Oplossing bij rode draad: het verraderlijke virus

De laatste jaren bracht Covid-19 heel wat onzekerheid met zich mee. Om deze onzekerheid te beheersen, werden er modellen opgesteld, die op basis van voorgaande en beschikbare informatie, voorspelling kunnen doen over de nabije toekomst. Een belangrijke eerste stap in deze modellen is het verzamelen van data. Met deze data, worden er kansen berekend en gepubliceerd voor het risico op besmetting, symptomen, hospitalisatie en overlijden. Het is belangrijk om deze gegeven kansen en percentages correct te interpreteren en te verwerken. Zo vermijden we misverstanden en foutieve communicatie.

We beschouwen Ω de verzameling van alle mogelijke uitkomsten van een experiment of meting. Elke deelverzameling van Ω is een gebeurtenis. Indien A en B gebeurtenissen zijn, dan noteren we:

A^c : De uitkomsten die *niet* bevat zijn door A.

$A \cap B$: De uitkomsten die bevat zijn door A *en* door B.

$A \cup B$: De uitkomsten die bevat zijn door A *of* door B.

Aan elke gebeurtenis A kennen we de waarde $P(A)$ toe tussen 0 en 1. Deze waarde drukt de kans uit dat de gebeurtenis A zal plaatsvinden. De afbeelding $P(\cdot)$ is een *kansmaat* indien de volgende eigenschappen voldaan zijn:

1. $P(\emptyset) = 0$
2. $P(\Omega) = 1$
3. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

De eerste voorwaarde drukt uit dat de kans dat er 'niets' gebeurt, gelijk is aan nul. De tweede regel zegt dat de kans dat er 'iets' gebeurt, gelijk is aan 1. De laatste voorwaarde zegt dat indien de gebeurtenissen A en B geen mogelijkheden gemeenschappelijk hebben, dan is de kans op A of B precies gelijk aan de som van de kansen.

Een bijkomend belangrijk concept is de *voorwaardelijke kans* $P(B|A)$. Deze drukt de kans uit op een gebeurtenis B gegeven dat de gebeurtenis A reeds heeft plaatsgevonden. Deze wordt als volgt gedefinieerd,

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

De afbeelding $P(\cdot|A)$ is opnieuw een kansmaat. We beschouwen een voorbeeld om de notatie en de concepten toe te lichten.

Voorbeeld

We hebben een zak met 15 knikkers. Elke knikker heeft een maat (groot of klein) en een kleur (rood of wit). Er zijn 10 grote knikkers en van deze grote knikkers zijn er 8 wit gekleurd. Indien we nu een willekeurige knikker nemen uit de zak, noteren we de gebeurtenis dat de knikker groot is met G en de gebeurtenis dat ze wit gekleurd is door W. Uit de gegevens halen we meteen dat $P(G) = 2/3$ en dat $P(W|G) = 4/5$. Uit de formule voor de voorwaardelijke kans leiden we af dat de kans op het trekken van een knikker die zowel groot als wit is, gegeven is door $P(G \cap W) = 8/15$. Verder is de gebeurtenis G^c overeenkomstig met het nemen van een kleine knikker en W^c overeenkomstig met het nemen van een rode knikker. Zo weten we meteen dat $P(G^c) = 1/3$ en $P(W^c|G) = 1/5$.

Het voorbeeld toont aan dat je met beperkte informatie de kans op verschillende gebeurtenissen kan afleiden door gebruik te maken van de eigenschappen van een kansmaat. Nu is het aan jullie.

Opdracht A

Vind en bewijs een formule voor $P(B)$ en $P(A|B)$ indien je enkel over de volgende gegevens beschikt: $P(A)$, $P(B|A)$ en $P(B|A^c)$.

Een oplossing

We vinden volgende formules, de zogenaamde regel van Bayes:

1.

$$
\begin{aligned}
P(B) &= P((A \cup A^c) \cap B) = P((A \cap B) \cup (A^c \cap B)) \\
&= P(A \cap B) + P(A^c \cap B) = \frac{P(A \cap B)P(A)}{P(A)} + \frac{P(A^c \cap B)P(A^c)}{P(A^c)} \\
&= P(B|A)P(A) + P(B|A^c)P(A^c) = P(B|A)P(A) + P(B|A^c)(1 - P(A))
\end{aligned}
$$

2.

$$
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)(1 - P(A))}
$$

Deze formule werd voor het eerste beschreven door Thomas Bayes en met hulp van Richard Price gepubliceerd in 1763. Het vormt het basisbeginsel van de Bayesiaanse statistiek. Het laat ons toe om de initiële onzekerheid van een hypothese statistisch bij te stellen op basis van observaties.

