Bedingte Wahrscheinlichkeit

In Bemerkung hatten wir schon erwähnt, dass man ein gegebenes Wahrscheinlichkeitsmaß als Voraussage für ein Zufallsexperiment interpretieren kann. Wenn man nun zusätzliche Informationen über das Experiment erhält, so kann man diese Voraussage „verbessern“. Z.B. hat man nach einem einfachen Experiment wie Münzwurf die Information, wie das Experiment ausgegangen ist, und man kann mit dieser vollständigen Information im Nachhinein sogar eine deterministische „Voraussage“ (die dann ihren Namen eigentlich nicht mehr verdient) machen, d.h. man wird nicht mehr das a priori gegebene Wahrscheinlichkeitsmaß betrachten, sondern vielmehr ein anderes (deterministisches), das jedem Ereignis entweder die Wahrscheinlichkeit 0 oder 1 zuordnet. Im allgemeinen erhält man keine vollständige Information, sondern nur eine solche der Art, dass bestimmte Ereignisse sicher eintreten. Dementsprechend geht man zu einem neuen Wahrscheinlichkeitsmaß über.

Beispiel F.12

(Voraussage für den zweifachen Münzwurf bei zusätzlicher Information)
Wir betrachten zwei aufeinanderfolgende Münzwürfe mit einer fairen Münze. Wie groß ist die Wahrscheinlichkeit dafür, dass „zweimal Kopf“ fällt (Ereignis A), wenn man weiß, dass

1. Fall: der erste Wurf das Ergebnis „Kopf“ hat (Ereignis B_1).
2. Fall: mindestens ein Wurf gleich „Kopf“ ist (Ereignis B_2).

Als Ergebnisraum wählen wir

$$\Omega := \{(K,K),(K,Z),(Z,K),(Z,Z)\}.$$

Da wir die Münze als fair annehmen, hat jedes Elementarereignis die Wahrscheinlichkeit $1/4$. Für unsere speziell betrachteten Ereignisse gilt

$$\begin{aligned}
A &= \{(K,K)\},\\
P(A) &= \frac{1}{4},\\
B_1 &= \{(K,K),(K,Z)\},\\
P(B_1) &= \frac{1}{2},\\
B_2 &= \{(K,K),(K,Z),(Z,K)\},\\
P(B_2) &= \frac{3}{4}.
\end{aligned}$$

1. Fall: Aufgrund der zusätzlichen Informationen, dass das Ereignis B_1 eintritt, können die Elementarereignisse (Z,Z) und (Z,K) völlig ausgeschlossen werden. Es können also nur (K,K) oder (K,Z) eintreten. Ohne jegliche weitere Information sind diese beiden als gleichwahrscheinlich anzunehmen. Durch diese Überlegungen ordnen wir insbesondere dem Ereigneis (K,K) eine neue Wahscheinlichkeit zu:
$$P(A|B_1) = \frac{1}{2}.$$
Wir bezeichnen diese als **die bedingte Wahrscheinlichkeit des Ereignisses** (K,K) **bei gegebenem** B_1.
2. Fall: Es können nur $(K,K),(K,Z),(Z,K)$ eintreten. Wieder sehen wir diese Elementarereignisse als gleichwahrscheinlich an. Also
$$P(A|B_2) = \frac{1}{3}.$$

In beiden Fällen werden die möglichen Elementarereignisse auf eine Menge $B_i \subset \Omega$ reduziert. Wie wir sehen, ist die bedingte Wahrscheinlichkeit für das Ereignis A bei gegebenem B gleich

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

Mit Hilfe des letzten Ausdrucks definieren wir allgemein die bedingte Wahrscheinlichkeit.

Definition F.13 (Bedingte Wahrscheinlichkeit)

Seien (Ω, P) ein endlicher Wahrscheinlichkeitsraum, $B \subset \Omega$ mit $P(B) > 0$ und $A \in \Omega$. Die **bedingte Wahrscheinlichkeit von A bei gegebenen** B ist

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

Bemerkung

Es folgt

$$P(A \cap B) = P(B) \cdot P(A|B). \qquad (1)$$

Satz F.14 (zur bedingten Wahrscheinlichkeit)

Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum.

1. **(Die bedingte Wahrscheinlichkeit ist ein Wahrscheinlichkeitsmaß)**
Sei $P(B) > 0$. Durch

$$P_B(A) := P(A|B)$$

ist ein Wahrscheinlichkeitsmaß auf Ω definiert. Ist $A \subset B^c$ oder $P(A) = 0$, so ist $P(A|B) = 0$.

2. **(Formel der totalen Wahrscheinlichkeit)**
Sei $\Omega = \bigcup_{i=1}^{n} B_i$ mit $B_i \cap B_j = \emptyset$ für $i \neq j$ (disjunkte Zerlegung von Ω).
Dann gilt für jedes $A \subset \Omega$:

$$P(A) = \sum_{\substack{1 \leq k \leq n, \\ P(B_k) > 0}} P(B_k) \cdot P(A|B_k). \qquad (2)$$

Daher wird über alle Indizes k summiert, für die $P(B_k) > 0$. Wir schreiben der Kürze halber auch „$\sum_{k=1}^{n}$“ anstatt „$\sum_{\substack{1 \leq k \leq n, \\ P(B_k) > 0}}$“, wobei wir im Fall $P(B_k) = 0$ das Produkt als 0 definieren.

