Aufgabe 8.3: Formel von Bayes (7 Punkte)

Eine bestimmte Krebsart kommt bei ca. 0,5 % der Bevölkerung vor. Ein Test zur Auffindung der Krebserkrankung führt bei 99 % der Kranken zu einer Reaktion, aber auch bei 2 % der Gesunden.

(i) Wie groß ist die Wahrscheinlichkeit, dass eine Person, bei der die Reaktion eintritt, die Krebserkrankung wirklich hat?

(ii) Wie groß sind die Sensitivität und die Spezifität des Tests?

Hinweis: Die Definitionen der beiden Begriffe finden Sie im Internet oder im Buch, z.B.: „Sachs, L. und Hedderich, J. (2016): Angewandte Statistik, Methodensammlung mit R. Berlin, Heidelberg: Springer Verlag.“, das über die Bibliothek als E-Book ausleihbar ist.

(iii) Weisen Sie allgemein nach, dass die Chancen (engl.: odds) für eine Erkrankung bei positivem Test um den Faktor

$$\frac{\text{Sensitivität}}{1-\text{Spezifität}} = \frac{\text{richtig-positiv Rate}}{\text{falsch-positiv Rate}}$$

größer sind als vor dem Test.

Hinweis: Mit „odds“ bezeichnet man das Verhältnis zwischen der Wahrscheinlichkeit für „krank“ (=„*Treffer*“) und der Wahrscheinlichkeit für „gesund“ (=„*kein Treffer*“). Der Begriff wird auch bei Sportwetten verwendet.

(iv) Prüfen Sie die Gültigkeit der Formel aus Teil (iii) anhand der Zahlen aus der Aufgabe und dem Ergebnis aus Teil (i).

Lösung:

(i) $\Omega = \{1, \ldots, N\}$ Bevölkerung
$\mathcal{F} = \mathcal{P}(\Omega)$
$P =$ Gleichverteilung auf $\{1, \ldots, N\} = U_{\{1,\ldots,N\}}$

Geg.: $P(K) = 0{,}5\ \%$ $\quad K := \{\text{Menschen mit Krebs}\}$
$P(PT \mid K) = 99\ \%$ $\quad PT := \{\text{Menschen, bei denen Test positiv ist}\}$
$P(PT \mid \overline{K}) = 2\ \%$

Ges.: $P(K \mid PT)$

Lsg.:

$$\begin{aligned} P(K \mid PT) &\overset{\text{Bayes}}{=} \frac{P(PT \mid K) \cdot P(K)}{P(PT \mid K) \cdot P(K) + P(PT \mid \overline{K}) \cdot P(\overline{K})} \\ &= \frac{\frac{99}{100} \cdot \frac{5}{1000}}{\frac{99}{100} \cdot \frac{5}{1000} + \frac{2}{100} \cdot \frac{995}{1000}} \\ &= \frac{99 \cdot 5}{99 \cdot 5 + 2 \cdot 995} \approx \frac{500}{500 + 2000} = 20\ \% \end{aligned}$$

3

Patienten mit positiven Test haben „nur“ zu 20 % Krebs.

Alternative mit kartesischem Produkt:

$\Omega = \underbrace{\{K, G\}}_{=:\ \Omega_1} \times \underbrace{\{PT, NT\}}_{=:\ \Omega_2}$ mit:

K:	krank
G:	gesund
PT:	pos. Test
NT:	neg. Test

$\mathcal{F} = \mathcal{P}(\Omega)$
$P(\{K\} \times \Omega_2) = 0{,}5\ \%$
$P(\Omega_1 \times \{PT\} \mid \{K\} \times \Omega_2) = 99\ \%$
$P(\Omega_1 \times \{PT\} \mid \{g\} \times \Omega_2) = 2\ \%$

