

Entwurf



DISKRETE STRUKTUREN

Institut für Informatik

Katharina Klost

Wintersemester 2024/25

Entwurf

Vorwort

Dieses Skript ist ein Entwurf. Es enthält sehr wahrscheinlich noch Fehler, über deren Meldung ich mich per Mattermost oder per Mail an katharina.klost@fu-berlin.de sehr freue.

Über dieses Skript

Dieses Skript ist entstanden als Begleitmaterial für die Veranstaltung *Diskrete Strukturen für Informatik*, welche Bestandteil des Bachelorstudiums Informatik und Bioinformatik an der Freien Universität Berlin ist. Die Zielgruppe des Skripts sind daher Studierende im ersten Semester ihres Studiums. Das Skript versucht dabei, den Übergang von der Mathematik, wie sie in der Schule gelehrt wird zur Mathematik an der Universität so weit wie möglich zu vereinfachen. Gerade in den vorderen Kapiteln wird dabei viel Wert darauf gelegt auch eine Intuition für die Konzepte zu geben. In den späteren Kapiteln wird dann mehr und mehr vorausgesetzt sich selber Beispiele erarbeiten zu können.

Mathematische Texte und hier ist dieses Skript keine Ausnahme sind häufig sehr „dicht“ geschrieben. Drei bis viel Seiten im Skript entsprechen häufig einer Vorlesung von 90 Minuten. Es ist also zu erwarten, dass Aussagen und Definitionen zum Teil mehrfach gelesen werden müssen. Im Anhang werden einige Hilfestellungen dazu gegeben, welche Techniken helfen können dieses Skript, oder auch andere mathematische Texte zu lesen.

Die Abschnitte des Skripts bauen zum Teil nicht linear aufeinander auf. Daher kann es auch sein, dass die Themen in der Vorlesung in einer anderen Reihenfolge behandelt werden. Daher ist am Anfang jedes größeren Abschnitts angegeben, welche anderen Abschnitte als Voraussetzung bekannt sein sollten. Grundlegende mathematische Fähigkeiten aus der Schule, wie Bruchrechnen, das Umformen von Gleichungen und ähnliches werden für das gesamte Skript vorausgesetzt. Einige der Grundlagen werden in ?? wiederholt.

Ebenso zu Beginn der größeren Abschnitte Lernziele angegeben. Diese geben an, welche Kompetenzen in Bezug auf die in den Abschnitten behandelten Inhalte von den Studierenden erwartet werden.

Gewisse Konzepte im Skript sind farblich kodiert. Definitionen von Begriffen finden sich meistens in grünen Boxen.

Definition 0.1. Dies ist eine Definition. Der definierte **Begriff** ist fett und grün angegeben.

Einige Begriffe werden nur im Fließtext definiert, diese sind dann auch **fett und grün** hervorgehoben.

Wichtige Aussagen, zu denen es Beweise gibt, finden sich in blauen Boxen. Die Beweise, also Argumente, dass diese Aussagen gelten finden sich dann meistens direkt unter den Boxen.

Klost:
muss
noch
ge-
schrie-
ben
werden

Entwurf

Satz 0.1. *Dies ist eine wichtige Aussage.*

Beweis. Dies ist der Beweis für die Aussage. □

Häufig wird es zu den Definitionen oder Aussagen Beispiele werden. Diese sind wie folgt gesetzt.

Beispiel. Dies ist ein Beispiel. ◀

Warum eigentlich Mathematik im Informatikstudium?

Ich wollte doch Informatik studieren und nicht Mathematik!

Es gibt verschiedene Gründe, warum es interessant und wichtig für Informatiker:innen ist, sich mit der Mathematik zu beschäftigen. Auf der einen Seite hat sich die Informatik historisch aus der Mathematik entwickelt. Viele Personen, die wir heute als frühe Informatiker:innen bezeichnen würden haben ursprünglich Mathematik studiert.

Zudem gibt es auch praktische Anwendungen der Mathematik in der Informatik. Aus der Informatik ergeben sich viele Fragestellungen, welche dann mit mathematischen Werkzeugen gelöst werden können und zu neuen Erkenntnissen sowohl in der Informatik als auch in der Mathematik führen können. Es gibt auch praktische Gesichtspunkte. Dadurch, dass die Mathematik schon viele Lösungen für mathematische Probleme gefunden haben, können diese bei passender Modellierung direkt für die Lösung von Problemen aus der echten Welt verwendet werden. Viele der Probleme der künstlichen Intelligenz und Informationssicherheit lassen sich auf mathematische Probleme zurückführen.

Inhaltsverzeichnis

Vorspann	i
Vorwort	ii
Inhaltsverzeichnis	v
I Aussagenlogik, Mengen und Beweise	1
1 Logik	2
1.1 Aussagenlogik	2
1.1.1 Einführung Aussagen	2
1.1.2 Von zusammengesetzten Aussagen zu logischen Termen	3
1.1.3 Eigenschaften von Termen und Boolesche Gesetze	7
1.2 Prädikatenlogik	10
1.3 Syntax der Aussagenlogik, Boolesche Algebra und Boolesche Funktionen	19
1.3.1 Syntax der Aussagenlogik	19
1.3.2 Boolesche Algebra	21
1.3.3 Interpretation eines Booleschen Terms	22
1.3.4 Boolesche Funktionen	23
1.3.5 Vollständige logische Signaturen	28
2 Beweistechniken	32
2.1 Allgemeines zu mathematischen Beweisen	32
2.2 Beweise von Implikationen	33
2.2.1 Direkter Beweis	34
2.2.2 Indirekte Beweise	35
2.3 Beweis durch vollständige Induktion	37
2.3.1 Charakterisierung der natürlichen Zahlen	37
2.3.2 Beweis durch vollständige Induktion	37
2.4 Das Schubfachprinzip	43
3 Mengen	47
3.1 Mengen	47
3.1.1 Einführung	47
3.1.2 Eigenschaften von Mengen	49

3.2	Relationen	54
3.2.1	Operationen auf Relationen	56
3.2.2	Äquivalenzrelationen	60
3.2.3	Halbordnungsrelationen und totale Ordnung	63
3.3	Funktionen	69
3.4	Abzählbarkeit von Mengen	76
 II Kombinatorik und Wahrscheinlichkeitsrechnung		83
4	Kombinatorik	84
4.1	Grundregeln und Inklusion-Exklusion	84
4.2	Permutationen	88
4.3	Binomialkoeffizienten	89
4.3.1	Monotone Gitterwege	93
4.3.2	Zahlpartitionen und geordnete Summendarstellungen	96
4.4	Doppeltes Abzählen	98
4.5	Zusammenfassung und eine Anwendung	100■
4.5.1	Ein Kartentrick	100■
5	Diskrete Wahrscheinlichkeitstheorie	102■
5.1	Grundlegende Begriffe und Beispiele	103■
5.2	Bedingte Wahrscheinlichkeiten	107■
5.2.1	Unabhängigkeit von Ereignissen	111■
5.2.2	Satz der totalen Wahrscheinlichkeit	114■
5.3	Zufallsvariablen und Erwartungswert	115■
5.3.1	Zufallsvariablen	115■
5.3.2	Erwartungswert	116■
5.4	Besondere Verteilungen	121■
5.4.1	Bernoulli-Verteilung	121■
5.4.2	Binomialverteilung	122■
5.4.3	Geometrische Verteilung	123■
 III Graphentheorie		125■
6	Graphen	126■
6.1	Grundlegende Definitionen	126■
6.2	Bäume	137■
6.3	Graphentraversierung	140■
6.3.1	Breitensuche	140■
6.3.2	Tiefensuche	144■
6.4	Gerichtete azyklische Graphen	147■
6.5	Planare Graphen	150■

Entwurf

Inhaltsverzeichnis

6.6	Eulerkreise und Eulerpfade	153■
	Anhang	156■
	Todo	156■

Entwurf

I

Aussagenlogik, Mengen und Beweise

1.1 Aussagenlogik

Lernziele

Die Studierenden ...

