Wahrscheinlichkeiten

Ideen und Konzepte der Informatik

Kurt Mehlhorn

max planck institut
informatik

SIC Saarland Informatics Campus

Wahrscheinlichkeitsverteilungen

- Sei S eine endliche Menge (Ereignisraum). Eine Wahrscheinlichkeitsverteilung auf S ist eine Funktion p, die jedem Element von e eine nicht-negative Zahl p_e zuordnet (man schreibt gern p_e statt $p(e)$), so dass $\sum_{e \in S} p_e = 1$.
- Werfen von zwei Würfeln: Dann besteht S aus allen Paaren (x, y) mit $1 \leq x, y \leq 6$. Für jedes Paar e ist $p_e = 1/36$.
- Eine Teilmenge $E \subseteq S$ heißt Ereignis. Die Wahrscheinlichkeit eines Ereignisses ist definiert als $p(E) = \sum_{e \in E} p_e$.
- Beim Würfeln:
  - Augensumme = 7: Für $E_1 = \{(x, y); x + y = 7\} = \{(1,6), (2,5), \ldots, (6,1)\}$ ist $p_{E_1} = 6/36 = 1/6$ und für
  - Augensumme ≥ 8: Für $E_2 = \{(x, y); x + y \geq 8\}$ gilt $p_{E_2} = 15/36 = 5/12$.

Bedingte Wahrscheinlichkeiten und Bayessche Regel

Sei E und F Ereignisse. Dann ist die die bedingte Wahrscheinlichkeit von E unter der Voraussetzung F definiert als

$$\mathrm{prob}(E|F) = \frac{\mathrm{prob}(E \cap F)}{\mathrm{prob}(F)}.$$

Bayessche Regel (Thomas Bayes): Sie erlaubt die Berechung von $\mathrm{prob}(E|F)$ aus $\mathrm{prob}(F|E)$ nach folgender Formel:

$$\mathrm{prob}(E|F) = \frac{\mathrm{prob}(F|E) \cdot \mathrm{prob}(E)}{\mathrm{prob}(F|E)\mathrm{prob}(E) + \mathrm{prob}(F|\neg E)\mathrm{prob}(\neg E)}$$

Dabei steht $\neg E$ für das Ereignis $S \setminus E$, das Gegenereignis zu E.

Nach Definition $\mathrm{prob}(E \cap F) = \mathrm{prob}(E|F)\mathrm{prob}(F)$. Dann auch

$$\mathrm{prob}(E \cap F) = \mathrm{prob}(F|E)\mathrm{prob}(E),$$

und nun zum Nenner

$$\begin{aligned}\mathrm{prob}(F) &= \mathrm{prob}(F \cap E) + \mathrm{prob}(F \cap \neg E)\\ &= \mathrm{prob}(F|E)\mathrm{prob}(E) + \mathrm{prob}(F|\neg E)\mathrm{prob}(\neg E).\end{aligned}$$

Anwendungsbeispiel der Regel von Bayes

Es gibt einen Test auf eine Krankheit. 2% der Bevölkerung sind krank. Bei einer kranken Person ist der Test mit Wahrscheinlichkeit 0.95 positiv, bei einer gesunden Person ist er mit 99% Wahrscheinlichkeit negativ. Wie groß ist die Wahrscheinlichkeit, dass man krank ist, wenn man einen positiven Test hat.

Wir definieren zuerst die Ereignisse: E = krank sein
F = positiver Test.

Dann extrahieren wir die relevanten Wahrscheinlichkeiten:

$$\begin{aligned} \mathrm{prob}(E) &= 2/100 = 0.02 & \qquad \mathrm{prob}(\neg E) &= 0.98 \\ \mathrm{prob}(F|E) &= 0.95 & \qquad \mathrm{prob}(F|\neg E) &= 0.01. \end{aligned}$$

Dann setzen wir in die Formel von Bayes ein.

$$\begin{aligned} \mathrm{prob}(E|F) &= \frac{\mathrm{prob}(F|E) \cdot \mathrm{prob}(E)}{\mathrm{prob}(F|E)\mathrm{prob}(E) + \mathrm{prob}(F|\neg E)\mathrm{prob}(\neg E)} \\ &= \frac{0.95 \cdot 0.02}{0.95 \cdot 0.02 + 0.01 \cdot 0.98} = \frac{190}{288} \approx 0.66 \end{aligned}$$

Und nun ohne die Formel

Es gibt einen Test auf eine Krankheit. 2% der Bevölkerung sind krank. Bei einer kranken Person ist der Test mit Wahrscheinlichkeit 0.95 positiv, bei einer gesunden Person ist er mit 99% Wahrscheinlichkeit negativ. Wie groß ist die Wahrscheinlichkeit, dass man krank ist, wenn man einen positiven Test hat.

Legen wir eine Population von 10000 Personen zu Grunde. Davon sind 200 krank und 9800 gesund. Bei den 200 kranken Personen ist der Test bei 190 positiv. Bei den 9800 Gesunden ist der Test bei 98 positiv. Also gibt es insgesamt 288 positive Tests. Von der Personen mit positiven Tests sind 190 tatsächlich krank. Also

$$\text{prob}(\text{krank}|\text{positiver Test}) = \frac{190}{288} \approx 0.66.$$