Applied Statistics for Data Science

Serie 6

Aufgabe 6.1

In einem Land der Dritten Welt leiden 1 % der Menschen an einer bestimmten Infektionskrankheit. Ein Test zeigt die Krankheit bei den tatsächlich Erkrankten zu 98 % korrekt an. Leider zeigt der Test auch 3 % der Gesunden als erkrankt an.

Wir bezeichnen mit K eine kranke Person und T eine positiv getestete Person.

a) Interpretieren (nicht berechnen!) Sie die Wahrscheinlichkeiten

$$P(K), \qquad P(\overline{T}), \qquad P(K \mid T), \qquad P(T \mid K), \qquad P(\overline{T} \mid \overline{K})$$

b) Bezeichnen Sie die Wahrscheinlichkeit, die in der Aufgabe gegeben sind, wie in a).

c) Berechnen Sie $P(\overline{K})$.

d) Mit welcher Wahrscheinlichkeit zeigt der Test bei einer zufällig ausgewählten Person ein positives Ergebnis?

   Verwenden Sie das Gesetz der totalen Wahrscheinlichkeit.

e) Mit welcher Wahrscheinlichkeit ist eine positiv getestete Person auch tatsächlich krank? Kommentieren Sie das Ergebnis.

   Verwenden Sie das Bayes Theorem.

f) Mit welcher Wahrscheinlichkeit ist eine als negativ getestete Person gesund? Kommentieren Sie das Ergebnis.

   Verwenden Sie das Bayes Theorem.

Aufgabe 6.2

Bei einer Sportveranstaltung wird ein Dopingtest durchgeführt. Wenn ein Sportler gedopt hat, dann fällt der Test zu 99 % positiv aus.

Hat ein Sportler aber nicht gedopt, zeigt der Test trotzdem zu 5 % ein positives Ergebnis an. Aus Erfahrung weiss man das 20 % der Sportler gedopt sind.

a) Wie gross ist die Wahrscheinlichkeit, dass eine Dopingprobe positiv ausfällt.

b) Wie gross ist die Wahrscheinlichkeit, dass der Test negativ ausfällt, obwohl der Sportler gedopt hat?

c) Wie gross ist die Wahrscheinlichkeit, dass ein Sportler gedopt hat, falls seine Dopingprobe negativ ausgefallen ist.

Aufgabe 6.3

Ein Lügendetektortest wird routinemässig bei Angestellten durchgeführt, die in sensitiven Positionen arbeiten. Bezeichnen wir nun mit $+$ das Ereignis, dass der Test positiv ist, also dass der Lügendetektor angibt, dass der Angestellte gelogen habe. Mit W wird das Ereignis bezeichnet, dass der Angestellte die Wahrheit sagte und mit L, dass der Angestellte gelogen hat.

Aus Untersuchungen über Lügendetektoren wissen wir, dass

$$P(+|L) = 0.88 \quad \text{und} \quad P(-|W) = 0.86$$

Des weiteren wissen wir, dass folgendes gilt

$$P(W) = 0.99$$

a) Interpretieren Sie die Wahrscheinlichkeiten $P(W)$ und $P(+|L)$.

b) Bei einer Person zeigt der Detektor an, dass gelogen wurde. Mit welcher Wahrscheinlichkeit hat diese Person effektiv gelogen?

c) Interpretieren Sie das Resultat aus b) in 2-3 Sätzen. Für wie aussagekräftig halten Sie den Lügendetektor.

Aufgabe 6.4

Der Serumtest untersucht schwangere Frauen auf Babys mit Down-Syndrom. Der Test ist ein sehr guter, aber nicht perfekter Test. Ungefähr 1 % der Babys haben das Down-Syndrom. Wenn das Baby das Down-Syndrom hat, besteht eine 90-prozentige Wahrscheinlichkeit, dass das Ergebnis positiv ausfällt. Wenn das Baby nicht betroffen ist, besteht immer noch eine 1 % Chance, dass das Ergebnis positiv ist. Eine schwangere Frau wurde getestet und das Ergebnis ist positiv. Wie gross ist die Wahrscheinlichkeit, dass ihr Baby tatsächlich das Down-Syndrom hat?

Aufgabe 6.5

Die Rauchsensoren in einer Fabrik melden ein Feuer mit Wahrscheinlichkeit 0.95. An einem Tag ohne Brand geben sie mit Wahrscheinlichkeit 0.01 falschen Alarm. Pro Jahr rechnet man mit einem Brand.

a) Die Alarmanlage meldet Feuer. Mit welcher Wahrscheinlichkeit brennt es tatsächlich?

b) In einer Nacht ist es ruhig (kein Alarm). Mit welcher Wahrscheinlichkeit brennt es tatsächlich nicht?

Applied Statistics for Data Science

Musterlösungen zu Serie 6

Lösung 6.1

a)
- $P(K)$: Wahrscheinlichkeit, dass eine getestete Person wirklich krank ist.
- $P(\overline{T})$: Wahrscheinlichkeit, dass eine getestete Person negativ getestet wurde (Krankheit wird nicht angezeigt).
- $P(K \mid T)$: Wahrscheinlichkeit, dass eine positiv getestete Person auch wirklich krank ist.

  Oder: Eine Person ist positiv getestet. Wahrscheinlichkeit, dass sie wirklich krank ist.
- $P(T \mid K)$: Wahrscheinlichkeit, dass eine kranke Person positiv getestet wird.

  Oder: Eine Person ist krank. Wahrscheinlichkeit, dass sie auch positiv getestet wird.
- $P(\overline{T} \mid \overline{K})$: Wahrscheinlichkeit, dass eine gesunde Person negativ getestet wird.

  Oder: Eine Person ist gesund. Wahrscheinlichkeit, dass sie auch negativ getestet wird.

b)
- In einem Land der Dritten Welt leiden 1 % der Menschen an einer bestimmten Infektionskrankheit:
$$P(K) = 0.01$$
- Ein Test zeigt die Krankheit bei den tatsächlich Erkrankten zu 98 % korrekt an:
$$P(T \mid K) = 0.98$$
- Leider zeigt der Test auch 3 % der Gesunden als erkrankt an:
$$P(T \mid \overline{K}) = 0.03$$

c) Es gilt
$$P(\overline{K}) = 1 - P(K) = 1 - 0.01 = 0.99$$

d) Gesucht: $P(T)$

$$\begin{aligned} P(T) &= P(T \mid K) \cdot P(K) + P(T \mid \overline{K}) \cdot P(\overline{K}) \\ &= P(T \mid K) \cdot P(K) + P(T \mid \overline{K}) \cdot (1 - P(K)) \\ &= 0.98 \cdot 0.01 + 0.03 \cdot 0.99 \\ &= 0.0395 \end{aligned}$$

Das heisst, 3.95 % der Getesteten werden positiv getestet.

e) Gesucht: $P(K \mid T)$

$$\begin{aligned} P(K \mid T) &= \frac{P(T \mid K) \cdot P(K)}{P(T)} \\ &= \frac{0.98 \cdot 0.01}{0.0395} \\ &= 0.2481 \end{aligned}$$

Das heisst, nur etwa 25 % aller positiv Getesteten ist effektiv auch krank.

f) Gesucht: $P(\overline{K} \mid \overline{T})$:

$$\begin{aligned} P(\overline{K} \mid \overline{T}) &= \frac{P(\overline{T} \mid \overline{K}) \cdot P(\overline{K})}{P(\overline{T})} \\ &= \frac{(1 - P(T \mid \overline{K})) \cdot P(\overline{K})}{P(\overline{T})} \\ &= \frac{(1 - 0.03) \cdot 0.99}{1 - 0.0395} \\ &= 0.999792 \end{aligned}$$

Das heisst, wenn der Test negativ ist, so sind wir sehr sicher, dass wir auch gesund sind. Während ein positiver Test nicht sehr viel aussagt, sagt ein negativer Test sehr viel aus.