3. **(Formel von Bayes)**
Sei neben den Voraussetzungen in 2. zusätzlich noch $P(A) > 0$ erfüllt.
Dann gilt für jedes $1 \leq i \leq n$:

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum\limits_{k=1}^{n} P(B_k) \cdot P(A|B_k)}.$$

Bemerkung

Interpretation der Formel von Bayes
Wie durch das weiter unten folgende Beispiel F.15 illustriert wird, werden in der Formel von Bayes, die Ereignisse B_k als mögliche „Ursachen“ für das beobachtete Ereignis („Symptom“) A aufgefasst.
Für jedes Ereignis B_k wird die A-priori-Wahrscheinlichkeit $P(B_k)$ als bekannt vorausgesetzt und ebenso die bedingten Wahrscheinlichkeiten dafür, dass bei Eintreten von Ursache B_k auch das Symptom A eintritt.
Mit Hilfe der Formel von Bayes wird für ein B_i die A-posteriori-Wahrscheinlichkeit berechnet unter der zusätzlichen Information, dass das Symptom A beobachtet wird.
Diese Vorgehensweise der Korrektur von A-priori-Wahrscheinlichkeiten aufgrund von Beobachtungen spielt in der *Bayesischen Statistik* ein wichtige Rolle.

Beispiel F.15 (Diagnostischer Test, vgl. [Krengel])

Eine Krankheit komme bei etwa $0,5\%$ der Bevölkerung vor. Ein Test zur Auffindung der Krankheit führe bei 99% der Kranken zu einer Reaktion, aber auch bei 2% der Gesunden. Wir möchten die Wahrscheinlichkeit dafür ermitteln, dass eine Person, bei der die Reaktion eintritt, die Krankheit tatsächlich hat, und des Weiteren die Wahrscheinlichkeit, dass eine Person, bei der keine Reaktion eintritt, in Wirklichkeit krank ist. Dazu definieren wir mögliche Ereignisse:

$$\begin{aligned} B_1 &: \text{„Die Person hat die Krankheit.“}, \\ B_2 = B_1^C &: \text{„Die Person hat die Krankheit nicht.“}, \\ A_1 &: \text{„Test positiv“}, \\ A_2 = A_1^C &: \text{„Test negativ“}. \end{aligned}$$

Nach der Formel von Bayes gilt

$$\begin{aligned} P(B_1|A_1) &= \frac{P(B_1) \cdot P(A_1|B_1)}{P(B_1) \cdot P(A_1|B_1) + P(B_2) \cdot P(A_1|B_2)} \\ &= \frac{5 \cdot 10^{-3} \cdot 0.99}{5 \cdot 10^{-3} \cdot 0.99 + (1 - 5 \cdot 10^{-3}) \cdot 0.02} \approx 0.2. \end{aligned}$$

Die gesuchte bedingte Wahrscheinlichkeit für eine tatsächliche Erkrankung einer Person, bei der der Test positiv ist. beträgt etwa 0.2.

Auch die Wahrscheinlichkeit dafür, dass eine negativ getestete Person tatsächlich krank ist, berechnen wir nach der Formel von Bayes:

$$\begin{aligned} P(B_1|A_2) &= \frac{P(B_1)\cdot P(A_2|B_1)}{P(B_1)\cdot P(A_2|B_1)+P(B_2)\cdot P(A_2|B_2)} \\ &= \frac{5\cdot 10^{-3}\cdot 0.01}{5\cdot 10^{-3}\cdot 0.01+(1-5\cdot 10^{-3})\cdot 0.98} \approx 5.1\cdot 10^{-5}. \end{aligned}$$

Definition F.16 (Effizienz diagnostischer Tests, s. [Sachs])

Wir betrachten wie in Beispiel F.15 einen diagnostischen Test für eine Krankheit. Der getestete Patient kann gesund (Ereignis K^C) oder tatsächlich krank sein (Ereignis K). Der Test kann positiv ausfallen, d.h. der Patient wird als krank getestet (Ereignis T_+), oder negativ (Ereignis $T_- = T_+^C$).

1. Die **Spezifität** des Tests ist die bedingte Wahrscheinlichkeit $P(T_-|K^C)$ für einen negativen Test, wenn der Patient gesund ist.
2. Die **Sensitivität** des Tests ist die bedingte Wahrscheinlichkeit $P(T_+|K)$ für einen positiven Test, wenn der Patient krank ist.

Spezifizität und Sensitivität können wir als Gütekriterium eines Tests ansehen. Sie sollten beide nahe bei 1 liegen. Die bedingte Wahrscheinlichkeit $P(K|T_+)$ ist der **Voraussagewert** eines positiven Testergebnisses bei Kranken, und $P(K^C|T_-)$ ist der Voraussagewert eines negativen Testergebnisses bei Gesunden. Diese sollten idealerweise ebenfalls nahe bei 1 liegen. Sie hängen nach der Formel von Bayes allerdings auch von der A-priori-Wahrscheinlichkeit für die Krankheit ab, welche als die relative Häufigkeit „Anzahl der Kranken geteilt durch die Gesamtzahl der Menschen“ (z.B. in einem bestimmten Land) definiert ist, der so genannten **Prävalenz** der Krankheit. Diese Abhängigkeit kann wie in Beispiel F.15 zu niedrigen Voraussagewerten führen, wenn die Krankheit nur sehr selten ist, also zu typischem „Fehlalarm bei seltenen Ereignissen“.