3 Wahrscheinlichkeitstheorie

3.1 Wahrscheinlichkeit

• Die Wahrscheinlichkeitstheorie modelliert Situationen, in denen Unsicherheit über bestimmte Aspekte der Umwelt vorherrscht. Diese Unsicherheit kann daher rühren, dass bestimmte Aspekte der Umwelt sich erst in der Zukunft realisiern, oder daher, dass man nicht alles beobachten kann, also ein Mangel an Information herrscht.

Ereignisse Wir abstrahieren von den inhärenten Quellen (Zukunft, Informationsmangel) der Unsicherheit und modellieren diese einfach durch eine Menge von Umweltzuständen bzw. möglicher Elementarreignisse. Im folgenden bezeichen wir einen solche Menge von Umweltzuständen mit

$$\Theta = \{\theta_1, \ldots, \theta_N\}$$

(und beschränken uns auf endliche Mengen).[1]
Beispiel: Würfelwurf: $\Theta = \{1, \ldots, 6\}$.

Wir unterstellen, dass genau ein Umweltzustand eintritt bzw. wahr ist. Wir nennen diesen den wahren Zustand. In einer Situation von Unsicherheit wissen wir aber nicht, welcher Zustand wahr ist.

• Ein Ereignis A ist eine Menge von Umweltzuständen: $A \subseteq \Theta$.
Beispiel: Das Ereignis “gerade Augenanzahl wird gewürfelt” ist $\{2, 4, 6\}$.

• Ein Ereignis tritt ein bzw. ist wahr, wenn es den wahren Zustand enthält. Entsprechend nennen wir das Ereignis $A = \Theta$ das sichere Ereignis. Wir sagen “sicher”, weil der wahre Zustand immer in Θ enthalten ist. Das Ereignis $A = \emptyset$ ist das unmögliche Ereignis. Wir sagen “unmöglich”, weil der wahre Zustand niemals in $\emptyset$ enthalten ist.

Wahrscheinlichkeit Eine Wahrscheinlichkeit ordnet jedem Ereignis auf einer Skala zwischen 0 und 1 eine Zahl zu, die angibt, mit wie großer Sicherheit das Ereignis eintritt bzw. wie viel man über das Eintreten des Ereignisses weiß. Wir definieren die Wahrscheinlichkeit auf der

[1] Wir nennen Θ auch den Zustands*raum*.

Menge der Elementareignisse und erweitern sie sodann für die Menge der Ereignisse durch "Aufsummieren".

• Eine Wahrscheinlichkeitsverteilung p auf Θ ist eine Funktion

$$p : \Theta \rightarrow [0,1], \quad \theta \mapsto p(\theta)$$

mit der Eigenschaft, dass

$$p(\theta_1) + \ldots p(\theta_N) = 1.$$

Beispiele: Würfelwurf: $p(\theta) = 1/6$ für alle $\theta \in \Theta$.

Gelegentlich schreiben man eine W-Verteilung p auch als Vektor $p = (p_1, \ldots, p_N)$.

• Ausgehend davon ist die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Theta$ gegeben durch die Summe der Wahrscheinlichkeiten aller in ihm enthaltenen Elementarereignisse. Wir schreiben:

$$p(A) = \sum_{\theta \in A} p(\theta).$$

Beispiel: Würfelwurf: $p($ gerade Augenanzahl wird gewürfelt $) = 1/6 + 1/6 + 1/6 = 1/2$.

Insbesondere gilt: $p(\emptyset) = 0$, $p(\Theta) = 1$

3.2 Bedingte Wahrscheinlichkeit

• Eine Wahrscheinlichkeitsverteilung beschreibt das Ausmaß an Unsicherheit oder Information über die Umwelt. Häufig bekommt man zusätzliche Information, welche die Wahrscheinlichkeit von Ereignissen verändern kann.

Beispiel: Wenn wir (aus irgendwelchen Gründen) die Information bekommen, dass bei einem Würfelwurf eine gerade Augenzahl eingetreten ist, dann wissen wir, dass die Augenzahl nicht 1, 3, oder 5 sein kann. Deshalb sollten wir die Wahrscheinlichkeit für diese Elementarereignisse von $1/6$ auf jeweils 0 anpassen. Hingegen sollten wir die Wahrscheinlichkeit, dass eine, sagen wir, 6 gefallen ist, erhöhen. Da das Ereignis "gerade Augenzahl" drei gleich wahrscheinliche Elementarereignisse enthält, sollte die neue Wahrscheinlichkeit gleich $1/3$ sein.

