## Opgave 2

I et klinisk forsøg har et lægemiddel vist sig at helbrede 67.3% af patienter, der får lægemiddelet ordineret. I et land gives lægemiddelet til 27.361 personer.

- i) Bestem antalsparameteren  $n$  og sandsynlighedsparameteren  $p$  for den binomialfordelte stokastiske variabel  $X_b$ , der beskriver antallet af helbredte

- ii) Bestem middelværdien og spredningen for  $X_b$ .
- iii) Bestem en normalfordelt stokastisk variabel  $X_n$ , der kan bruges til at approksimere  $X_b$
- iv) Brug fordelingsfunktionen for  $X_n$  til at approksimere sandsynligheden for at mere end 20.000 helbredes.

## Opgave 3

En producent af printplader har bestilt 150.000 kondensatorer. Af disse har han fået 2377 defekte kondensatorer, hvilket han synes er lidt for mange. Producenten af kondensatorer lover, at kun 1% af kondensatorerne er defekte.

- i) Bestem antalsparameteren  $n$  og sandsynlighedsparameteren  $p$  for den binomialfordelte stokastiske variabel  $X_b$ , der beskriver antallet af defekte kondensatorer, hvis vi antager, at producenten af kondensatorer taler sandt.
- ii) Bestem middelværdien og spredningen for  $X_b$ .
- iii) Bestem en normalfordelt stokastisk variabel  $X_n$ , der kan bruges til at approksimere  $X_b$
- iv) Brug fordelingsfunktionen for  $X_n$  til at approksimere sandsynligheden for at få 2377 eller flere defekte kondensatorer. Har producenten af printplader grund i sin mistanke?

## Opgave 4

I har en aflevering (: