Kapitel 3.

Modelle für Zähldaten

> Everybody speaks of probability, but no one is able to say what it is, in a way which is satisfactory for others.
>
> *(Garrett Birkhoff)*

3.1. Einführung

In diesem Kapitel werden wir uns mit Modellen von Zähldaten beschäftigen. Von Zähldaten sprechen wir, wenn der Ausgang einer Messung eine *natürliche* Zahl ergibt. Als Beispiel betrachten wir folgende Fragestellungen:

- Wir werfen einen Würfel 300 mal. Wie viele Male kommt die Augenzahl 4 vor?
- Wir wählen zufällig 100 männliche Personen aus. Wie viele Personen haben eine Körperlänge von mehr als 173 cm?

In diesen Fällen wird die Messung einmal durchgeführt. Führt man die Messung ein zweites Mal durch, so werden die Resultate verschieden zu jener der ersten Messung sein. Sind im zweiten Beispiel das erste Mal zufällig sehr viele Asiaten dabei und bei einer zweiten Messung zufällig viele Basketballspieler, so wird es sicherlich Unterschiede in der gemessenen Anzahl von Personen mit Körpergrössen von mehr als 173 cm geben. Wird aber die Messung aus dem zweiten Beispiel 1000 Mal wiederholt, dürfte sich herauskristallisieren, was die zu *erwartende* Anzahl Personen mit einer Körpergrösse von mehr als 173 cm ist.

Können wir aber theoretisch vorhersagen, wie gross der Wert einer zukünftigen Messung wahrscheinlich sein wird? Um eine solche Vorhersage machen zu können, brauchen wir *Modelle* für die Messungen. Nehmen wir an, wir würfeln mit einem „fairen“ Würfel. Dann gehen wir davon aus, dass die Wahrscheinlichkeit eine 4 zu würfeln, genau $1/6$ ist. Dies ist allerdings nur ein *Modell*. Wenn wir den Würfel 1.2 Million mal werfen, dann wird die Anzahl der auftretenden 4 wohl nie *genau* 200 000 sein. Es ist durchaus möglich, wenn auch sehr unwahrscheinlich, dass wir in all diesen Würfen nie eine 4 würfeln. Wir haben also eine - in diesem Fall vernünftige - Idealisierung gemacht. Wir sprechen auch von einer *Modellierung* der Problemstellung. Die Vorhersagen, die wir machen können, sind aber mit Unsicherheiten behaftet, und wir können nur sagen, mit welcher *Wahrscheinlichkeit* ein Resultat eintritt.

Es folgt noch ein physikalisches Beispiel dieser Überlegungen, das uns auch als Standardbeispiel dienen soll.

Der Alpha-Zerfall

Ein radioaktiver Atomkern, der beim Zerfall Helium-4-Atomkerne aussendet, wird als *Alphastrahler* bezeichnet. Der vom zerfallenden Atomkern emittierte Helium-4-Atomkern wird *Alphateilchen* genannt. Diese Alphateilchen kann man messen. Die Emission von Alphateilchen von einer radioaktiven Quelle pro Zeiteinheit ist *nicht konstant*, sondern fluktuiert auf *zufällige* Art und Weise. Der Strahler kann also in einem 10-Sekundenintervall 10 Alphateilchen aussenden, dann in den nächsten 10 Sekunden gar keine und dann wieder 100. Ein typischer Alphastrahler ist Americium 241, ein Zerfallsprodukt von Plutonium-241, das in radioaktiven Abfällen häufig anzutreffen ist.

Anzahl Zerfälle	Beobachtet in Anzahl Experimente
0-2	18
3	28
4	56
5	105
6	126
7	146
8	164
9	161
10	123
11	101
12	74
13	53
14	23
15	15
16	9
17+	5

Tabelle 3.1.: Anzahl Zerfälle in 10 Sekunden und Anzahl Experimente (von insgesamt 1207), in denen diese Anzahl Zerfälle beobachtet wurde

In einem Experiment wurde nun die Anzahl Zerfälle von Americium 241 in einem Intervall von 10 Sekunden gemessen. Das Experiment wurde 1207 Mal wiederholt, jedes Mal wurde die Anzahl Zerfälle in 10 Sekunden gemessen. In Tabelle 3.1 ist in der ersten Spalte die Anzahl Zerfälle aufgeführt, in der zweiten Spalte, wie oft diese Anzahl Zerfälle in den 1207 Experimenten beobachtet worden ist. So sind in 18 der 1207 Experimente 0, 1 oder 2 Alphateilchen gemessen worden. In 28 der 1207 Experimente wurden 3 Alphateilchen gemessen.

Wie viele Zerfälle sich in einem bestimmten Zeitintervall ereignen werden, kann mit einem physikalischen Gesetz nicht vorausgesagt werden. Wir können aber versuchen, die Wahrscheinlichkeit für jede Anzahl Zerfälle im Zeitintervall von 10 Sekunden anzugeben. Wir sind also bestrebt, ein Wahrscheinlichkeitsmodell für den Alpha-Zerfall zu *konstruieren*.

Es stellt sich nun die Frage, wie wir als erstes die Zerfallswahrscheinlichkeiten aus den Daten *schätzen* können. Wie gross ist also beispielsweise die Wahrscheinlichkeit, dass wir in 10 Sekunden 12 Zerfälle messen? Wir stellen dazu die Häufigkeiten der Anzahl Zerfälle graphisch dar. Dazu betrachten wir das Histogramm unserer Daten (siehe Abbildung 3.1 links). Wir erhalten dabei eine sogenannte *Häufigkeitsverteilung*. Daraus können wir zum Beispiel ablesen, dass in 74 Experimenten 12 Zerfälle beobachtet wurden. Wir sprechen in diesem Fall von *absoluten Häufigkeiten*. Diese Zahl

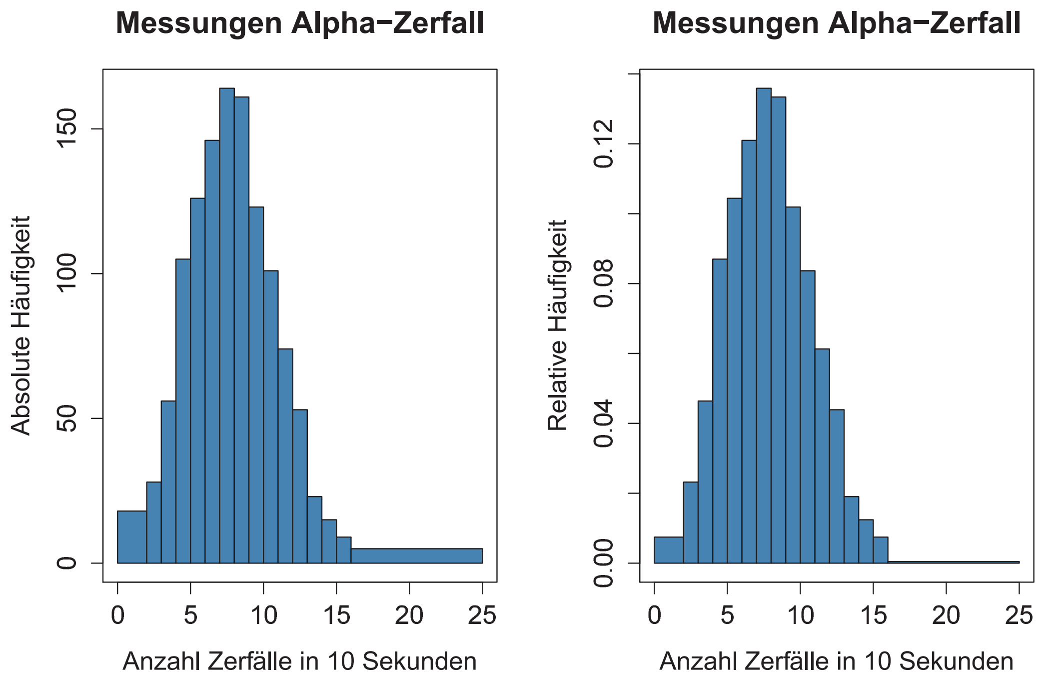

Abbildung 3.1.: Häufigkeitsverteilung des Alphazerfalles

74 in Bezug auf die beobachtete Anzahl Experimente ist ohne die Angabe der Gesamtzahl 1207 beobachteter Experimente allerdings nicht sehr aussagekräftig. Diese

Zahl lässt sich nur schwer mit einer anderen Versuchsreihe vergleichen, zum Beispiel mit einer, bei der 2389 Experimente durchgeführt wurden. Mehr Aufschluss gibt die *relative Häufigkeit* einer bestimmten Anzahl Zerfälle. Dabei dividieren wir die (absolute) Häufigkeit der Experimente, bei denen eine bestimmte Anzahl Zerfälle beobachtet wurde, durch die Gesamtzahl durchgeführter Experimente. Dies ergibt dann den *prozentualen* Anteil der Experimente, bei denen eine bestimmte Anzahl Zerfälle beobachtet wurde. Wie wir im Kapitel über Histogramme gesehen haben, entspricht die Fläche eines Balkens der relativen Häufigkeit von Messwerten im entsprechenden Intervall. So lesen wir aus Abbildung 3.1 rechts, dass in rund 12 % der Experimente 7 Zerfälle in 10 Sekunden stattfanden. Die Gesamtfläche der Balken ergibt in diesem Fall eins. Dies führt uns jetzt natürlich direkt zum Begriff der Zerfallswahrscheinlichkeit. Wir interpretieren also die relativen Häufigkeiten der beobachteten Anzahl Zerfälle pro 10 Sekunden als (geschätzte) *Wahrscheinlichkeiten*: die Wahrscheinlichkeit, dass in 10 Sekunden eine entsprechende Anzahl Zerfälle beobachtet wird. Wahrscheinlichkeiten werden in der Mathematik in der Regel nicht als Prozente angegeben, sondern mit Zahlen zwischen 0 und 1. Die Wahrscheinlichkeit, dass in 10 Sekunden 7 Zerfälle stattfinden, beträgt also 0.12. Je mehr Experimente wir durchführen, desto mehr nähert sich diese geschätzte Wahrscheinlichkeit der wahren Wahrscheinlichkeit an. Weiter stellen wir fest, dass sich alle Zerfälle auf eins aufsummieren.

Als nächstes stellt sich die Frage: wie sind diese Wahrscheinlichkeiten auf die Anzahl Zerfälle verteilt? Gibt es ein mathematisches Modell, das zur beobachteten Verteilung der relativen Häufigkeiten passt, mit dem wir dann auch Vorhersagen machen können?

3.2. Wahrscheinlichkeitsmodelle

Wir betrachten **Zufallsexperimente**, bei denen der Ausgang nicht exakt vorhersagbar ist. Beispiele dafür sind

- Anzahl Zerfälle eines Alphastrahlers
- geworfene Augenzahl bei einem Würfelwurf

Ein **Wahrscheinlichkeitsmodell** beschreibt, welche Ergebnisse in einem solchen Experiment möglich sind und welche Wahrscheinlichkeiten, die verschiedenen Ergebnisse haben. Beim Würfel sind die möglichen Ergebnisse $1, 2, 3, 4, 5, 6$ und die Wahrscheinlichkeit eine dieser Zahlen zu werfen, beträgt $1/6$, sofern der Würfel fair ist.

Ein Wahrscheinlichkeitsmodell erlaubt uns dann, gewisse Vorhersagen zu machen, die wir experimentell überprüfen können. Damit können wir uns zum Beispiel bei Glücksspielen eine gute Spielstrategie erarbeiten. Ein Wahrscheinlichkeitsmodell hat die folgenden Komponenten:

- **Grundraum** Ω, bestehend aus den **Elementarereignissen** ω,

- **Ereignisse** $A, B, C, \ldots,$
- **Wahrscheinlichkeiten** P

Elementarereignisse sind mögliche Ergebnisse oder Ausgänge des Experiments, die zusammen den Grundraum bilden:

$$\Omega = \{\underbrace{\text{mögliche Elementarereignisse } \omega}_{\text{mögliche Ausgänge/Resultate}}\}$$

Beispiel 3.2.1

Beim Würfelwurf ist der Grundraum gegeben durch die möglichen Ergebnisse des Zufallsexperiments

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Das Element $\omega = 2$ ist ein Elementarereignis. Dieses hat die Bedeutung, dass beim Würfeln die Zahl 2 geworfen wurde.

□

Beispiel 3.2.2

Beim Alphazerfall ist der Grundraum gegeben durch

$$\Omega = \{0, 1, 2, 3, 4, \ldots\}$$

da hier beliebig viele Zerfälle in einem 10-Sekundenintervall möglich sind. Das Elementarereignis $\omega = 6$ bedeutet, dass in 10 Sekunden 6 Zerfälle gemessen wurden.

□

Bei der Durchführung eines Experiments wird aus der Menge aller Elementarereignisse (Grundraum) ein Elementarereignis zufällig „ausgewählt“.

Beispiel 3.2.3 2-maliges Werfen einer Münze

Wir wählen die Bezeichnungen K für „Kopf“ und Z für „Zahl“ (dies gilt für das gesamte Kapitel). Dann sind alle möglichen Ergebnisse des Experimentes gegeben durch

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

wobei Ω wieder den Grundraum bezeichnet. Ein Elementarereignis ist z.B. $\omega = KZ$.

□

Unter einem **Ereignis** A versteht man eine Teilmenge von Ω:

$$A \subset \Omega$$

„Ein Ereignis A tritt ein“ bedeutet, dass das Ergebnis ω des Experiments zu A gehört.

Beispiel 3.2.4 2-maliges Werfen einer Münze

Wir betrachten das Ereignis A, bei welchem genau einmal K geworfen wird. Dieses Ereignis besteht aus den Elementarereignissen KZ und ZK. Das Ereignis A ist dann gegeben durch die Menge

$$A = \{KZ, ZK\}$$

Würfeln wir ZZ, so trifft das Ereignis A *nicht* ein.

□

Beispiel 3.2.5 Würfeln

Das Ereignis A bezeichne „eine ungerade Augenzahl würfeln“. Dann ist

$$A = \{1, 3, 5\}$$

Das Ereignis A tritt ein, wenn z.B. die Zahl 5 gewürfelt wird.

Wir bezeichnen mit B das Ereignis, eine Zahl kleiner als 7 zu würfeln. Das ist natürlich immer der Fall und somit ist in diesem Fall

$$B = \Omega$$

Wir sprechen vom *sicheren* Ereignis.

Weiter sei C das Ereignis „die Zahl sieben würfeln". Dies ist unmöglich und wir schreiben

$$C = \emptyset$$

Dabei stellt $\emptyset$ die leere Menge dar, die kein Element enthält. Wir sprechen in diesem Fall vom *unmöglichen* Ereignis.

□

Für den Umgang mit Ereignissen ist es nützlich, sich die Operationen der Mengenlehre und deren Bedeutung in Erinnerung zu rufen. Die Operationen der Mengenlehre (Vereinigung, Durchschnitt, Komplement) werden verwendet, um aus bereits definierten Ereignissen neue Ereignisse zu gewinnen (siehe Tabelle 3.2).

Name	Symbol	Bedeutung
Vereinigung	$A \cup B$	A **oder** B
Durchschnitt	$A \cap B$	A **und** B
Komplement	$\overline{A}$	**nicht** A
Differenz	$A \backslash B = A \cap \overline{B}$	A **ohne** B

Tabelle 3.2.: Operationen der Mengenlehre

Bemerkungen:

i. Das „oder" der Mengenoperation *Vereinigung* ist nicht exklusiv: Ein Element kann also in A *und* in B liegen. Umgangssprachlich wird „oder" meist als „entweder ... oder ..." gedeutet. Das ist beim „oder" der Mengenoperation Vereinigung nicht so.

ii. Anstelle des Begriffs *Durchschnitt* verwendet man oft auch den Begriff „Schnittmenge".

iii. Für das Komplement $\overline{A}$ wird auch die Bezeichnung A^c verwendet.

Die Mengen A und B heissen **disjunkt**, wenn sich A und B gegenseitig ausschliessen und daher nicht zusammen eintreten können. In diesem Fall gilt

$$A \cap B = \emptyset$$

Dieses Ereignis ist also unmöglich.

Ferner gelten die **De Morgan - Regeln**

1. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
2. $\overline{A \cup B} = \overline{A} \cap \overline{B}$

Alle diese Begriffe, Operationen und Regeln lassen sich einfach mit Venn-Diagrammen illustrieren (siehe Abbildung 3.2).

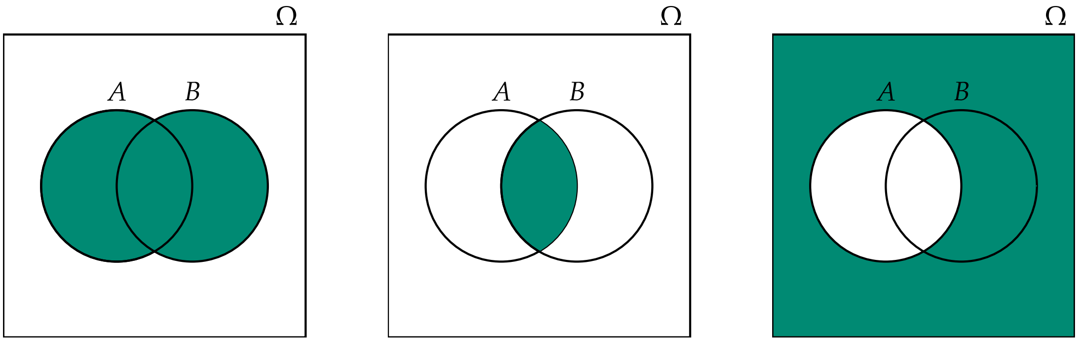

Abbildung 3.2.: *Links:* $A \cup B$, *Mitte:* $A \cap B$ und *Rechts:* $\overline{A}$

Beispiel 3.2.6

Wir betrachten die Ereignisse

- A: „morgen scheint die Sonne“
- B: „morgen regnet es“

Dann bedeuten folgende Ereignisse:

1. $A \cup B$:

   „morgen scheint die Sonne *oder* morgen regnet es“, und dies kann auch bedeuten: „morgen scheint die Sonne *und* morgen regnet es“.

2. $A \cap B$:

   „morgen scheint die Sonne *und* morgen regnet es“

3. $\overline{A}$:

   „morgen scheint die Sonne nicht“

□

Die Wahrscheinlichkeitsrechnung wird auf folgenden drei Grundregeln (Axiome) aufgebaut:

Kolmogorov Axiome der Wahrscheinlichkeitsrechnung

Jedem Ereignis A wird eine **Wahrscheinlichkeit** $P(A)$ zugeordnet. Dabei müssen die folgenden drei grundlegenden Regeln - **Axiome der Wahrscheinlichkeitsrechnung** - erfüllt sein:

A1: $P(A) \geq 0$

A2: $P(\Omega) = 1$

A3: $P(A \cup B) = P(A) + P(B) \quad$ falls $A \cap B = \emptyset$

Bemerkungen:

i. Die Bezeichnung $P(A)$ steht für die Wahrscheinlichkeit, dass das Ereignis A eintritt. Ist A das Ereignis, eine ungerade Zahl zu würfeln, so gilt bei einem fairen Würfel
$$P(A) = \frac{1}{2}$$

ii. Der Buchstabe P steht für das englische Wort *probability*.

iii. Wahrscheinlichkeiten sind nie negativ.

iv. Mit $P(\Omega) = 1$ wird festgelegt, dass die Wahrscheinlichkeiten eines Ereignisses zwischen 0 und 1 liegen müssen.

Beispiel 3.2.7

Beim Wurf zweier Münzen ist es plausibel, dass alle 4 Elemente von

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

gleich wahrscheinlich sind. Wegen $P(\Omega) = 1$ müssen sich die Wahrscheinlichkeiten zu eins aufaddieren. Somit gilt

$$P(KK) = P(KZ) = P(ZK) = P(ZZ) = \frac{1}{4}$$

□

Weitere Regeln können aus den drei Axiomen von Kolmogorov hergeleitet werden:

Rechenregeln

Sind A, B und $A_1, \ldots A_n$ Ereignisse, dann gilt

$P(\overline{A}) = 1 - P(A)$	für jedes A
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	für beliebige A und B
$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$	für beliebige $A_1, \ldots, A_n$
$P(B) \leq P(A)$	für beliebige A und B mit $B \subseteq A$
$P(A \backslash B) = P(A) - P(B)$	für beliebige A und B mit $B \subseteq A$

Wir können uns die Wahrscheinlichkeiten als Flächen im Venn-Diagramm vorstellen. Dabei ist die Totalfläche von Ω gleich 1 oder $|\Omega| = 1$. Dann erscheinen diese Rechenregeln ganz natürlich. So folgt die erste Regel sofort aus Abbildung 3.2 rechts. Dabei ist $P(A)$ der Flächeninhalt der Fläche A und $P(\overline{A})$ ist der Flächeninhalt der restlichen Fläche in Ω. Es gilt also offensichtlich

$$P(A) + P(\overline{A}) = |\Omega| = 1$$

und damit

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

Überprüfen Sie die restlichen Regeln als Übung.

Grundsätzlich werden in der Wahrscheinlichkeitstheorie die Wahrscheinlichkeiten gewisser Ereignisse A festgelegt (auf Grund von Plausibilitäten, Symmetrieüberlegungen, wissenschaftlichen Theorien, Fachwissen und Daten) und daraus die Wahrscheinlichkeiten von gewissen anderen Ereignissen B aus den obigen Gesetzen hergeleitet.

Die Statistik geht umgekehrt vor: aus Daten, d. h. aus der Information, dass gewisse Ereignisse eingetreten sind, versucht man Rückschlüsse auf ein unbekanntes Wahrscheinlichkeitsmodell (unbekannte Wahrscheinlichkeiten) zu ziehen.

Interpretation von Wahrscheinlichkeiten:

- **Frequentistisch:** Idealisierung der relativen Häufigkeiten bei vielen unabhängigen Wiederholungen

  Aus Messungen zum Alphazerfall (Experiment) schliessen wir, dass in 12 % der Beobachtungen 7 Zerfälle in 10 Sekunden vorkommen.

- **Bayes'sch:** Mass für den Glauben, dass ein Ereignis eintreten wird

  „Ich bin mir zu 90 % sicher, dass ich die nächste Prüfung bestehen werde."

Wir behandeln in diesem Kapitel *diskrete* Wahrscheinlichkeitsmodelle, bei denen der Grundraum endlich oder unendlich und diskret ist. Mit dem Begriff „diskret" meinen wir zum Beispiel folgende Menge

$$\Omega = \{0, 1, \ldots, 10\}$$

die *endlich und deshalb diskret* ist, während

$$\Omega = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$$

zwar *unendlich*, aber trotzdem *diskret* ist.

Die Menge $\Omega = \mathbb{R}$ ist *nicht* diskret. Sie wird später für Messdaten eine sehr wichtige Rolle spielen.

Im diskreten Fall ist die Wahrscheinlichkeit eines Ereignisses

$$A = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

durch die Wahrscheinlichkeiten der zugehörigen Elementarereignisse $P(\omega)$ festgelegt:

$$P(A) = P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = \sum_{\omega_i \in A} P(\omega_i)$$

Dies folgt aus Axiom A3.

Beispiel 3.2.8

Wir haben einen Würfel, der nicht fair ist. Die Wahrscheinlichkeiten, unterschiedliche Zahlen zu werfen, sind also nicht gleich. In Tabelle 3.3 sind die entsprechenden Wahrscheinlichkeiten für alle Zahlen angegeben.

ω	1	2	3	4	5	6
$P(\omega)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$

Tabelle 3.3.: Wahrscheinlichkeiten für einen nicht-fairen Würfel.

Es gilt

$$\begin{aligned} P(\Omega) &= P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{4} + \frac{1}{12} + \frac{1}{12} \\ &= 1 \end{aligned}$$

Für das Ereignis $A = \{1, 2, 4\}$ gilt

$$\begin{aligned} P(\Omega) &= P(1) + P(2) + P(4) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

□

Beispiel 3.2.9 2-maliges Werfen einer Münze

Es sei das Ereignis A : „genau einmal K werfen“, also

$$A = \{KZ, ZK\}$$

Dann erhalten wir für die Wahrscheinlichkeit $P(A)$, dass das Ereignis A eintritt

$$P(A) = P(KZ) + P(ZK) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Für das Ereignis B: „mindestens einmal Kopf werfen“ erhalten wir die zugehörige Wahrscheinlichkeit, dass dieses Ereignis eintritt

$$P(B) = P(KZ) + P(ZK) + P(KK) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

da $B = \{KZ, ZK, KK\}$. Diese Wahrscheinlichkeit können wir in diesem Fall einfacher mit der sogenannten *Gegenwahrscheinlichkeit* berechnen. Das Komplement $\overline{B}$ von B ist gegeben durch

$$\overline{B} = \{ZZ\}$$

Und damit erhalten wir mit der ersten Rechenregel (siehe oben):

$$P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{4} = \frac{3}{4}$$

□

Laplace-Wahrscheinlichkeit

In vielen Fällen ist es plausibel anzunehmen, dass jedes Elementarereignis die gleiche Wahrscheinlichkeit hat. Ein solches Wahrscheinlichkeitsmodell wird **Modell von Laplace** genannt. Das Standardbeispiel hierfür ist der Wurf eines fairen Würfels. Da wird *jede* Zahl mit der Wahrscheinlichkeit $1/6$ geworfen. In diesen Fällen gibt es eine besonders einfache Möglichkeit, die Wahrscheinlichkeit eines Ereignisses zu berechnen. Ein Ereignis E besteht aus den g verschiedenen Elementarereignissen

$$E = \{\omega_1, \omega_2, \ldots, \omega_g\}$$

und der Grundraum Ω aus m Elementarereignissen.

Damit sich die Wahrscheinlichkeiten aller Elementarereignisse zu 1 addieren (siehe oben), haben wir hier

$$P(\omega_k) = \frac{1}{|\Omega|} = \frac{1}{m}$$

Somit gilt für $P(E)$:

$$\begin{aligned} P(E) &= P(\omega_1) + P(\omega_2) + \ldots + P(\omega_g) \\ &= \frac{1}{m} + \frac{1}{m} + \ldots + \frac{1}{m} \\ &= g \cdot \frac{1}{m} \\ &= \frac{g}{m} \end{aligned}$$

Für ein Ereignis E im Laplace Modell gilt also

$$P(E) = \sum_{\omega_k \in E} P(\omega_k) = \frac{g}{m}$$

Man teilt die Anzahl der „**günstigen**“ Elementarereignisse durch die Anzahl der „**möglichen**“ Elementarereignisse.

Beispiel 3.2.10

Es werden zwei Würfel geworfen. Wie gross ist die Wahrscheinlichkeit, dass die Augensumme 7 ergibt?

Ein Elementarereignis beschreibt die Augenzahlen auf beiden Würfeln. Dieses Ereignis können wir in der Form $(1,4)$ schreiben, wenn der eine Würfel eine 1 und der andere eine 4 zeigt. Es sind insgesamt 36 Elementarereignisse möglich:

$$\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$$

Wir bezeichnen mit E das Ereignis, dass die Augensumme 7 erreicht wird. Es gibt davon 6 Elementarereignisse:

$$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

Da alle Elementarereignisse gleich wahrscheinlich sind, ist die Wahrscheinlichkeit für das Ereignis E:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

□

3.3. Der Begriff der Unabhängigkeit

Wenn wir die Wahrscheinlichkeiten $P(A)$ und $P(B)$ kennen, so lässt sich die Wahrscheinlichkeit $P(A \cap B)$ im Allgemeinen nicht aus $P(A)$ und $P(B)$ berechnen: Es sind alle Werte zwischen 0 und dem Minimum von $P(A)$ und $P(B)$ möglich. Ein wichtiger Spezialfall liegt vor, wenn die Berechnung von $P(A \cap B)$ aus $P(A)$ und $P(B)$ mit Hilfe folgender Produktformel möglich ist.

Die Ereignisse A und B heissen **stochastisch unabhängig**, wenn

$$P(A \cap B) = P(A)P(B)$$

Beispiel 3.3.1

Es sei A: „K im 1. Wurf“ und B: „K im 2. Wurf“. Dann gilt

$$P(A) = P(KK) + P(KZ) = \frac{1}{2}$$

und analog

$$P(B) = P(KK) + P(ZK) = \frac{1}{2}$$

Mit $A \cap B$ wird das Ereignis „K im 1. *und* im 2. Wurf" beschrieben. Es gilt dann für die Wahrscheinlichkeit dieses Ereignisses

$$P(A \cap B) = P(KK) = \frac{1}{4}$$

Da

$$P(A)P(B) = \frac{1}{4}$$

gilt also

$$P(A \cap B) = P(A)P(B)$$

Somit erfüllen die Ereignisse A und B die Bedingung für stochastische Unabhängigkeit. Die Ereignisse A und B sind also unabhängig.

$\square$

Viel wichtiger ist jedoch der umgekehrte Schluss. Wenn zwischen den Ereignissen A und B kein *kausaler Zusammenhang* besteht - d.h. es gibt keine gemeinsamen Ursachen oder Ausschliessungen - dann nehmen wir stochastische Unabhängigkeit an, und damit *muss* obige Produktformel gelten. In diesem Fall kann also $P(A \cap B)$ aus $P(A)$ und $P(B)$ berechnet werden.

Die stochastische Unabhängigkeit soll durch die folgenden Beispiele veranschaulicht werden.

Beispiel 3.3.2

Wir werfen zuerst einen Würfel und ziehen nachher eine Karte aus einem vollständigen Stapel Jasskarten. Wie gross ist die Wahrscheinlichkeit, zuerst eine 3 zu werfen und nachher einen König zu ziehen?

Dazu bezeichnen wir dieses Elementarereignis mit $3K$. Die Wahrscheinlichkeit, eine 3 zu werfen ist $1/6$, also

$$P(3) = \frac{1}{6}$$

Die Wahrscheinlichkeit, einen König zu ziehen ist $1/9$:

$$P(K) = \frac{1}{9}$$

Der Würfelwurf hat auf das Ziehen einer Jasskarte keinen Einfluss. Zur Berechnungen der Wahrscheinlichkeit, dass beide Ereignisse eintreffen, müssen wir $1/9$ von einem Sechstel nehmen. Dies entspricht einer Multiplikation der beiden Wahrscheinlichkeiten und es gilt

$$P(3K) = P(3) \cdot P(K) = \frac{1}{6} \cdot \frac{1}{9} = \frac{1}{54}$$

Die Produktformel gilt, da die beiden Ereignisse unabhängig sind.

□

Beispiel 3.3.3

Es sei A: „K im 1. Wurf" und B: „K im 2. Wurf".

Es ist plausibel, dass es keinen kausalen Zusammenhang zwischen dem Ergebnis des ersten und des zweiten Wurfs gibt. Mit anderen Worten hat das Ergebnis des 1. Wurfes (Ereignis A) keinen Einfluss auf das Ergebnis des 2. Wurfes (Ereignis B). Die Ereignisse A und B sind also unabhängig. Deshalb kann man $P(A \cap B)$ wie folgt berechnen:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

□

Beispiel 3.3.4

Angenommen, wir werfen zwei Würfel. Sei E das Ereignis, dass der erste Würfel die Augenzahl 4 zeigt und F das Ereignis, dass die *Augensumme* 6 ist. Sind die Ereignisse E und F unabhängig voneinander? Um dies festzustellen, müssen wir überprüfen, ob die Produktregel

$$P(E \cap F) = P(E) \cdot P(F)$$

gilt oder nicht.

Das Ereignis $E \cap F$ besteht nur aus dem Elementarereignis $(4, 2)$. Damit gilt für dessen Wahrscheinlichkeit

$$P(E \cap F) = P((4, 2)) = \frac{1}{36}$$

Es gilt natürlich $P(E) = {}^1\!/_6$. Das Ereignis F besteht aus den Elementen

$$F = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$$

und damit gilt

$$P(F) = \frac{5}{36}$$

Nun können wir das Produkt dieser Wahrscheinlichkeiten berechnen und damit die Produktformel überprüfen:

$$P(E) \cdot P(F) = \frac{1}{6} \cdot \frac{5}{36} = \frac{5}{216} \neq \frac{1}{36} = P(E \cap F)$$

Deshalb sind E und F nicht unabhängig.

Wir können die Abhängigkeit der beiden Ereignisse E und F auch ohne Rechnung leicht einsehen. Werfen wir den ersten Würfel und erhalten eine 6, dann haben wir keine Möglichkeit mehr in zwei Würfen die Augensumme 6 zu erreichen : Die Wahrscheinlichkeit ist 0. Werfen wir zuerst irgendeine andere Zahl, so haben wir noch die Möglichkeit, auf die Augensumme 6 zu kommen: Die Wahrscheinlichkeit ist ungleich 0. Damit hängt die Wahrscheinlichkeit, die Augensumme 6 zu erzielen von der Augenzahl des ersten Wurfes ab. Also können E und F nicht unabhängig voneinander sein.

□

Das vorherige Beispiel führt uns direkt zum Begriff der *bedingten Wahrscheinlichkeit*. Die Wahrscheinlichkeit, in zwei Würfen die Augensumme 6 zu erreichen, *gegeben* im ersten Wurf wurde eine 6 gewürfelt, ist null.

Haben wir drei Ereignisse A_1, A_2 und A_3, so sind diese Ereignisse unabhängig, wenn

$$P(A_1 \cap A_2) = P(A_1)P(A_2),\ P(A_1 \cap A_3) = P(A_1)P(A_3),\ P(A_2 \cap A_3) = P(A_2)P(A_3)$$

und

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

Für n Ereignisse $A_1, \ldots, A_n$ wird die Unabhängigkeit analog definiert. Alle möglichen Schnittmengen von Ereignissen $A_1, \ldots, A_n$ müssen die Produktformel erfüllen.

3.4. Wahrscheinlichkeitsbäume

Ein Zufallsprozess, der aus mehreren nacheinander ablaufenden Zufallsexperimenten besteht, wird *mehrstufiges Zufallsexperiment* genannt. Die Beispiele aus dem vorangehenden Kapitel sind solche mehrstufigen Zufallsexperimente. Wird eine Münze zweimal geworfen, handelt es sich um einen Zufallsprozess, bei dem nacheinander zwei einfache Zufallsexperimente ablaufen (*2-stufiges* Zufallsexperiment). Wird zuerst ein Würfel geworfen und danach eine Jasskarte aus einem Stapel gezogen, so stellt auch dieser Zufallsprozess ein 2-stufiges Zufallsexperiment dar.

Ein äusserst anschauliches Hilfsmittel bei der Berechnung von Wahrscheinlichkeiten bei *mehrstufigen* Zufallsexperimenten ist der sogenannte *Wahrscheinlichkeitsbaum* (häufig auch als *Ereignisbaum* oder *Baumdiagramm* bezeichnet). Er besteht aus einer *Wurzel* (Ausgangspunkt) und mehreren *Verzweigungspunkten* und einer Vielzahl von *Zweigen*. In Abbildung 3.3 ist ein solcher Wahrscheinlichkeitsbaum dargestellt. Die Verzweigungspunkte A_1 und A_2 charakterisieren dabei die möglichen Zwischenergebnisse nach der 1. Stufe. Die von diesen Verzweigungspunkten ausgehenden Zweige führen zu den jeweils möglichen Ergebnissen der nachfolgenden 2. Stufe : Hier

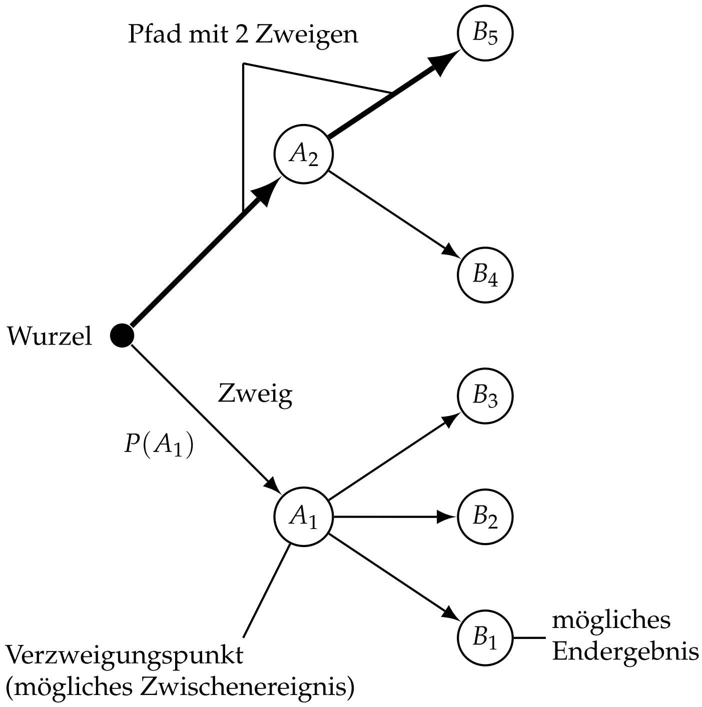

Abbildung 3.3.: Wahrscheinlichkeitsbaum mit den Verzweigungspunkten A_1 und A_2 (mögliche Ergebnisse der 1. Stufe) und den möglichen Endergebnissen B_1 bis B_5.

sind es die möglichen *Endergebnisse* B_1 bis B_5. Die Wahrscheinlichkeit eines bestimmten Ereignisses schreibt man an den betreffenden Zweig. So ist z.B. $P(A_1)$ die Wahrscheinlichkeit des Ereignisses A_1. Ein mögliches *Endergebnis* erreicht man dann immer von der Wurzel ausgehend längs eines bestimmten Pfades. Dieser besteht meistens aus mehreren Zweigen, wie z.B. der in Abbildung dick eingezeichnete Pfad, der von der Wurzel über das Zwischenergebnis A_2 zum Endergebnis B_5 führt. Für die Berechnung der Wahrscheinlichkeit eines bestimmten Endergebnisses gelten dann die folgenden Regeln:

Regeln für die Berechnung von Wahrscheinlichkeiten bei mehrstufigen Zufallsexperimenten mit Hilfe des Wahrscheinlichkeitsbaumes

Bei einem in mehreren Stufen ablaufenden Zufallsexperiment lässt sich jedes mögliche Endergebnis durch einen bestimmten Pfad im zugehörigen *Wahrscheinlichkeitsbaum* anschaulich darstellen. Die Berechnung der Wahrscheinlichkeiten für die möglichen Endergebnisse erfolgt dabei mit Hilfe der folgenden Regeln:

1. Die Wahrscheinlichkeiten längs eines *Pfades* werden miteinander *multipliziert*.
2. Führen *mehrere* Pfade zum *gleichen* Endergebnis, so *addieren* sich ihre Wahrscheinlichkeiten.

Beispiel 3.4.1 2-maliges Werfen einer Münze

Es sei A: „K im 1. Wurf" und B: „K im 2. Wurf". Wir möchten mit Hilfe des Wahrscheinlichkeitsbaumes die Wahrscheinlichkeit des Ereignisses $E = A \cap B$ berechnen.

Die Wahrscheinlichkeit für „K im 1. Wurf" beträgt $1/2$, die Wahrscheinlichkeit für „K im 2. Wurf" beträgt wieder $1/2$. Somit ist die Wahrscheinlichkeit des Endergebnisses E (K im 1. Wurf *und* K im 2. Wurf) gegeben durch die Multiplikation der Wahrscheinlichkeiten längs des Pfades, also

$$P(E) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Wir lesen aus dem Wahrscheinlichkeitsbaum (siehe Abbildung 3.4), dass es für die Wahrscheinlichkeit von „K im 2. Wurf" keine Rolle spielt, ob wir im 1. Wurf „Z" oder „K" erhalten haben. Die Ereignisse „K im 1. Wurf" und „K im 2. Wurf" sind stochastisch unabhängig.

□

Beispiel 3.4.2 Würfel und Jasskarte

Betrachten wir das Beispiel, in welchem zuerst ein (fairer) Würfel geworfen und dann eine Jasskarte gezogen wird. Uns interessiert die Wahrscheinlichkeit, zuerst eine 3 zu werfen und nachher einen König zu ziehen. Wir können dazu den in Abbildung 3.5 dargestellten Wahrscheinlichkeitsbaum zur Berechnung verwenden. In diesem Fall haben wir nicht nur ein Endergebnis, sondern deren vier: auf der 2. Stufe sind „König Schellen", „König Schilten", „König Eichel" und „König Rosen" Elemente des Ereignisses, zuerst eine 3 zu werfen und nachher einen König zu ziehen. Daher addieren wir die Wahrscheinlichkeiten aller Pfade, die zum selben Endergebnis führen. Dies ergibt dann die Wahrscheinlichkeit $4 \cdot \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{54}$.

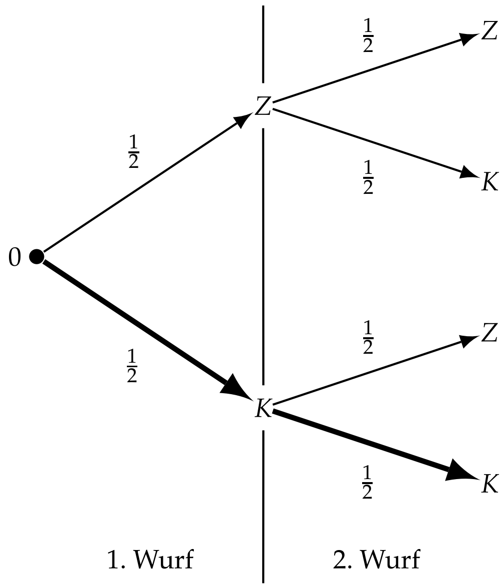

Abbildung 3.4.: Wahrscheinlichkeitsbaum für das zweimalige Werfen einer Münze.

Würden wir den Wahrscheinlichkeitsbaum vervollständigen, wäre ersichtlich, dass die Wahrscheinlichkeit, auf der 2. Stufe einen König zu ziehen, unabhänigig vom Zwischenergebnis der 1. Stufe ist. Dies zeigt wiederum, dass die Augenzahl auf der 1. Stufe und die gezogene Jasskarte unabhänige Ereignisse sind.

□

Beispiel 3.4.3

Betrachten wir wiederum das Beispiel, in dem wir zwei Würfel werfen. Uns interessiert das Endergebnis, wo die *Augensumme* der beiden Würfel 6 beträgt. Wir bezeichnen mit F das Ereignis „Augenzahl des 2. Würfels, so dass Augensumme der beiden Würfel gleich 6 ist". Der entsprechende Wahrscheinlichkeitsbaum ist in der Abbildung 3.6 dargestellt.

Nun sehen wir, dass die Wahrscheinlichkeiten für die Zweige auf der 2. Stufe vom Ausgang der vorangehenden Stufe abhängen. Die Wahrscheinlichkeit des Ereignisses F am Zweig nach dem Zwischenergebnis „6 beim 1. Würfel" beträgt 0, während die Wahrscheinlichkeit für F an den übrigen Zweigen $^1/_6$ beträgt. Es handelt sich hier um *bedingte* Wahrscheinlichkeiten, auf die wir nun im folgenden Kapitel eingehen.

□

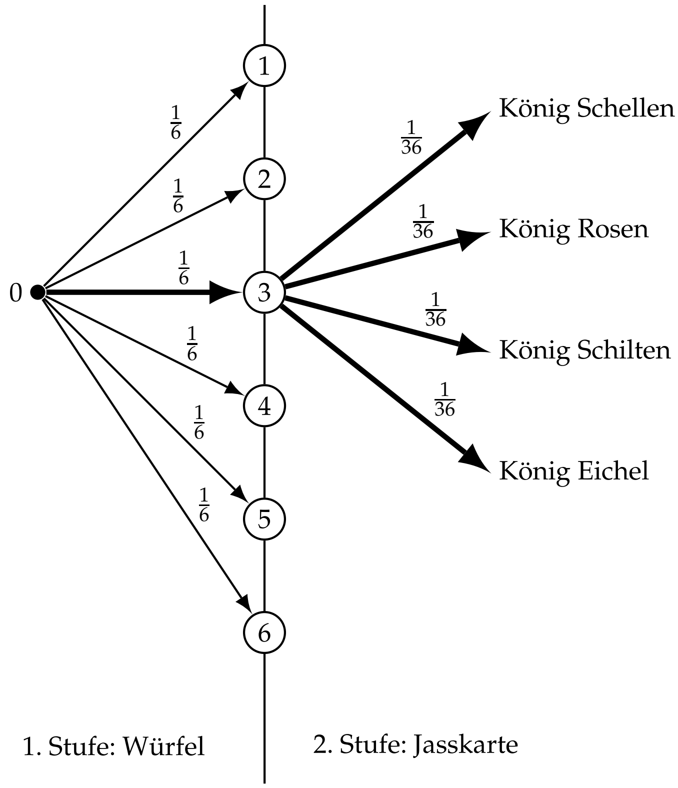

Abbildung 3.5.: Wahrscheinlichkeitsbaum für das Werfen einer (fairen) Münze und dann das Ziehen einer Jasskarte.

3.5. Bedingte Wahrscheinlichkeit

Wir wollen zuerst den Begriff der bedingten Wahrscheinlichkeit an einem kleinen Beispiel kennenlernen. Dazu betrachten wir einen fairen Würfel, bei dem die ungeraden Zahlen blau und die geraden Zahlen rot gefärbt sind. Wir werfen nun den Würfel und sehen, dass eine blaue Seite oben liegt. Damit wissen wir, dass eine ungerade Zahl geworfen wurde. Wie gross ist nun die Wahrscheinlichkeit, dass die Zahl 3 geworfen wurde? Es gibt die drei ungeraden Zahlen 1, 3 und 5 auf dem Würfel. Die 3 ist eine davon. Wenn wir davon ausgehen, dass jede Zahl auf dem Würfel gleich wahrscheinlich ist, ist das Ergebnis $^1/_3$. Natürlich können wir auch nach der Wahrscheinlichkeit fragen, die Zahl 2 zu werfen, wenn eine blaue Seite oben ist. Diese Wahrscheinlichkeit ist natürlich 0, da wir ja schon wissen, dass eine ungerade Zahl geworfen wurde. Der entsprechende Wahrscheinlichkeitsbaum ist in Abbildung 3.7 dargestellt. Wir bemerken bei der Bestimmung der Wahrscheinlichkeiten für die Zweige des Wahrscheinlichkeitsbaumes, dass von der 2. Stufe ab alle Ereignisse vom Ausgang der *vorangegangen* 1. Stufe abhängen.

Fragen dieser Art kann man auch in komplizierteren Sachverhalten mit dem Begriff

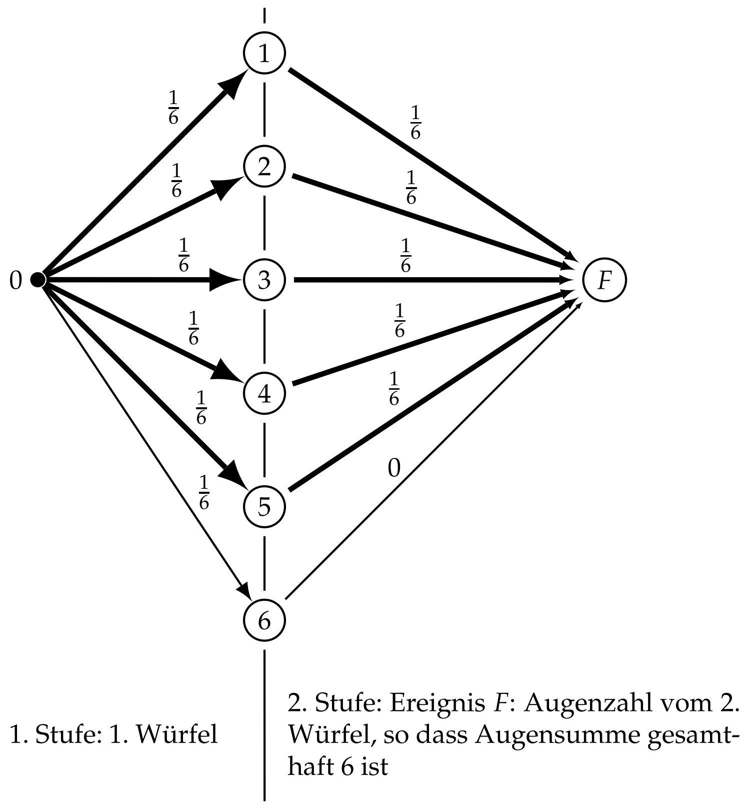

Abbildung 3.6.: Wahrscheinlichkeitsbaum für die Augensumme 6 in zwei Würfen.

der **bedingten Wahrscheinlichkeit** beantworten.