Wir bezeichnen die revidierte Wahrscheinlichkeit auch mit bedingter Wahrscheinlichkeit, bedingt auf das Ereignis "gerade Augenzahl".

Bedingte Wahrscheinlichkeit Seien A und B Ereignisse mit $p(B) > 0$. Die bedingte Wahrscheinlichkeit von A, bedingt auf das Ereignis B, ist definiert durch

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}.$$

Die bedingte Wahrscheinlichkeit gibt die aktualisierte Wahrscheinlichkeit an, wenn man weiß, dass B eingetreten ist.

• Beispiel: Wir wollen den Sieger eines Autorennens mit drei Teilnehmern prognostizieren. Die Elementarereignisse sind $\Theta = \{\theta_1, \theta_2, \theta_3\}$, wobei θ_i das Ereignis ist, dass Fahrer i gewinnt. Nehmen wir an, dass Fahrer 1 favorisiert und Fahrer 3 der Außenseiter ist. Sagen wir, unsere Ausgangsprognose entspricht der W-Verteilung (1/2, 3/8, 1/8).

Unmittelbar vor dem Rennen wird bekannt, dass Fahrer 2 einen Motorschaden hat und nicht teilnehmen kann. D.h. wir wissen, dass das Ereignis $B = \{\theta_1, \theta_3\}$ eintreten wird. Um die neue Prognose zu erstellen, berechnen wir die bedingte W-keit, bedingt auf B. Durch Anwendung der obigen Formel erhalten wir:

$$p(\theta_1 \mid B) = \frac{1/2}{1/2 + 1/8} = 4/5, \quad p(\theta_2 \mid B) = \frac{0}{1/2 + 1/8} = 0, \quad p(\theta_3 \mid B) = \frac{1/8}{1/2 + 1/8} = 1/5.$$

Beachte: die Wahrscheinlichkeiten von θ_1 und θ_2 erhöhen sich zwar beide, aber ihr Verhältnis bleibt gleich dem Verhältnis ihrer Ausgangswahrscheinlichkeiten ($\frac{1/2}{1/8} = \frac{4/5}{1/5} = 4$)! Mit anderen Worten: wenn wir erfahren, dass Fahrer 2 nicht gewinnt, sollte das die *absolute* Siegeswahrscheinlichkeit der beiden anderen Fahrer jeweils erhöhen. Dies sollte aber unsere Einschätzung der relativen Stärke von Fahrer 1 zu Fahrer 3, also der *relativen* Siegeswahrscheinlichkeit, nicht verändern.

Unabhängigkeit Zwei Ereignisse sind unabhängig, falls der Eintritt des einen Ereignisses die Wahrscheinlichkeit des anderen nicht verändert. Wir definieren: die Ereignisse A und B sind stochastisch unabhängig, falls

$$p(A \cap B) = p(A) \cdot p(B).$$

Für $p(B) > 0$ gilt dann, dass $p(A \mid B) = p(A)$. Wir sagen auch: Ereignis B enthält keine Information darüber, ob Ereignis A eintritt oder nicht (und umgekehrt).

3.3 W-Verteilungen auf Produktmengen

• Häufig hat ein Zustand mehrere Ausprägungen. Solche Situationen lassen sich durch Produktmengen modellieren. Beispiel: Wurf von zwei Würfeln: $\Theta = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.

• Sei im folgenden $\Theta = X \times Y$ eine Produktmenge mit $X = \{x_1, \dots, x_N\}$ und $Y = \{y_1, \dots, y_M\}$.

Gemeinsame Verteilung Sei p eine W-Verteilung auf Θ. $p((x_i, y_j))$ ist die W-keit dass der Umweltzustand (x_i, y_j) eintritt. Wir nennen p auch *gemeinsame* Verteilung von x und y.

Randverteilung Wir können fragen, wie hoch die Wahrscheinlichkeit ist, dass eine Komponente x oder y eine bestimmte Ausprägung annimmt, unbesehen der Ausprägung der anderen Komponente. Diese Auskunft gibt die Randverteilung. Wir definieren die Randverteilungen

$$p^x(x_i) = p(\{x_i\} \times Y) \quad und \quad p^y(y_j) = p(X \times \{y_j\}).$$

Die Randverteilung p^x (bzw. p^y) ist eine W-Verteilung auf dem Zustandsraum X (bzw. Y).

