

Betinget sandsynlighed og uafhængighed

Betinget sandsynlighed

Det er særligt når man begynder at introducere betingelser i sandsynligheder der, hvor man hurtigt kommer til at drage forkerte konklusioner. Fra Georg Mohr 2022 lød en opgave, at 150 1.g-ere havde tilkendegivet at 80% af dem var tilfredse med maden i kantinen, hvorimod 100 2.g-elever havde tilkendegivet, at 40% var tilfredse med maden. Vi skal bestemme sandsynligheden for, at en stemmeseddel med svaret ”ja” er afleveret af en 1.g’er. Vi bestemmer først de absolutte tal for ja og nej; $150 \cdot 0.8 = 120$ 1.g-elever har svaret ja og $100 \cdot 0.4 = 40$ 2.g-elever har svaret ja. Resultaterne kan ses af Tabel 1

	Tilfreds	Utilfreds	
1.g	120	30	150
2.g	40	60	100
	160	90	

Tabel 1: Resultater af spørgeskema

Vi skal bestemme sandsynligheden

$$P(\{\text{Svar er fra 1.g'er givet at der står ja på sedlen}\}).$$

I sandsynlighedsregning skriver vi typisk denne slags sandsynligheder på følgende vis. Lad A og B betegne følgende hændelser:

$$\begin{aligned} B &= \{\text{Der står ja på sedlen}\}, \\ A &= \{\text{Sedlen er fra en 1.g'er}\}. \end{aligned}$$

Vi ønsker så at bestemme sandsynligheden for udfaldet

$$P(A \mid B),$$

hvor den lodrette streg læses som ”hvor” eller ”betinget af”. Hvis sedlen er valgt tilfældigt kan vi nu udregne sandsynligheden. Vi har 160 forskellige sedler med ”ja” og af disse er 120 afleveret af en 1.g’er. Derfor er sandsynligheden

$$P(A \mid B) = \frac{120}{160} = \frac{3}{4}.$$

Vi vil nu definere *betinget sandsynlighed*.

Definition 1.1 (Betinget sandsynlighed). For to hændelser A og B , hvor $P(B) > 0$ defineres betingede sandsynlighed for A forudsat B som

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

Eksempel 1.2. Bruger vi definitionen til at udregne sandsynligheden fra spørgeskemaet om tilfredshed fås

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{120}{250}}{\frac{160}{250}} = \frac{3}{4}.$$

Med begrebet om betinget sandsynlighed kan vi også tale om uafhængighed af to hændelser. Vi vil sige, at to hændelser er uafhængige, hvis den ene hændelse ikke påvirker udfaldet af den anden.

Eksempel 1.3. Lad os betragte de to hændelser

$$\begin{aligned} A &= \{\text{Du slår krone med en mønt}\}, \\ B &= \{\text{Det er regnvejr}\}. \end{aligned}$$

På en tilfældig dag er sandsynligheden for, at det regner ca. 50%. Sandsynligheden for at du slår krone er

$$P(A) = 1/2.$$

Sandsynligheden for at du slår krone givet at det regner er

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2},$$

så sandsynligheden for, at du slår krone har altså ikke ændret sig af, at det regner.

I eksemplet udnyttede vi, at $P(A \cap B) = P(A) \cdot P(B)$. Dette er faktisk vores definition på, at to hændelser er uafhængige.

Definition 1.4 (Uafhængighed). To hændelser A og B siges at være uafhængige, hvis

$$P(A \cap B) = P(A) \cdot P(B).$$

Vi har følgende sætning, der fortæller os, at denne definition er ækvivalent med vores intuitive forståelse af uafhængighed.

Sætning 1.5. *To hændelser A og B er uafhængige hvis og kun hvis*

$$P(A \mid B) = P(A)$$

eller tilsvarende

$$P(B \mid A) = P(B).$$

Bevis. Opgave. ■

Bayes formel

En tilfældig udvalgt 50-årig kvinde screenes for brystkræft. Hun tester positiv og vil nu gerne vide, hvad sandsynligheden for, at hun rent faktisk er syg er. Vi får yderligere følgende information.