Die **bedingte Wahrscheinlichkeit** ist die Wahrscheinlichkeit, dass das Ereignis A eintritt, wenn wir schon wissen, dass B eingetreten ist. Diese Wahrscheinlichkeit wird mit

$$P(A|B)$$

bezeichnet. Der Längsstrich wird als „unter der Bedingung“ gelesen.

Die bedingte Wahrscheinlichkeit $P(A|B)$ wird definiert durch

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

Die Interpretation ist folgendermassen: $P(A|B)$ ist die Wahrscheinlichkeit für das Ereignis A, wenn wir wissen, dass das Ereignis B schon eingetroffen ist.

Wie ist diese Formel zu verstehen? Da wir wissen, dass B schon eingetreten ist, haben wir einen neuen Grundraum $\Omega' = B$. Damit müssen wir von A nur noch denjenigen

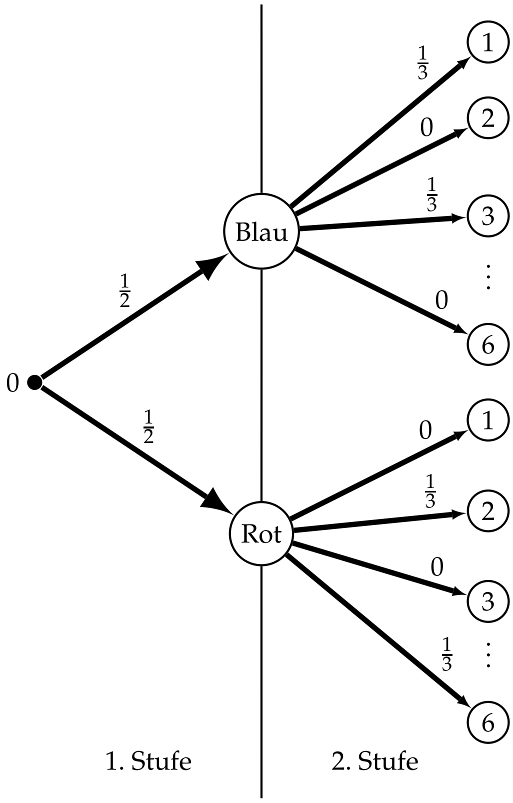

Abbildung 3.7.: Wahrscheinlichkeitsbaum für Wurf eines Würfels, wobei Seiten mit ungerader Augenzahl blau, Seiten mit gerader Augenzahl rot gefärbt sind.

Teil anschauen, der sich in B abspielt (daher $A \cap B$). Dies müssen wir jetzt noch in Relation zur Wahrscheinlichkeit von B bringen. Das lässt sich anschaulich einfach im Venn-Diagramm in Abbildung 3.8 verdeutlichen.

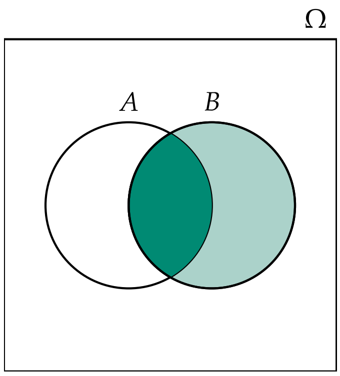

Abbildung 3.8.: Hilfsillustration für bedingte Wahrscheinlichkeiten.

Wenn wir die Wahrscheinlichkeiten wieder als Flächen denken ($|\Omega| = 1$), dann ist die bedingte Wahrscheinlichkeit $P(A|B)$ der Flächeninhalt der dunkel gefärbten Flächen, während $P(B)$ der Flächeninhalt der gesamten gefärbten Fläche B ist. Der Anteil der dunkelgefärbten Fläche zur gefärbten Fläche ist dann

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

Bemerkungen:

i. In der Definition für die bedingte Wahrscheinlichkeit wird stillschweigend davon ausgegangen, dass $P(B) > 0$ gilt. Das muss auch so sein, da wir sonst eine Division durch 0 hätten.

Beispiel 3.5.1

Wir kommen nochmals auf das Beispiel vom Anfang dieses Unterkapitels zurück und betrachten es mit den Begriffen der bedingten Wahrscheinlichkeit.

Sei A : „Die 3 wird geworfen" und B : „Eine ungerade Zahl wird geworfen" (blau gefärbte Würfelseite). Für die bedingte Wahrscheinlichkeit $P(A|B)$ gilt dann

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Dies ist die Wahrscheinlichkeit, dass eine 3 geworfen wird, *wenn* wir schon wissen, dass eine ungerade Zahl geworfen wurde. Die Wahrscheinlichkeiten $P(A|B)$ stehen an den Zweigen im Wahrscheinlichkeitsbaum, die vom Ereignis „B" ausgehen und zu den ungeraden Augenzahlen führen (Abbildung 3.7).

□

Bedingte Wahrscheinlichkeiten sind Wahrscheinlichkeiten für spezielle Situationen. Es gelten wieder entsprechende Rechenregeln.

Rechenregeln

$0 \leq P(A|B) \leq 1$	für jedes Ereignis A
$P(B|B) = 1$	für jedes Ereignis B
$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$	für A_1, A_2 disjunkt (d.h. $A_1 \cap A_2 = \emptyset$)
$P(\overline{A}|B) = 1 - P(A|B)$	für jedes Ereignis A

Im Allgemeinen ist $P(A|B)$ **nicht** gleich wie $P(B|A)$, da diese Wahrscheinlichkeiten von völlig verschiedenen Ereignissen sind [1]. So ist $P(A|B)$ im Beispiel vorher die Wahrscheinlichkeit, eine 3 zu werfen, wenn wir schon wissen, dass eine ungerade Zahl geworfen wurde. Auf der anderen Seite ist $P(B|A)$ hingegen die Wahrscheinlichkeit, dass wir ein ungerade Zahl werfen, wenn wir wissen, dass schon die Zahl 3 geworfen wurde. Diese Wahrscheinlichkeit ist natürlich 1, da 3 sowieso ungerade ist.

Wir wollen die bedingte Wahrscheinlichkeit nun an einem realistischen Beispiel untersuchen, das ein überraschendes Ergebnis hat.

Beispiel 3.5.2

Ein medizinischer Test auf eine (tödliche) Krankheit soll feststellen, ob eine Person an dieser Krankheit erkrankt ist oder nicht. Natürlich ist dieser Test nicht ganz genau. Manchmal zeigt er die Krankheit an, obwohl die Person gesund ist, oder er zeigt die Krankheit nicht an, obwohl die Person krank ist. Uns interessiert folgende Frage: Sie gehen zum Arzt und machen diesen Test auf die Krankheit. Der Test ist positiv: Sie haben gemäss dem Test die Krankheit. Wie gross ist die Wahrscheinlichkeit, dass Sie auch wirklich diese Krankheit haben?

Um diese Frage zu beantworten, führen wir folgende Bezeichnungen ein:

- D: Krankheit ist vorhanden; $\overline{D}$: Krankheit ist nicht vorhanden
- $+$: Test zeigt Krankheit an; $-$: Test zeigt Krankheit nicht an

Die Wahrscheinlichkeiten in Tabelle 3.4 sind durch Versuche bekannt. Z.B. ist die Wahrscheinlichkeit, dass die Krankheit vorhanden ist *und* der Test positiv ausfällt

$$P(D \cap +) = 0.009$$

Diese Wahrscheinlichkeit ist recht klein. Der Grund dafür liegt darin, dass nur ein kleiner Prozentsatz der Bevölkerung diese Krankheit hat.

	D	D^c
$+$	0.009	0.099
$-$	0.001	0.891

Tabelle 3.4.: Wahrscheinlichkeit für eine Krankheit.