Vanaf nu baseren we ons op een fictief onderzoek bij mensen die recent besmet geraakten met het fictieve virus X-07. Hieruit bleek dat 15% van de niet-gevaccineerde mensen symptomen ontwikkelde, terwijl dit slechts 5% bedroeg voor de gevaccineerde personen. Tijdens dit onderzoek was 90% van de deelnemers reeds gevaccineerd.

Gebruik de volgende notatie voor je gebeurtenissen:

V : De persoon is gevaccineerd.

S : De persoon ontwikkelt symptomen.

Opdracht

Indien we de percentages uit het onderzoek doortrekken naar de volledige bevolking, beantwoord dan de volgende vragen:

1. Wat is de kans dat een willekeurig gekozen X-07 besmet persoon symptomen ontwikkelt?
2. Wat is de kans dat een willekeurig gekozen X-07 besmet persoon met symptomen, gevaccineerd is?
3. Vergelijk de kans uit (2) met de omgekeerde conditionele kans, i.e. de kans dat een willekeurig gekozen X-07 besmet en gevaccineerd persoon, symptomen ontwikkelt? Voor welke hypothetische vaccinatiegraad zouden deze gelijk zijn?

Een oplossing

- $P(S) = P(S|V)P(V) + P(S|V^c)P(V^c) = \frac{5}{100}\frac{9}{10} + \frac{15}{100}\frac{1}{10} = \frac{3}{50}$
- $P(V|S) = \frac{P(S|V)P(V)}{P(S|V)P(V)+P(S|V^c)P(V^c)} = \frac{3}{4}$
- $P(V|S) = \frac{1}{20} = P(S|V) = \frac{\frac{5}{100}P(V)}{\frac{5}{100}P(V)+\frac{15}{100}(1-P(V))} \Leftrightarrow P(V) = \frac{3}{22}$

Het vervolg van het onderzoek beperkte zich tot de groep van deelnemers met X-07 symptomen. Binnen deze groep werd 30% opgenomen in het ziekenhuis. Bovendien was 50% van de opgenomen deelnemers met X-07 symptomen gevaccineerd. Gebruik de volgende notatie voor je gebeurtenissen:

V : Een persoon is gevaccineerd.

S : Een persoon ontwikkelt symptomen.

H : Een persoon wordt gehospitaliseerd.

Opdracht

We zijn geïnteresseerd in de efficiëntie van de vaccins met betrekking tot de kans op hospitalisatie bij mensen met reeds X-07 symptomen. Hiervoor vragen we om de volgende verhoudingen te berekenen:

1.
$$\frac{P(H^c|V\cap S)}{P(H|V\cap S)}$$

2.
$$\frac{P(H^c|V^c\cap S)}{P(H|V^c\cap S)}$$

3.
$$\frac{P(H|V^c\cap S)}{P(H|V\cap S)}$$

Geef een interpretatie aan deze verhoudingen.

Een oplossing

1.
$$\begin{aligned}\frac{P(H^c|V\cap S)}{P(H|V\cap S)} &= \frac{1}{P(H|V\cap S)}-1\\ &= \frac{P(V\cap S)}{P(V\cap H\cap S)}-1\\ &= \frac{P(V|S)P(S)}{P(V|H\cap S)P(H|S)P(S)}-1\\ &= 4\end{aligned}$$

2.
$$\begin{aligned}\frac{P(H^c|V^c\cap S)}{P(H|V^c\cap S)} &= \frac{P(V^c|S)P(S)}{P(V^c|H\cap S)P(H|S)P(S)}-1\\ &= \frac{2}{3}\end{aligned}$$

3.
$$\begin{aligned}\frac{P(H|V^c\cap S)}{P(H|V\cap S)} &= \frac{P(V^c|H\cap S)P(V|S)}{P(V|H\cap S)P(V^c|S)}\\ &= 3\end{aligned}$$

Interpretatie:

1. Voor elke gehospitaliseerde persoon binnen de groep van gevaccineerden met symptomen, zijn er gemiddeld 4 niet-gehospitaliseerde personen binnen deze groep.

2. Voor elke drie gehospitaliseerde personen binnen de groep van gevaccineerden met symptomen, zijn er binnen deze groep gemiddeld slechts twee niet-gehospitaliseerde personen.

3. De kans om gehospitaliseerd te worden indien je gevaccineerd bent en symptomen hebt, is 3 keer kleiner dan wanneer je niet-gevaccineerd bent met symptomen.