# Middelværdi og varians

## Middelværdi

Har vi en stokastisk variabel $X$, så kunne vi godt tænke os at beskrive det mest sandsynlige udfald af den stokastiske variabel. Betragter vi 1000 slag med en terning, så vil det gennemsnitlige slag være summen af alle de slåede slag delt med 1000. Lad os se på et eksempel:

**Eksempel 1.1.** Vi slår 1000 gange med en terning og får følgende resultater

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 157 | 162 | 165 | 176 | 173 | 167 |

Vi vil så bestemme det gennemsnitlige slag som

$$\frac{1 \cdot 157 + 2 \cdot 162 + 3 \cdot 165 + 4 \cdot 176 + 5 \cdot 173 + 6 \cdot 167}{1000} = 3,445.$$

Vi har derfor eksperimentelt bestemt den forventede værdi, når vi slår med en terning.

Vi vil bruge mere eller mindre det samme princip, når vi definerer *middelværdien for en stokastisk variabel.*

**Definition 1.2** (Middelværdi)**.** Lad $X$ være en (diskret) stokastisk variabel med billedmængde $\{x_1, x_2, \ldots, x_n\}$ og fordelingsfunktion $P$. Så defineres middelværdien $\mathbb{E}(X)$ for en stokastisk variabel som

$$\begin{aligned}\mathbb{E}(X) = \mu &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \cdots + x_n \cdot P(X = x_n)\\ &= \sum_{i=1}^{n} x_i P(X = x_i).\end{aligned}$$

Middelværdien giver os et indblik i, hvad den stokastiske variabel mest sandsynligt antager af værdi, men den siger ikke noget om, hvor meget værdierne, variablen kan antage varierer omkring middelværdien $\mu$. Vi vil definere variansen som den "gennemsnitlige variation fra middelværdien". Derfor ville det give mening at definere variansen som

$$\mathrm{Var} = \mathbb{E}(X - \mu),$$

men dette vil tage højde for, at $X - \mu$ nogle gange er negativ og nogle gange er positiv. Vi vil bare gerne vide, hvor langt $X$ ligger fra $\mu$, så derfor definerer vi variansen på følgende vis:

**Definition 1.3** (Varians)**.** Lad $X$ være en (diskret) stokastisk variabel med billedmængde $\{x_1, x_2, \ldots, x_n\}$ og fordelingsfunktion $P$. Så defineres *Variansen* for $X$ som

$$\begin{aligned}\mathrm{Var}(X) = \sigma^2 &= E[(X-\mu)^2] \\ &= (x_1-\mu)^2 \cdot P(X=x_1) + (x_2-\mu)^2 \cdot P(X=x_2) + \cdots + (x_n-\mu)^2 \cdot P(X=x_n) \\ &= \sum_{i=1}^{n} (x_i-\mu)^2 \cdot P(X=x_i).\end{aligned}$$

Tallet $\sigma = \sqrt{\mathrm{Var}(X)}$ kaldes for spredningen af $X$.

**Eksempel 1.4.** Lad os fortsætte eksemplet med terningekastet. Vi vil bestemme middelværdien $\mathbb{E}(X)$ for den stokastiske variabel $X$, der beskriver et terningekast. Dette bestemmes ifølge definitionen som

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3,5,$$

hvilket er tæt på det vi eksperimentelt havde bestemt gennemsnittet til at være. Variansen er så

$$\begin{aligned}\mathrm{Var}(X) &= (1-3,5)^2 \cdot \frac{1}{6} + (2-3,5)^2 \cdot \frac{1}{6} + (3-3,5)^2 \cdot \frac{1}{6} \\ &\quad + (4-3,5)^2 \cdot \frac{1}{6} + (5-3,5)^2 \cdot \frac{1}{6} + (6-3,5)^2 \cdot \frac{1}{6} \approx 2,92\end{aligned}$$

Spredningen er derfor

$$\sigma \approx \sqrt{2,92} \approx 1,71.$$

Det er svært at give en god fortolkning af spredningen ud over at det er den gennemsnitlige afstand et slag vil have fra middelværdien.

Hvis vi kender sandsynlighedsparametren $p$ og antalsparametren $n$ for en binomialfordelt stokastisk variabel, så er det muligt at bestemme middelværdien og variansen.

**Sætning 1.5.** *Lad $X$ være en binomialfordelt stokastisk variabel med antalsparameter $n$ og sandsynlighedsparameter $p$. Så er middelværdien $\mu$ og spredningen $\sigma$ for $X$ givet ved*

$$\mathbb{E}(X) = \mu = n \cdot p,$$

*og*

$$\sqrt{\mathrm{Var}(X)} = \sigma = \sqrt{n \cdot p \cdot (1-p)}$$

# Opgave 1

i) Slå plat og krone, og lad $X = 1$, hvis udfaldet er plat og $X = 0$ ellers. Hvad er middelværdien og variansen for $X$?

ii) Slå plat og krone to gange og lad $X$ være antallet af plat. Hvad er middelværdien og variansen af $X$?

iii) Kast en 12-sidet terning og lad $X$ være antallet af øjne. Hvad er middelværdien og variansen af $X$?

iv) Lad $X$ beskrive antallet af piger i en familie på 4. Hvad er middelværdien og variansen af $X$?

# Opgave 2

i) Lad $X$ være en Bernoulli-fordelt stokastisk variabel med sandsynlighedsparameter $p$. Hvad er middelværdien og variansen af $X$?

ii) Forklar med ord, hvorfor det giver god mening, at middelværdien for en binomialfordelt stokastisk variabel er $n \cdot p$.

# Opgave 3

i) Et lægemiddel helbreder 15% af behandlede personer. Lad $X$ beskrive antallet af helbredte personer, når vi prøver at helbrede 10000 personer. Hvad er middelværdien og variansen af $X$?

ii) Vi slår fem gange med en terning, og lader $X$ beskrive antallet af gange, vi slår mere end 4. Hvad er middelværdien og variansen for $X$?