Hier können wir verschiedene bedingte Wahrscheinlichkeiten berechnen:

- $P(+|D)$: W'keit, dass ein Kranker auch wirklich positiv getestet wird
- $P(+|\overline{D})$: W'keit, dass ein Gesunder fälschlicherweise positiv getestet wird

[1] Diese Verwechslung wird vor Gericht (und nicht nur da) so oft gemacht, dass er schon einen eigenen Namen hat: Prosecutor's fallacy.

- $P(-|D)$: W'keit, dass ein Kranker fälschlicherweise negativ getestet wird
- $P(-|\overline{D})$: W'keit, dass ein Gesunder richtigerweise negativ getestet wird
- $P(D|+)$: W'keit, dass ein positiv Getesteter auch wirklich krank ist
- $P(D|-)$: W'keit, dass ein negativ Getesteter fälschlicherweise krank ist
- $P(\overline{D}|+)$: W'keit, dass ein positiv Getesteter fälschlicherweise gesund ist
- $P(\overline{D}|-)$: W'keit, dass ein negativ Getesteter wirklich auch gesund ist

Wir berechnen zuerst die Wahrscheinlichkeit $P(+|D)$ wie folgt:

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9$$

Dabei haben wir für $P(D)$ folgende Tatsache benützt

$$P(D) = P(D \cap +) + P(D \cap -) = 0.009 + 0.001$$

Dies ist die Summe der Einträge in der Tabelle in der Spalte unter D. Die Kranken sind entweder positiv oder negativ getestet.

Die bedingte Wahrscheinlichkeit $P(-|\overline{D})$ ist die Wahrscheinlichkeit, dass eine gesunde Person auch wirklich negativ (gemäss Test ist Krankheit nicht vorhanden) getestet wird:

$$P(-|\overline{D}) = \frac{P(- \cap \overline{D})}{P(\overline{D})} = \frac{0.891}{0.891 + 0.099} = 0.9$$

Dieser Test scheint recht genau zu sein. Kranke Personen werden zu 90 % als positiv eingestuft, und gesunde Personen werden zu 90 % als negativ eingestuft.

Wir können die Fragestellung aber auch umkehren. Angenommen, Sie gehen zu einem Test, und dieser wird als positiv eingestuft. Wie gross ist die Wahrscheinlichkeit, dass Sie die Krankheit wirklich haben? Die meisten Leute würden 0.9 antworten. Müssen Sie sich also grosse Sorgen machen und das Testament schreiben oder einer Sterbehilfeorganisation beitreten?

Schauen wir uns nun die *richtige* Antwort an und dies ist die bedingte Wahrscheinlichkeit $P(D|+)$:

$$P(D|+) = \frac{P(+ \cap D)}{P(+)} = \frac{0.009}{0.009 + 0.099} = 0.08$$

Was bedeutet nun dieses Resultat? Die bedingte Wahrscheinlichkeit $P(D|+)$ ist die Wahrscheinlichkeit, dass man bei einem positiven Test auch wirklich krank ist. Diese beträgt aber nur 8 %. Man hat bei einem positiven Test also nur zu 8 % auch wirklich die Krankheit. Ein positiver Test sagt hier also sehr wenig darüber aus, ob man die Krankheit hat oder nicht.

□

Die Lektion ist: Bei bedingten Wahrscheinlichkeiten darf man seiner Intuition nicht trauen, sondern muss die entsprechenden Wahrscheinlichkeiten wirklich ausrechnen.

Bayes Theorem

Zwischen $P(A|B)$ und $P(B|A)$ gibt es folgenden Zusammenhang:

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

Die letzte Gleichung erhalten wir wegen

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

Beispiel 3.5.3

Das Bayes Theorem liefert die gleiche Lösung wie unsere obige Rechnung:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.9 \cdot (0.009 + 0.001)}{0.009 + 0.099} = \frac{0.009}{0.009 + 0.099} = 0.08.$$

□

Ein weiterer nützlicher Begriff ist die *totale Wahrscheinlichkeit* . Dabei wird eine Menge A zunächst in Mengen $A_1, \ldots, A_k$ unterteilt, die miteinander keine Schnittmenge haben und zusammen (Vereinigung) die ganze Menge A bilden. Eine solche Aufteilung nennen wir eine **Partitionierung**.

Beispiel 3.5.4

Für den Würfelwurf ist folgende Partitionierung möglich:

$$A_1 = \{1\}, \qquad A_2 = \{2,4\}, \qquad A_3 = \{3,5,6\}$$

Es gilt also

$$A_1 \cap A_2 = \emptyset; \quad A_1 \cap A_3 = \emptyset; \quad A_2 \cap A_3 = \emptyset$$

und

$$A_1 \cup A_2 \cup A_3 = A$$

□

Es gilt dann das

Gesetz der totalen Wahrscheinlichkeit

Für die Partitionierung $A_1, \ldots, A_k$ und für jedes beliebige Ereignis B gilt:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_k)P(A_k) = \sum_{i=1}^{k} P(B|A_k)P(A_k)$$

Das Gesetz der totalen Wahrscheinlichkeit lässt sich mit Hilfe eines Wahrscheinlichkeitsbaumes sehr gut verstehen (Abbildung 3.9) : Ein Ereignis B trete stets in Verbindung mit genau einem der sich paarweise ausschliessenden Ereignisse A_i ($i = 1, 2, \ldots, n$) auf, d.h. die Ereignisse A_i sind die möglichen Zwischenstationen auf dem Weg zum Ereignis B.

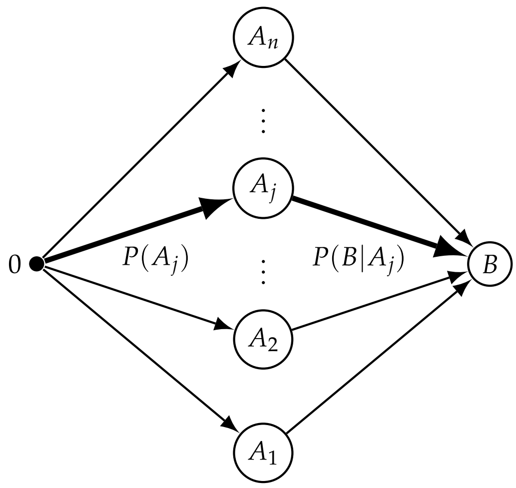

Abbildung 3.9.: Um die Wahrscheinlichkeit $P(B)$ zu berechnen, wird über alle $P(B|Aj) \cdot P(A_j)$ summiert.

Beispiel 3.5.5

Ich teile meine Emails in drei Kategorien ein:

$$A_1 : \text{„spam“}, \qquad A_2 : \text{„niedrige Priorität“}, \qquad A_3 : \text{„hohe Priorität“}$$

Aus früheren Beobachtungen weiss ich, dass

$$P(A_1) = 0.7, \qquad P(A_2) = 0.2, \quad \text{und} \quad P(A_3) = 0.1$$

Es gilt

$$P(A_1) + P(A_2) + P(A_3) = 1$$

wie es bei einer Partitionierung auch sein sollte.

Sei B das Ereignis, dass das Wort „free“ in der Email auftaucht. Dieses Wort kommt sehr oft in Spam-Mails vor, aber auch in den anderen. Von früheren Beobachtungen weiss ich, dass

$$P(B|A_1) = 0.9, \qquad P(B|A_2) = 0.01, \quad \text{und} \quad P(B|A_3) = 0.01$$

Hier ergibt die Summe nicht 1. Dies sind die Wahrscheinlichkeiten, mit denen das Wort „free“ in den drei Mailkategorien vorkommt. Angenommen, ich erhalte eine Email, die das Wort „free“ enthält. Wie gross ist die Wahrscheinlichkeit, dass es sich um Spam handelt? Das Bayes Theorem zusammen mit dem Gesetz der totalen Wahrscheinlichkeit liefert die Lösung:

$$\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\
&= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\
&= \frac{0.9 \cdot 0.7}{(0.9 \cdot 0.7) + (0.01 \cdot 0.2) + (0.01 \cdot 0.1)} \\
&= 0.995
\end{aligned}$$

Viele Spamfilter basieren tatsächlich auf diesem Prinzip: Die Mails werden nach Worten wie „free“, „credit“, etc. durchsucht, die häufig in Spam-Mails vorkommen, in den anderen aber eher nicht.

□