Prævalens:

Sandsynligheden for at en tilfældig valgt 50-årig kvinde har brystkræft er 1%,

Sensitivitet:

Sandsynligheden for at teste positiv givet kræft er 90%

Specifitet:

Sandsynligheden for at teste negativ givet ingen kræft er 91%

En række praktiserende læger blev stillet dette spørgsmål og skulle svare på, hvad de troede sandsynligheden for, at hun rent faktisk er syg er. De havde valgmulighederne

a) $\frac{9}{10}$

b) $\frac{8}{10}$

c) $\frac{1}{10}$

d) $\frac{1}{100}$

og over halvdelen svarede $\frac{9}{10}$. Vi vil bruge *Bayes formel* til at bestemme svaret.

Sætning 1.6 (Bayes formel). *For to hændelser A og B gælder der, at*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Bevis. Opgave. ■

Vi har altså hændelserne

$$A = \{\text{Kvinden har kræft}\},$$

$$B = \{\text{Testen er positiv}\},$$

og vi ønsker at bestemme

$$P(A | B).$$

Vi indsætter dette i Bayes formel og får

$$\begin{aligned} P(A | B) &= \frac{P(B | A)P(A)}{P(B)} \\ &= \frac{P(\{\text{Positiv}\} | \{\text{Syg}\})P(\{\text{Syg}\})}{P(\{\text{Positiv}\})} \\ &= \frac{P(\{\text{Positiv}\} | \{\text{Syg}\})P(\{\text{Syg}\})}{P(\{\text{Positiv}\} | \{\text{Rask}\})P(\{\text{Rask}\}) + P(\{\text{Positiv}\} | \{\text{Syg}\})P(\{\text{Syg}\})} \\ &= \frac{0.9 \cdot 0.01}{0.09 \cdot 0.99 + 0.9 \cdot 0.01} \\ &\approx 0.09 \end{aligned}$$

Sandsynligheden for, at hun rent faktisk er syg er derfor under 10 procent!

Opgave 1

- i) For to hændelser A og B med $P(A) = 0.5$, $P(B) = 0.4$ og $P(A \cap B) = 0.2$ bestem da $P(A | B)$ og $P(B | A)$.

Opgave 2

Vi betragter følgende hændelser fra eksemplet med spørgeskemaet.

$$A = \{\text{Der står ja på sedlen}\},$$

$$B = \{\text{Der står nej på sedlen}\},$$

$$C = \{\text{En 1.g'er har svaret på sedlen}\},$$

$$D = \{\text{En 2.g'er har svaret på sedlen}\}.$$

- i) Bestem $P(A | C)$.
- ii) Bestem $P(C | A)$.
- iii) Bestem $P(D)$.
- iv) Bestem $P(A | D)$.
- v) Bestem $P(D | A)$.

Opgave 3

Vi trækker et kort fra et kortspil uden jokere og betragter hændelserne

$$A = \{\text{Kortet er et hjerte}\},$$
$$B = \{\text{Kortet er et billedkort}\}.$$

- i) Bestem sandsynlighederne $P(A)$ og $P(B)$.
- ii) Bestem sandsynligheden $P(A \cap B)$.
- iii) Bestem sandsynligheden for $P(A \mid B)$ og $P(B \mid A)$ og argumentér for, at A og B er uafhængige hændelser

Opgave 4

- i) Hvis du ved, at et par med to børn har mindst én datter, hvad er så sandsynligheden for, at de har to døtre?
- ii) Du trækker et kort fra et kortspil hvor spar 5 mangler. Betragt hændelserne

$$A = \{\text{Dit kort er et es}\},$$
$$B = \{\text{Dit kort er et klør}\}.$$

Bestem sandsynlighederne $P(A)$ og $P(B)$ og $P(A \cap B)$. Er de to hændelser uafhængige?

Opgave 5

Bevis Sætning 1.5.

Opgave 6

En Covid-antigentest har en sensitivitet på 65.3% og en specificitet på 99.5%. Lad os sige, at 0.2% af Danmarks befolkning har Covid lige nu. Hvad er så sandsynligheden for, at en tilfældig person, der tester positiv med en antigen-test rent faktisk er syg?

Opgave 7

En test for cannabis har en sensitivitet på 90% og en specificitet på 80%. Lad os sige, at 5% af befolkningen ryger cannabis jævnligt. Hvad er sandsynligheden for, at en person, der tester positiv rent faktisk ryger cannabis?

Opgave 8

Bevis Bayes formel.