KAPITEL 2

Bedingte Wahrscheinlichkeiten, Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeiten

Definition 2.1. *Im diskreten Wahrscheinlichkeitsraum* $(\Omega, \mathbf{P})$ *sei ein Ereignis* $A \subset \Omega$ *mit* $\mathbf{P}(A) > 0$ *gegeben.* $\mathbf{P}(\cdot|A) : \mathcal{P}(\Omega) \to [0,1]$ *definiert durch*

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

*heißt **bedingte Wahrscheinlichkeitsverteilung** gegeben* A*.*

Bemerkung. $\mathbf{P}(\cdot|A)$ *ist ein Wahrscheinlichkeitsmaß auf* Ω *mit* $\mathbf{P}(B|A) = 1$ *für alle* $B \subset \Omega$ *mit* $A \subset B$*, wie eine kurze Übung zeigt.*

Interpretation:

1. Wir wissen, dass ein gegebener Zufallsmechanismus durch $(\Omega, \mathbf{P})$ beschrieben wird.
2. Wir erhalten die Information, dass der Ausgang $\omega \in A$ erfüllt.
3. Beschreibe den Zufallsmechanismus jetzt durch $(\Omega, \mathbf{P}(\cdot|A))$.

Beispiel 9. *Wir betrachten eine Krankheit mit einer Prävalenz von* 1%*. Für eine zufällig ausgewählte Person gilt also*

$$\mathbf{P}(\textit{Person hat die Krankheit}) = 0.01.$$

Für die Krankheit existiert ein Test mit Sensitivität 98%*:*

$$\mathbf{P}(\textit{Test positiv} \mid \textit{getestete Person hat Krankheit}) = 0.98$$

und Spezifizität 98.5%:

$$\mathbf{P}(\textit{Test positiv} \mid \textit{getestete Person hat Krankheit nicht}) = 0.015.$$

Wird nun im Rahmen einer Routineuntersuchung eine Person positiv getestet, stellt sich die Frage, wie hoch die Wahrscheinlichkeit ist, dass diese Person tatsächlich die fragliche Krankheit hat. Gesucht wird also

$$\mathbf{P}(\textit{getestete Person hat Krankheit} \mid \textit{Test positiv}).$$

Schreiben wir kurz p *bzw.* n *für einen positiven bzw. negativen Test und* $+$ *bzw.* $-$ *für "Person hat die Krankheit" bzw. "hat die Krankheit nicht", so berechnen wir die gesuchte Wahrscheinlichkeit als*

$$\begin{aligned}
\mathbf{P}(+ \mid p) &= \frac{\mathbf{P}(+,p)}{\mathbf{P}(p)} \\
&= \frac{\mathbf{P}(p \mid +)\mathbf{P}(+)}{\mathbf{P}(p,+) + \mathbf{P}(p,-)} \\
&= \frac{\mathbf{P}(p \mid +)\mathbf{P}(+)}{\mathbf{P}(p \mid +)\mathbf{P}(+) + \mathbf{P}(p \mid -)\mathbf{P}(-)} \\
&= \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.015 \cdot 0.99} \approx 0.4.
\end{aligned}$$

Es wäre also durchaus voreilig, auf den positiven Test hin eine Behandlung mit möglicherweise erheblichen Nebenwirkungen zu starten.
Etwas anders stellt sich die Situation dar, wenn nur Personen diesem Test unterzogen werden, die a priori ein stark erhöhtes Risiko für die Krankheit haben. Nehmen wir also an, die zu testende Person gehöre einer Risikogruppe an, innerhalb der die Prävalenz 10% *beträgt. Dann ist* $\mathbf{P}(+) = 0.1$ *und die Rechnung ergibt*

$$\mathbf{P}(+|p) = \frac{0.98 \cdot 0.1}{0.98 \cdot 0.1 + 0.015 \cdot 0.9} \approx 0.88.$$

Die in Beispiel 9 genutzte Rechnung ist ein Spezialfall des folgenden Satzes.

Satz 2.2. *Seien* $(\Omega, \mathbf{P})$ *ein diskreter Wahrscheinlichkeitsraum,* $A \subset \Omega$ *und* $\{B_1, \ldots, B_J\}$ *eine Zerlegung von* Ω*, d.h.,* $B_i \cap B_j = \emptyset$ *für* $i \neq j$ *und* $\Omega = \bigcup_{j=1}^J B_j$*. Ferner gelte* $\mathbf{P}(A) > 0$ *und* $\mathbf{P}(B_j) > 0$ *für* $j = 1, \ldots, J$*. Dann gelten:*

(i)

$$\mathbf{P}(A) = \sum_{j=1}^J \mathbf{P}(B_j)\mathbf{P}(A|B_j) \qquad \textbf{\textit{Formel von der totalen Wahrscheinlichkeit}}$$

(ii) und

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(B_i)\mathbf{P}(A|B_i)}{\sum_{j=1}^J \mathbf{P}(B_j)\mathbf{P}(A|B_j)} \quad \textit{für } 1 \leq i \leq J \qquad \textbf{\textit{Bayesformel}}$$

Beweis. Dies folgt ganz analog zur Rechnung im Beispiel durch elementare Umformungen aus der Definition der bedingten Wahrscheinlichkeit. □

Bemerkung. *Seien $\mathbf{P}_1, \mathbf{P}_2$ zwei Wahrscheinlichkeitsmaße auf demselben Stichprobenraum Ω und $\{B_1, \ldots, B_J\}$ eine disjunkte Zerlegung von Ω. Aus $\mathbf{P}_1(A|B_j) \leq \mathbf{P}_2(A|B_j)$ folgt im allgemeinen* <u>*nicht*</u> *$\mathbf{P}_1(A) \leq \mathbf{P}_2(A)$. Begründung mit der Formel von der totalen Wahrscheinlichkeit:*

$$\mathbf{P}_i(A) \quad = \quad \sum_j \mathbf{P}_i(B_j)\mathbf{P}_i(A|B_j), \qquad i = 1, 2$$

Die Ungleichung $\mathbf{P}_1(A|B_j) \leq \mathbf{P}_2(A|B_j)$ impliziert nicht viel für $\mathbf{P}_1(A)$ und $\mathbf{P}_2(A)$, da i.a. $\mathbf{P}_1(B_j) \neq \mathbf{P}_2(B_j)$.

Beispiel 10 (Das Paradoxon von Simpson)**.** *Wir wollen die Wirksamkeit einer Therapie für eine Krankheit untersuchen. Dazu beobachten wir eine Population von Personen, die an der Krankheit leiden. Wir nutzen folgende Bezeichnungen für Teilmengen unserer Population ein:*

B_j :	*Personen in Altersgruppe j*
A :	*Personen, die am Ende des Beobachtungszeitraums geheilt sind*
T :	*Personen, die die Therapie erhalten.*

Außerden bezeichnen wir mit $\mathbf{P}_1 = \mathbf{P}(\cdot|T)$ und $\mathbf{P}_2 = \mathbf{P}(\cdot|T^c)$ die Wahrscheinlichkeiten für die Ausgänge von behandelten bzw. unbehandelten Personen. Nicht selten werden wir feststellen, dass $\mathbf{P}_1(A) < \mathbf{P}_2(A)$, aber für alle Altersgruppen $\mathbf{P}_1(A|B_j) > \mathbf{P}_2(A|B_j)$ gilt.

2.2 Unabhängigkeit

Das scheinbare Paradoxon im letzten Beispiel war möglich, weil sowohl die Inanspruchnahme der Therapie als auch die Heilungschance von der Zugehörigkeit zur Altersgruppe B_j abhing. Das motiviert eine mathematische Behandlung des Konzepts der Unabhängigkeit von Ereignissen.

Definition 2.3. *Zwei Ereignisse $A, B \subset \Omega$ heißen unabhängig, wenn*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$$

gilt.

Bemerkung. *Wenn zwei Ereignisse A, B unabhängig sind und $\mathbf{P}(A) > 0$ bzw. $\mathbf{P}(A^c) > 0$ gilt, dann gilt*

$$\mathbf{P}(B|A) = \mathbf{P}(B) \text{ bzw. } \mathbf{P}(B|A^c) = \mathbf{P}(B),$$

d.h., ***die Wahrscheinlichkeit****, dass B eintritt, hängt nicht davon ab, ob A eintritt. Aber es kann natürlich sein, dass das Eintreten von B doch von A abhängt, wie man aus folgendem Beispiel sieht. Werfe Würfel zweimal:*

$$\begin{aligned} A &= \{ \text{ Erste Augenzahl ist gerade } \}, \\ B &= \{ \text{ Summe beider Augenzahlen ist gerade } \}. \end{aligned}$$

Man rechnet leicht nach, dass die beiden Ereignisse unabhängig sind.

Unabhängigkeit von mehr als zwei Mengen definiert man rekursiv.

Definition 2.4. *Die Ereignisse $A_1, \dots, A_n \subset \Omega$ mit $n \geq 3$ heißen unabhängig, wenn*

(i) für jedes $k = 1, \dots, n$ die Mengen A_i, $i = 1, \dots, n$, $i \neq k$ unabhängig sind und

(ii) $\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \cdots \mathbf{P}(A_n)$ gilt.