- ... entscheiden für einen gegebenen Satz, ob dieser eine Aussage ist.
- ... können den Wahrheitswert einer zusammengesetzten Aussage bestimmen
- ... können mithilfe der Äquivalenzregeln die Äquivalenz von zwei Aussagen feststellen



Anmerkung: Dieser Abschnitt gibt eine intuitive Einführung in die Aussagenlogik. Wir werden definieren, was eine Aussage ist und betrachten, wie Aussagen verknüpft werden können. Einen formelleren Zugang, bei dem die Syntax und Semantik eingeführt wird, gibt es dann in Abschnitt 1.3.



1.1.1 Einführung Aussagen


Dieser Abschnitt beschäftigt sich mit der Aussagenlogik. Zunächst einmal definieren wir das Konzept der *Aussage*.

Definition 1.1 (Aussage, nach Aristoteles). Eine **Aussage** ist ein Satz (sprachliches Gebilde) von dem es sinnvoll ist zu sagen, er sei entweder wahr oder falsch.

Aus dieser Definition ergibt sich direkt ein wichtiges Prinzip der Aussagenlogik, das **Zweiwertigkeitsprinzip**. Dieses besagt, dass eine Aussage immer entweder wahr oder falsch ist.

Beispiel. Im Folgenden betrachten wir einige sprachliche Gebilde und untersuchen, ob es sich jeweils um eine Aussage handelt oder nicht.

1. Die Sätze *Die Zahl 7 ist eine Primzahl*¹ und *Die Zahl 7 ist ungerade* sind Aussagen. Ihnen ist der Wahrheitswert *wahr* zugeordnet. Der Satz *Die Zahl 7 ist gerade* ist ebenfalls eine Aussage, sie hat jedoch den Wahrheitswert *falsch*.
2. Der Satz *Jede natürliche Zahl > 1 ist das Produkt von Primzahlen* ist eine wahre Aussage. Dass sie wirklich wahr ist, ist etwas aufwändiger zu zeigen, als die Aussagen im ersten Beispiel.
3. Der Satz *Jede gerade natürliche Zahl größer als 2 ist die Summe zweier Primzahlen* ist eine Aussage. Entweder gibt es eine natürliche Zahl, die sich nicht als Summe zweier Primzahlen darstellen lässt, dann wäre der Wahrheitswert *falsch*, oder es gibt keine solche Zahl, dann ist der Wahrheitswert *wahr*. Dieser Satz heißt Goldbach-Vermutung und es wird in der Mathematik davon ausgegangen, dass er wahr ist, es gibt aber bis jetzt keinen Beweis dafür.

Diese Aussage ist ein Beispiel für viele Sätze, die zwar Aussagen sind, von denen der Wahrheitswert aber (noch) nicht bekannt ist.
4. Der Satz *Der Mond ist aus Käse* ist eine falsche Aussage.
5. Der Satz *Sei ruhig!* ist keine Aussage, da ihr kein Wahrheitswert zugewiesen werden kann.
6. Der Satz *Dieser Satz ist falsch* ist als Russells Paradoxon bekannt und *keine Aussage*, da ihr kein eindeutiger Wahrheitswert zugewiesen werden kann. Wenn wir davon ausgehen, dass der Satz wahr ist, sagt der Satz aus, dass er falsch ist und umgekehrt. 

Aussagen der Form wie wir sie im Beispiel gesehen haben, kann in den Einzelwissenschaften (z.B. Physik, Biologie...) ein Wahrheitswert direkt zugewiesen werden. Solche Aussagen, nennt man auch **Elementaraussagen**.

Im Folgenden werden wir uns damit beschäftigen, wie neue Aussagen durch die Kombination von Elementaraussagen geschaffen werden können und wie über den Wahrheitswert dieser zusammengesetzten Aussagen argumentiert werden kann. Dafür ist das zweite zentrale Prinzip der Aussagenlogik wichtig, dass **Extensionalitätsprinzip**. Dieses besagt, dass der Wahrheitswert einer zusammengesetzten Aussage nur vom Wahrheitswert der Elementaraussagen abhängt.

1.1.2 Von zusammengesetzten Aussagen zu logischen Termen

Nun betrachten wir die Frage, wie Aussagen zu neuen Aussagen zusammengesetzt werden können. Dies geschieht durch verschiedene Verknüpfungen (logische **Junktionen**). Im Folgenden bezeichnen wir mit *Aussage* sowohl Elementaraussagen, als auch zusammengesetzte Aussagen.

¹**Erinnerung:** Eine Primzahl ist eine Zahl, die nur durch 1 und sich selber teilbar ist.

Definition 1.2 (Logische Junktoren). Seien α und β Aussagen.

Negation Die Negation „nicht α “ (kurz $\neg\alpha$) von α ist eine Aussage. $\neg\alpha$ ist wahr, wenn α falsch ist und $\neg\alpha$ ist falsch, wenn α wahr ist.

Beispiel. Die Negation zu Die Zahl 7 ist eine Primzahl ist Es gilt nicht, dass die Zahl 7 eine Primzahl ist oder einfacher Die Zahl 7 ist keine Primzahl.
Da die ursprüngliche Aussage wahr ist, ist die Negation falsch. ◀

Konjunktion Die Konjunktion „ α und β “ (kurz $\alpha \wedge \beta$) ist eine Aussage. $\alpha \wedge \beta$ ist wahr, wenn beide Aussagen wahr sind, sonst ist $\alpha \wedge \beta$ falsch.

Beispiel. Die Konjunktion von Die Zahl 7 ist eine Primzahl und Die Zahl 7 ist gerade ist Die Zahl 7 ist eine Primzahl und die Zahl 7 ist gerade.
Diese Konjunktion ist falsch, da die zweite Aussage falsch ist. ◀

Disjunktion Die Disjunktion „ α oder β “ (kurz $\alpha \vee \beta$) ist eine Aussage. $\alpha \vee \beta$ ist wahr, wenn mindestens eine der Aussagen wahr ist.^a

Beispiel. Die Disjunktion von Die Zahl 7 ist eine Primzahl und Die Zahl 7 ist gerade ist Die Zahl 7 ist eine Primzahl oder die Zahl 7 ist gerade.
Diese Disjunktion ist wahr, da die erste Aussage wahr ist.

Die Disjunktion von Die Zahl 7 ist eine Primzahl und Die Zahl 7 ist ungerade ist Die Zahl 7 ist eine Primzahl oder die Zahl 7 ist ungerade.
Diese Disjunktion ist wahr, da in diesem Fall sogar beide Aussagen wahr sind. ◀

Antivalenz Die Antivalenz „Entweder α oder β “ (kurz $\alpha \oplus \beta$) ist eine Aussage. Die Antivalenz wird manchmal auch als $\alpha \dot{\vee} \beta$ beschrieben. $\alpha \oplus \beta$ ist wahr, wenn genau eine der beiden Aussagen wahr ist.


Beispiel. Die Antivalenz von Die Zahl 7 ist eine Primzahl und Die Zahl 7 ist gerade ist Entweder die Zahl 7 ist eine Primzahl oder die Zahl 7 ist gerade.
Die Antivalenz der Aussagen ist wahr, da die erste Aussage wahr und die zweite Aussage falsch ist. ◀

Implikation Die Implikation „Aus α folgt β “ bzw. „Wenn α dann β “ (kurz $\alpha \rightarrow \beta$) ist eine Aussage. Die Implikation wird manchmal auch als $\alpha \Rightarrow \beta$ geschrieben. $\alpha \rightarrow \beta$ ist wahr, wenn α und β beide wahr sind, oder wenn α falsch ist.^b

Beispiel. Die Implikation von Die Zahl 7 ist eine Primzahl nach Die Zahl 7 ist gerade ist Wenn 7 eine Primzahl ist, dann ist 7 gerade.
Diese Implikation ist falsch, da die zweite Aussage falsch ist.

Die Implikation von Die Zahl 7 ist gerade nach Die Zahl 7 ist eine Primzahl ist Wenn 7 gerade ist, dann ist 7 eine Primzahl.
Diese Implikation ist richtig, da die erste Aussage falsch ist. ◀

Äquivalenz Die Äquivalenz „ α genau dann, wenn β “ (kurz $\alpha \leftrightarrow \beta$) ist eine Aussage. Die Äquivalenz wird manchmal auch als $\alpha \Leftrightarrow \beta$ geschrieben. $\alpha \leftrightarrow \beta$ ist wahr, wenn α und β beide wahr sind, oder wenn α und β beide falsch sind.

Beispiel. Die Äquivalenz von *Die Zahl 7 ist eine Primzahl* und *Die Zahl 7 ist gerade* ist *7 ist eine Primzahl* ist genau dann, wenn 7 gerade ist. Diese zusammengesetzte Aussage ist falsch, da die erste Aussage wahr, die zweite Aussage aber falsch ist. 

^aDies entspricht nicht Gebrauch von *oder* in der alltäglichen Sprache, dort wird *oder* meistens als *entweder oder* interpretiert.

^bAchtung: Bei der Implikation ist die Reihenfolge, in der die Aussagen betrachtet werden, wichtig. Details zu dieser Definition gibt es weiter unten

Während einige der oben eingeführten Junktoren in ihrer Bedeutung mit der Alltagssprache übereinstimmen, führen die Disjunktion (oder) sowie die Unterscheidung der Implikation und der Äquivalenz öfter zu Schwierigkeiten.

Wie schon in der Fußnote kurz angemerkt, wird mit der Disjunktion ein sogenanntes *inklusive oder* gemeint, es wird also auch der Fall, dass beide Teilaussagen wahr sind zu wahr ausgewertet. Das umgangssprachlich oft verwendete *exklusive oder* schließt diesen Fall aber explizit aus.

Etwas komplizierter ist die Denkweise hinter der Definition der Implikation, insbesondere der Fall, wenn α eine falsche Aussage ist. Schauen wir uns ein konkretes Beispiel an. Ein Dozierender trifft die Aussage:

Wenn Sie 100% der Punkte in der Klausur bekommen, dann bekommen Sie eine 1.0.

Wenn ein Studierender 100% der Punkte in der Klausur erreicht erwartet er, dass er eine 1,0 als Note bekommt, sonst kann er sich beim Dekanat beschweren. Was ist nun, wenn nicht 100% der Punkte erreicht werden? Dann ist es möglich, dass der Studierende eine 1,0 bekommt, weil er trotzdem genug Punkte erreicht hat, ebenso kann es sein, dass eine schlechtere Note erreicht wird. In beiden Fällen ist die ursprüngliche Aussage wahr.

In diesem Kontext ist auch die Abgrenzung zur Äquivalenz wichtig. Die Äquivalenz der Aussagen aus dem Beispiel von oben ist *Sie bekommen eine 1,0 genau dann, wenn Sie 100% der Punkte erreichen.*² Diese Aussage schließt den Fall aus, dass weniger als 100% der Punkte erreicht werden, aber trotzdem eine 1,0 erreicht werden kann. Ein realistischeres Beispiel für eine Äquivalenz wäre

Sie bestehen die Klausur, genau dann, wenn Sie mehr als 50% der Punkte erreichen.

In diesem Fall können Sie davon ausgehen, dass die Klausur bestanden ist, wenn Sie mehr als 50% der Punkte erreicht haben. Gleichzeitig können Sie davon ausgehen, dass wenn Sie die Klausur bestanden haben, Sie mehr als 50% der Punkte erreicht haben.

²Die Reihenfolge der Aussage kann hier für ein besseres Leseverständnis umgedreht werden, da die Reihenfolge bei der Äquivalenz egal ist.