Gesucht: $P(\{K\} \times \Omega_2 \mid \Omega_1 \times \{PT\})$

$$
\begin{aligned}
&P(\{K\} \times \Omega_2 \mid \Omega_1 \times \{PT\}) = \\
&\overset{\text{Bayes}}{\overset{\downarrow}{=}} \frac{P(\Omega_1 \times \{PT\} \mid \{K\} \times \Omega_2) \cdot P(K \times \Omega_2)}{P(\Omega_1 \times \{PT\} \mid \{K\} \times \Omega_2) \cdot P(K \times \Omega_2) + P(\Omega_1 \times \{PT\} \mid \{G\} \times \Omega_2) \cdot P(G \times \Omega_2)} \\
&= \frac{99\ \% \cdot 0{,}5\ \%}{99\ \% \cdot 0{,}5\ \% + 2\ \% \cdot 99{,}5\ \%} \approx 20\ \%
\end{aligned}
$$

Bemerkung: Beim Sportwettenanbieter Oddset werden die Quoten q je Ergebnis angegeben, das ist der Betrag, den man für einen Euro Einsatz erhält. Bei einem fairen Spiel gilt $q = \frac{1}{p} \Leftrightarrow p = \frac{1}{q}$ und $Gewinn = q - 1 = \frac{1-p}{p} = \frac{1}{odds}$.
Beispiel: SC Freiburg - FCA

Ergebnis	Quote	p	Gewinn	odds
1	1,97	0,51	0,97	1,03
0	3,7	0,27	2,7	0,37
2	3,6	0,28	2,6	0,38

Da die Summe der Wahrscheinlichkeiten 0,51+0,27+0,28=1,06 eins übersteigt, ist das Spiel nicht fair, sondern zugunsten von Oddset verzerrt.

(ii) Sensitivität („richtig positiv Rate“) bezeichnet den Anteil der test-positiven Personen unter allen Erkrankten einer Stichprobe, d. h. die Wahrscheinlichkeit, mit dem Test die Kranken auch als krank zu identifizieren.
In unserem Beispiel gilt also: Sensitivität = 99%. 1/2

Spezifität („richtig-negative Rate eines Tests) beschreibt den Anteil der Test-negativen Personen unter allen Nicht-Erkrankten einer Stichprobe, d. h. die Wahrscheinlichkeit, mit dem Test Nicht-Erkrankte auch als nicht-erkrankt zu identifizieren.
In unserem Beispiel gilt also: Spezifität = 98%. 1/2

(iii) Z.z.:

$$\text{„}odds\ nach\ Test\text{“} = \frac{\text{Sensitivität}}{1 - \text{Spezifität}} \cdot \text{„}odds\ vor\ Test\text{“}$$

Mit den Bezeichnung aus Teil (i) ist daher z.z.:

$$\underbrace{\frac{\mathrm{P}(K \mid PT)}{1 - \mathrm{P}(K \mid PT)}}_{odds\ nach\ Test} = \frac{\mathrm{P}(PT \mid K)}{1 - \mathrm{P}\left(\overline{PT} \mid \overline{K}\right)} \cdot \underbrace{\frac{\mathrm{P}(K)}{1 - \mathrm{P}(K)}}_{odds\ vor\ Test}$$

Dies gilt, da:

$$
\begin{aligned}
\frac{\mathrm{P}(K \mid PT)}{1 - \mathrm{P}(K \mid PT)} &= \frac{\mathrm{P}(K \mid PT)}{\mathrm{P}\left(\overline{K} \mid PT\right)} \\
&\overset{Bayes}{=} \frac{\mathrm{P}(PT \mid K) \cdot \mathrm{P}(K) \cdot \cancel{\mathrm{P}}(PT)}{\cancel{\mathrm{P}}(PT) \cdot \mathrm{P}(PT \mid \overline{K}) \cdot \mathrm{P}(\overline{K})} \\
&= \frac{\mathrm{P}(PT \mid K)}{1 - \mathrm{P}\left(\overline{PT} \mid \overline{K}\right)} \cdot \frac{\mathrm{P}(K)}{1 - \mathrm{P}(K)}
\end{aligned}
$$

2

(iv) Laut Angabe gilt:

$$\frac{\text{richtig positiv Rate}}{\text{falsch positiv Rate}} = \frac{0.99}{0.02} \approx 50$$

Tatsächlich erhöhen sich die Chance auf eine Krebserkrankung bei positivem Test von $\frac{0{,}5\%}{99{,}5\%} \approx \frac{1}{200}$ auf $\frac{20\%}{80\%} = \frac{1}{4}$ (siehe Teil (i)), also ungefähr um den Faktor 50. 1