Bedingte Verteilung Die bedingte Verteilung gibt die Wahrscheinlichkeit an, mit der eine Komponente eintritt, wenn wir die andere bereits kennen:

$$p(x_i \mid y_j) = \frac{p((x_i, y_j))}{p^y(y_j)} \quad und \quad p(y_j \mid x_i) = \frac{p((x_i, y_j))}{p^x(x_i)}.$$

Interpretation: Für festes y_j ist $p(\cdot \mid y_j)$ eine neue W-Verteilung auf dem Zustandsraum X. Diese W-Verteilung entspricht der revidierten W-Verteilung, wenn wir wissen bzw. beobachten können, dass die zweite Komponente des Umweltzustands gleich y_j ist.

• Die bedingte Verteilung $p(\cdot \mid y_j)$ gibt also die Wahrscheinlichkeit von x an, wenn wir wissen, dass y gleich y_j ist, wohingegen die Randverteilung p^x die Wahrscheinlichkeit von x angibt, wenn wir y nicht kennen.

Unabhängigkeit Wir sagen, x und y sind unabhängig, falls die gemeinsame Verteilung gleich dem Produkt der Randverteilung ist. D.h. für alle $(x_i, y_j) \in \Theta$ gilt:

$$p((x_i, y_j)) = p^x(x_i) \cdot p^y(y_j).$$

In diesem Falle enthält eine Komponente keine weitere Information über das Eintreten der anderen Komponente, denn dann gilt:

$$p(x_i \mid y_j) = p^x(x_i) \quad und \quad p(y_j \mid x_i) = p^y(y_j).$$

Beispiel Das Wetter kann entweder gut ($x = G$) oder schlecht ($x = S$) werden, und die Ernte kann entweder hoch ($y = 1$) oder niedrig ($y = 0$) ausfallen. Der Zustandsraum ist $\{S, G\} \times \{0, 1\}$. Wir können die gemeinsame Verteilung übersichtlich in einer Tabelle darstellen:

p	$y = 0$	$y = 1$
$x = S$	$3/10$	$2/10$
$x = G$	$1/10$	$4/10$

Die Randverteilungen von x gibt an, wie wahrscheinlich gutes oder schlechtes Wetter ist:

$$p^x(S) = p(\{G\} \times Y) = p((G, 0)) + p((G, 1)) = 2/10 + 3/10 = 1/2.$$

Ebenso gibt die Randverteilung von y an, wie wahrscheinlich eine hohe oder niedrige Ernte ist:

$$p^y(0) = p(X \times \{0\}) = p((S, 0)) + p((G, 0)) = 3/10 + 1/10 = 2/5.$$

Wir erhalten also die Randverteilung einfach durch Aufsummieren der entsprechenen Zeilen oder Spalten der obigen Tabelle.

Die bedingte Verteilung von y, bedingt auf x, gibt an, mit welcher Ernte wir rechnen müssen, wenn wir das Wetter kennen:

$$p(y = 0 \mid S) = \frac{3/10}{3/10 + 2/10} = 3/5 \quad \textit{und} \quad p(y = 1 \mid S) = 2/5,$$
$$p(y = 0 \mid G) = \frac{1/10}{1/10 + 4/10} = 1/5 \quad \textit{und} \quad p(y = 1 \mid G) = 4/5.$$

Die W-Verteilung spiegelt also die plausible Annahme wider, dass eine hohe Ernte bei gutem Wetter wahrscheinlicher als bei schlechtem Wetter ist. Insbesondere sind somit Wetter und Ernte nicht stochastisch unabhängig, denn wenn wir wissen, dass das Wetter gut ist, verändert das die Wahrscheinlichkeit einer hohen Ernte verglichen mit der unbedingten (bzw. Rand-) Wahrscheinlichkeit.

Umgekehrt können wir auch Rückschlüsse auf die Wetterqualität ziehen, wenn wir nur die Erntehöhe beobachten können. Diese Auskunft gibt uns die bedingte Verteilung von x, bedingt auf y. Etwa

$$p(S \mid y = 0) = \frac{3/10}{3/10 + 1/10} = 3/4.$$

Bayes' Rule Die beiden bedingten Verteilungen sind wie folgt verknüpft:

$$\begin{aligned} p(x_i \mid y_j) &= \frac{p((x_i, y_j))}{p^y(y_j)} \\ &= \frac{p(y_j \mid x_i) \cdot p^x(x_i)}{p^y(y_j)} \\ &= \frac{p(y_j \mid x_i) \cdot p^x(x_i)}{p(y_j \mid x_1)p^x(x_1) + \ldots + p(y_j \mid x_N)p^x(x_N)}. \end{aligned}$$

Diese Formel ist als Bayes' Rule bekannt.[2] Sie besagt, dass wir die bedingte Verteilung von x, bedingt auf y, berechnen können, wenn wir die bedingte Verteilung von y, bedingt auf x, und die Randverteilung von x kennen.