Bemerkung. *Äquivalent sind $A_1, \dots, A_n$ unabhängig, wenn für jede mindestens zweielementige Teilmenge $\{k_1, \dots, k_m\} \subset \{1, \dots, n\}$*

$$\mathbf{P}(A_{k_1} \cap \ldots \cap A_{k_m}) = \mathbf{P}(A_{k_1}) \cdots \mathbf{P}(A_{k_m})$$

gilt.
Im allgemeinen folgt aus paarweiser Unabhängigkeit der A_i nicht die Unabhängigkeit im Sinne der rekursiven Definition.

Unabhängigkeit können wir nicht nur für Ereignisse, sondern auch für Zufallsvariablen definieren.

Definition 2.5. *Ω_i'-wertige Zufallsvariablen X_i $(i = 1, \dots, n)$ auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathbf{P})$ heißen unabhängig, falls für beliebige Mengen $B_i \subset \Omega_i'$ gilt:*

$$\mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1) \cdot \dots \cdot \mathbf{P}(X_n \in B_n).$$

Wir überzeugen uns leicht, dass dies eine Verallgemeinerung von Definition 2.4 ist. Tatsächlich sind $A, \dots, A_n \subset \Omega$ nämlich genau dann unabhängig, wenn die Zufallsvariablen $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ (alle mit Werten in $\Omega_j' = \{0, 1\}$) unabhängig sind.

Bemerkung. *Dann gilt für $1 \leq j \leq n$:*

$$\mathbf{P}\big(X_j \in B_j | X_i \in B_i (i \neq j)\big) = \frac{\prod_i \mathbf{P}(X_i \in B_i)}{\prod_{i \neq j} \mathbf{P}(X_i \in B_i)} = \mathbf{P}(X_j \in B_j),$$

falls $\mathbf{P}(X_i \in B_i) > 0$ für $i \neq j$. Dabei haben wir $X_i \in B_i (i \neq j)$ kurz für

$$X_1 \in B_1, \dots, X_{j-1} \in B_{j-1}, X_{j+1} \in B_{j+1}, \dots, X_n \in B_n$$

geschrieben. Wissen über die Realisierungen von X_i mit $(i \neq j)$ nützt einem bei unabhängigen Zufallsvariablen nichts zur Beurteilung des stochastischen Verhaltens von X_j.

Beispiel 11. *Beim n-maligen Werfen eines Würfels bezeichnen wir mit X_j die Augenzahl des jten Wurfs. Die X_j sind dann die Projektionen $X_j : \Omega \to \Omega'_j$ von $\Omega = \{1, \ldots, 6\}^n$ auf $\Omega'_j = \{1, \ldots, 6\}$ (hier sind alle Ω'_j gleich), gegeben durch*

$$X_j((\omega_1, \ldots, \omega_n)) = \omega_j.$$

Dann sind $X_j^{-1}(B_j) = \{(\omega_1, \ldots, \omega_n) \in \Omega : \omega_j \in B_j\}$, und man rechnet leicht nach, dass die X_j unabhängig sind.

Bemerkung. *(1) Da wir mit diskreten Wahrscheinlichkeitsräumen arbeiten, ist die Definition der Unabhängigkeit äquivalent zu*

$$\mathbf{P}(X_1 = \omega'_1, \ldots, X_n = \omega'_n) = \mathbf{P}(X_1 = \omega'_1) \cdots \mathbf{P}(X_n = \omega'_n)$$

für alle $(\omega'_1, \ldots, \omega'_n) \in \Omega'_1 \times \Omega'_n$, d.h., die gemeinsame Verteilung der X_i ist das Produkt der Randverteilungen.

(2) Seien X_i $(i = 1, \ldots, n)$ Ω'_i-wertige unabhängige Zufallsvariablen, $f_1 : \Omega'_1 \to \Omega''_1, \ldots, f_n : \Omega'_n \to \Omega''_n$ Funktionen und $\Omega'_1, \Omega''_1, \ldots, \Omega'_n, \Omega''_n$ abzählbare Mengen. Dann sind auch

$$f_1(X_1), \ldots, f_n(X_n)$$

unabhängig.

Eine oft studierte Situation sieht folgendermaßen aus: Seien $X_1, \ldots, X_n$ unabhängig, reellwertig und identisch verteilt mit dem Wahrscheinlichkeitsmaß $\mathbf{P}^{X_i} = \mathbf{P}$. Spätestens hier nutzen wir, dass wir in der Regel nur die Informationen über die Verteilungen der X_i brauchen. Der genaue Grundraum Ω spielt dabei oft gar keine Rolle und wird in der

In der Wahrscheinlichkeitstheorie fragen wir dann zum Beispiel nach dem Verhalten der Summe $X_1 + \ldots + X_n$ für große n, genauer für $n \to \infty$.

In der Statistik interessieren wir uns vor allem darüber, was uns die Werte von $X_1, \ldots, X_n$ (Beobachtungen, eine Stichprobe) über $\mathbf{P}$ verraten kann.