Lösung 6.2

Bezeichnungen:

- D: gedopt
- T: positiv getestet

Es gilt aus der Aufgabenstellung

$$P(D) = 0.2, \qquad P(T \mid D) = 0.99, \qquad P(T \mid \overline{D}) = 0.05$$

a) Gesucht: $P(T)$.

Wir verwenden das Gesetz der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(T) &= P(T \mid D) \cdot P(D) + P(T \mid \overline{D}) \cdot P(\overline{D}) \\ &= P(T \mid D) \cdot P(D) + P(T \mid \overline{D}) \cdot (1 - P(D)) \\ &= 0.99 \cdot 0.2 + 0.05 \cdot 0.8 \\ &= 0.238 \end{aligned}$$

Das heisst, 23.8 % der getesteten Personen werden positiv getestet.

b) Gesucht: $P(\overline{T} \mid D)$

$$P(\overline{T} \mid D) = 1 - P(T \mid D) = 1 - 0.99 = 0.01$$

c) Gesucht: $P(D \mid \overline{T})$

Wir verwenden das Theorem von Bayes:

$$\begin{aligned} P(D \mid \overline{T}) &= \frac{P(\overline{T} \mid D) \cdot P(D)}{P(\overline{T})} \\ &= \frac{0.01 \cdot 0.2}{1 - 0.238} \\ &= 0.00262 \end{aligned}$$

Nur 0.262 % aller negativ Getesteten sind auch gedopt. Ein negativer Test ist also recht aussagekräftig.

Lösung 6.3

a) $P(W)$: Wahrscheinlichkeit, dass ein zufällig ausgewählter Angestellter die Wahrheit sagt.
$P(+|L)$: Wahrscheinlichkeit, dass bei einem zufällig ausgewählten angestellten Lügner der Test dies auch anzeigt.

b) Gesucht: $P(L|+)$

Es gilt

$$\begin{aligned}
P(L|+) &= \frac{P(+|L)P(L)}{P(+|L)P(L) + P(+|W)P(W)} \\
&= \frac{P(+|L)(1-P(W))}{P(+|L)(1-P(W)) + (1-P(-|W))P(W)} \\
&= \frac{0.88 \cdot (1-0.99)}{0.88 \cdot (1-0.99) + (1-0.86) \cdot 0.99} \\
&= 0.0597
\end{aligned}$$

c) Zeigt der Test also an, dass der Angestellte gelogen hat, so ist er bloss mit einer Wahrscheinlichkeit von 6 % auch wirklich ein Lügner. Oder in 94 % der positiv ausgefallenen Tests, sagen die (angeblich lügenden) Angestellten die Wahrheit.

Der Test zeigt also zu einem sehr hohen Prozentsatz ein falsches Ergebnis an und ist somit wertlos.

Das ist auch der Grund, warum Lügendetektortest vor Gericht nicht erlaubt sind.

Lösung 6.4

Nun, bezeichnen wir D als Down-Syndrom und $+$ mit einem positiven Testergebnis. Gegeben sind die folgenden Wahrscheinlichkeiten:

$$P(D) = 0,01, \qquad P(+ \mid D) = 0,9, \qquad P(+ \mid \overline{D}) = 0.01$$

Unter Anwendung des Bayes'schen Theorems und des Gesetzes der totalen Wahrscheinlichkeit folgt

$$\begin{aligned}
P(D \mid +) &= \frac{P(+ \mid D) \cdot P(D)}{P(+ \mid D) \cdot P(D) + P(+ \mid \overline{D}) \cdot P(\overline{D})} \\
&= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.01 \cdot 0.99} \\
&= 0.4761905
\end{aligned}$$

Bei einem positiven Testergebnis besteht eine 48 % Wahrscheinlichkeit, dass das Baby das Down-Syndrom hat.