3.4 Inferenz

• Bayes' Rule spielt eine wichtige Rolle in der Spieltheorie, wenn Situationen beschrieben werden, in denen die Spieler über unterschiedliche Information verfügen. Dort stellt sich die Frage, was das Verhalten eines Gegenspielers über einen Umweltzustand offenbart, den allein der Gegenspieler kennt, nicht aber ich.

• Zur Illustration betrachten wir eine Menge von Zuständen $\Theta = \{\theta_1, \theta_2, \theta_3\}$ und eine Menge von Aktionen $A = \{a_1, a_2\}$. Sei p eine W-Verteilung auf Θ. Nimm an, dass ein Spieler eine Aktion aus A wählt und davor den Zustand θ beobachten kann, er kann also seine Aktion vom Zustand abhängig machen. Sein Verhalten wird dann durch eine Strategie $\sigma : \Theta \to A$ beschrieben.

Als Beispiel seien die Zustände θ Einkommensniveaus $\theta_1 < \theta_2 < \theta_3$, und die Aktionen seien der Kauf einer Jeans, wobei a_1 einer No-Name-Jeans und a_2 einer Designer-Jeans entspreche. Nehmen wir an, der Käufer kauft die Marken-Jeans nur beim höchsten Einkommensniveau, befolgt also die Strategie:

$$\theta_1 \mapsto a_1, \quad \theta_2 \mapsto a_1, \quad \theta_3 \mapsto a_2.$$

Weiterhin nehmen wir an, dass die W-Verteilung auf Θ durch (1/6, 2/3, 1/6) gegeben ist.

Fragen wir nun, welche Rückschlüsse wir auf das Einkommen des Konsumenten ziehen können, wenn wir nur beobachten können, welche Jeans er trägt, nicht aber seinen Gehaltszettel

[2] Thomas Bayes, englischer Mathematiker, 1702-1761.

direkt. Wenn wir die Designer-Jeans beobachten, dann wissen wir mit Sicherheit, dass der Konsument das höchste Einkommen hat, denn andernfalls kauft er die No-Name-Jeans. Also

$$p(\theta_1 \mid a_2) = 0, \quad p(\theta_2 \mid a_2) = 0, \quad p(\theta_3 \mid a_2) = 1.$$

Die gleiche Antwort bekommen wir durch Anwenden von Bayes' Rule:

$$p(\theta_3 \mid a_2) = \frac{p(a_2 \mid \theta_3)p(\theta_3)}{p(a_2 \mid \theta_1)p(\theta_1) + p(a_2 \mid \theta_2)p(\theta_2) + p(a_2 \mid \theta_3)p(\theta_3)} = \frac{1 \cdot 1/6}{0 \cdot 1/6 + 0 \cdot 2/3 + 1 \cdot 1/6} = 1.$$

Hierbei ergeben sich die Wahrscheinlichkeiten $p(a \mid \theta)$ durch die Spezifikation der Strategie! Da der Konsument im Zustand θ_1 mit Wahrscheinlichkeit 1 die Aktion a_1 wählt, gilt: $p(a_2 \mid \theta_1) = 0$, usw.

Wenn wir die No-Name-Jeans beobachten, können wir mit Sicherheit ausschließen, dass der Käufer das Einkommen θ_3 hat. Zur Anpassung der Wahrscheinlichkeit dass der Käufer Einkommen θ_1 bzw θ_2 hat, wenden wir Bayes' Rule an. Dann erhalten wir:

$$p(\theta_1 \mid a_1) = 1/5, \quad p(\theta_2 \mid a_1) = 4/5, \quad p(\theta_3 \mid a_2) = 0.$$

In der Tat,

$$p(\theta_1 \mid a_1) = \frac{p(a_1 \mid \theta_1)p(\theta_1)}{p(a_1 \mid \theta_1)p(\theta_1) + p(a_1 \mid \theta_2)p(\theta_2) + p(a_1 \mid \theta_3)p(\theta_3)} = \frac{1 \cdot 1/6}{1 \cdot 1/6 + 1 \cdot 2/3 + 0 \cdot 1/6} = \frac{1}{5}.$$

Verglichen zur Ausgangsposition, in der wir nichts beobachten können, erhöht sich die Wahrscheinlichkeit, dass wir es mit mittleren Einkommenstypen θ_2 zu tun haben, wenn er die No-Name-Jeans trägt, also von 2/3 auf 4/5.