Beispiel 12. *Seien $X_1, \ldots, X_n$ unabhängige $\{0, 1\}$-wertige Zufallsvariablen mit $\mathbf{P}(X_i = 1) = 1 - \mathbf{P}(X_i = 0) = p$. Dann ist $S_n = X_1 + \cdots + X_n$ binomialverteilt zum Parameter (n, p). Dies gilt weil:*

$$\begin{aligned}
\mathbf{P}(S_n = k) &= \ldots + \mathbf{P}(X_1 = X_2 = 0, X_3 = \cdots = X_{k+2} = 1, X_{k+3} = \cdots = X_n = 0) \\
&\quad + \ldots \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

Kurz: Die Summe unabhängiger, identisch Bernoulli-verteilter Zufallsvariablen ist binomialverteilt.

Bemerkung. *Wie bei Ereignissen gilt auch für Zufallsvariablen: Aus paarweiser Unabhängigkeit folgt im allgemeinen nicht Unabhängigkeit.*

Beispiel 13. *Seien X_1, X_2 unabhängig und gleichverteilt auf $\{0, \ldots, k\}$ und sei*

$$X_3 = \begin{cases} X_1 + X_2 & \text{falls } 0 \leq X_1 + X_2 \leq k, \\ X_1 + X_2 - (k+1) & \text{sonst.} \end{cases}$$

Es gilt X_i, X_j sind unabhängig für jedes Paar $1 \leq i, j \leq 3$. Aber X_1, X_2, X_3 sind nicht unabhängig.

Bemerkung. *Im letzten wie auch schon in vorherigen Beispielen mussten wir über den Wahrschienlichkeitsraum $(\Omega, \mathbf{P})$, auf dem die Zufallsvariablen definiert sind, nichts wissen. Für alle Rechnungen und insbesondere für die Frage nach der Unabhängigkeit genügt es völlig, die gemeinsame Verteilung von $(X_1, \ldots, X_n)$ zu kennen. Daraus erhält man auch sofort die Randverteilungen gemäß:*

$$\mathbf{P}^{X_i}(\{k\}) = \sum_{\omega'_j \in \Omega'_j, i \neq j} \mathbf{P}^{(X_1, \ldots, X_n)}(\{(\omega'_1, \ldots, \omega'_{i-1}, k, \omega'_{i+1} \ldots, \omega'_n)\}).$$

Die Summation erfolgt also über alle $(\omega'_1, \ldots, \omega'_n) \in \Omega'_1 \times \ldots \times \Omega'_n$ mit $\omega'_i = k$, also alle möglichen Werte von $(X_1, \ldots, X_n)$, die $X_i = k$ realisieren.

Beispiel 14. *Für $\Omega'_1 = \{1, 2, 3\}$, $\Omega'_2 = \{1, 2\}$ sei die gemeinsame Verteilung von $X = (X_1, X_2)$ durch*

$$\mathbf{P}^X(\{(1,1)\}) = \mathbf{P}^X(\{(1,2)\}) = \frac{1}{4}, \quad \mathbf{P}^X(\{(2,1)\}) = \mathbf{P}^X(\{(3,2)\}) = 0$$

$$\mathbf{P}^X(\{(2,2)\}) = \frac{1}{6}, \quad \mathbf{P}^X(\{(3,1)\}) = \frac{1}{3}$$

gegeben. Dann sind die Randverteilungen:

$$\begin{aligned}
\mathbf{P}^{X_1}(\{1\}) &= \mathbf{P}^X(\{(1,1)\}) + \mathbf{P}^X(\{(1,2)\}) &&= \frac{1}{2}, \\
\mathbf{P}^{X_1}(\{2\}) &= \mathbf{P}^X(\{(2,1)\}) + \mathbf{P}^X(\{(2,2)\}) &&= \frac{1}{6}, \\
\mathbf{P}^{X_1}(\{3\}) &= \mathbf{P}^X(\{(3,1)\}) + \mathbf{P}^X(\{(3,2)\}) &&= \frac{1}{3}, \\
\mathbf{P}^{X_2}(\{1\}) &= \mathbf{P}^X(\{(1,1)\}) + \mathbf{P}^X(\{(2,1)\}) + P^X(\{(3,1)\}) &&= \frac{7}{12} \\
\mathbf{P}^{X_2}(\{2\}) &= \mathbf{P}^X(\{(1,2)\}) + \mathbf{P}^X(\{(2,2)\}) + P^X(\{(3,2)\}) &&= \frac{5}{12}.
\end{aligned}$$

Wegen (u.a.)

$$\mathbf{P}^X(\{(2,1)\}) = 0 \neq \frac{1}{6}\frac{7}{12} = \mathbf{P}^{X_1}(\{2\})\,\mathbf{P}^{X_2}(\{1\})$$

sind X_1 und X_2 nicht unabhängig.