Lösung 6.5

Wir bezeichnen mit:

$F :=$ Ereignis, dass Feuer ausbricht
$A :=$ Ereignis, dass der Alarm losgeht

Die Wahrscheinlichkeit, dass ein Feuer ausbricht, ergibt sich zu

$$P(F) = \frac{1}{365}$$

Die Wahrscheinlichkeit, dass der Alarm losgeht, gegeben es bricht ein Feuer aus, ist

$$P(A \mid F) = 0.95$$

Die Wahrscheinlichkeit, dass es einen Alarm gibt, gegeben dass kein Feuer ausgebrochen ist, lautet

$$P(A \mid \overline{F}) = 0.01$$

a) Die Wahrscheinlichkeit, dass ein Feuer ausgebrochen ist, gegeben es gab einen Alarm, ist

$$P(F \mid A) = \frac{P(A \mid F) \cdot P(F)}{P(A)}$$

Die Wahrscheinlichkeit, dass es einen Alarm gibt, lässt sich mit dem Gesetz der totalen Wahrscheinlichkeit ausdrücken:

$$P(A) = P(A \mid F) \cdot P(F) + P(A \mid \overline{F}) \cdot P(\overline{F})$$

Also ergibt sich für

$$\begin{aligned} P(F \mid A) &= \frac{P(A \mid F) \cdot P(F)}{P(A \mid F) \cdot P(F) + P(A \mid \overline{F}) \cdot P(\overline{F})} \\ &= \frac{0.95 \cdot \frac{1}{365}}{0.95 \cdot \frac{1}{365} + 0.01 \cdot (1 - \frac{1}{365})} \\ &= 0.207 \end{aligned}$$

Also nur zu etwa 20 % der Alarme bedeuten auch ein Feuer. Das heisst aber auch, dass 4/5 der Alarme Fehlalarme sind.

b) Die Wahrscheinlichkeit, dass kein Feuer ausgebrochen ist, gegeben es gab keinen Alarm, ist

$$P(\overline{F} \mid \overline{A})$$

Nun verwenden das Bayes Theorem:

$$P(\overline{F} \mid \overline{A}) = \frac{P(\overline{A} \mid \overline{F}) \cdot P(\overline{F})}{P(\overline{A})}$$

auf der rechten Seite der Gleichung haben wir 3 unbekannte Terme, die wir allerdings aus den bekannten berechnen können:

- Für $P(\overline{A} \mid \overline{F})$ gilt

$$\begin{aligned} P(\overline{A} \mid \overline{F}) &= 1 - P(A \mid \overline{F}) \\ &= 1 - 0.01 \\ &= 0.99 \end{aligned}$$

- Für $P(\overline{F})$ gilt

$$\begin{aligned} P(\overline{F}) &= 1 - P(F) \\ &= 1 - \frac{1}{365} \\ &= \frac{364}{365} \end{aligned}$$

- Für $P(\overline{A})$ gilt

$$P(\overline{A}) = 1 - P(A)$$

  Die Wahrscheinlichkeit $P(A)$ können wir mit der totalen Wahrscheinlkeit:

$$\begin{aligned} P(A) &= P(A \mid F)P(F) + P(A \mid \overline{F})P(\overline{F}) \\ &= 0.95 \cdot \tfrac{1}{365} + 0.01 \cdot \tfrac{364}{365} \\ &= 0.01257534 \end{aligned}$$

Also finden wir

$$\begin{aligned} P(\overline{F} \mid \overline{A}) &= \frac{P(\overline{A} \mid \overline{F}) \cdot P(\overline{F})}{P(\overline{A})} \\ &= \frac{\big(1 - P(A \mid \overline{F})\big) \cdot P(\overline{F})}{1 - P(A)} \\ &= \frac{(1 - 0.01) \cdot \frac{364}{365}}{1 - (0.95 \cdot \frac{1}{365} + 0.01 \cdot (1 - \frac{1}{365}))} \\ &= 0.999 \end{aligned}$$

Es ist also praktisch sicher, dass es kein Feuer gibt, wenn kein Alarm losgeht.