Definition 1.3 (Stelligkeit von Junktoren). Die **Stelligkeit** eines Junktors gibt an, wie viele Aussagen mit diesem verknüpft werden. Werden k Aussagen verknüpft spricht man von einem k -stelligen Junktor. Hierbei werden Elementaraussagen auch als 0-stellige Junktoren bezeichnet.

Beispiel. Die Negation ist ein 1-stelliger Junktor, alle anderen genannten Junktoren sind 2-stellig. Die immer falsche Aussage *false* und die immer wahre Aussage *true* sind 0-stellige Junktoren. ◀

Hierbei sind wir natürlich nicht darauf beschränkt einmalig einen der Junktoren anzuwenden, da ja eine zusammengesetzte Aussage auch wieder eine Aussage ist.

Beispiel. Der Satz „(8 ist eine Primzahl und 8 ist gerade) oder 7 ist ungerade“ ist also eine zusammengesetzte (wahre) Aussage. ◀

Selbsttest: Bestimmen Sie den Wahrheitsgehalt der folgenden Aussagen:

- (a) 8 ist eine Primzahl oder 8 ist gerade.
- (b) Wenn 9 durch 4 teilbar ist, dann ist 9 gerade.
- (c) Wenn 9 durch 4 teilbar ist, dann ist 9 ungerade.

In der Aussagenlogik interessieren wir uns mehr für das Verhalten von zusammengesetzten Aussagen, als für die Elementaraussagen selber. Wir wollen also die Strukturen analysieren, die sich ergeben, wenn Aussagen zusammengesetzt werden. Daher werden wir im Folgenden die Elementaraussagen durch Variablen ersetzen, denen jeweils wahr oder falsch als Wahrheitswert zugewiesen werden kann. Statt Aussage nennen wir diese Konstrukte auch **logische Terme**. Um die Darstellung zu verkürzen, schreiben wir im Folgenden meistens 1 und 0 statt wahr und falsch.

Definition 1.4 (Belegung, Auswertung).³ Eine feste Zuweisung von Wahrheitswerten zu allen Variablen in einem Term, nennt man auch eine **Belegung** der Variablen. Wird dem Term dann Schritt für Schritt nach den oben genannten Regeln ein Wahrheitswert zugewiesen nennt man dies die **Auswertung** des Terms bezüglich einer festen Belegung.

Beispiel. Wir schauen uns nun Beispiele von Termen an, sowie eine Belegung und Auswertung zu einem der Terme.

Angenommen a, b, c und d sind Variablen. Dann sind die folgenden beide Konstrukte logische Terme:


$$(a \wedge b) \rightarrow (a \vee b)$$

$$(a \wedge b) \vee (c \wedge d)$$

³Dies ist eine eher informelle Definition von Belegung und Auswertung, eine formellere Definition findet sich in Abschnitt 1.3

a	b	$a \wedge b$	$a \vee b$	$a \oplus b$	$a \rightarrow b$	$a \leftrightarrow b$
0	0	0	0	0	1	1
0	1	0	1	1	1	0
1	0	0	1	1	0	0
1	1	1	1	0	1	1

Tabelle 1.1: Wahrheitstabelle für die eingeführten logischen Junktoren

$a = 0, b = 1, c = 1, d = 1$ ist eine Belegung der Variablen des Terms $(a \wedge b) \vee (c \wedge d)$. Der Term wird mit dieser Belegung zu 1 ausgewertet. 

Mithilfe der Begriffe der Belegung und der Auswertung, können wir nun eine kompaktere Darstellung für den Wahrheitswert von zusammengesetzten Aussagen angeben, die sogenannte **Wahrheitstabelle**. In einer Wahrheitstabelle werden alle möglichen Belegungen der Aussagen in einer zusammengesetzten Aussage aufgezählt und der entsprechende Wahrheitswert dazu angegeben. Die Wahrheitstabelle für die oben eingeführten Junktoren findet sich in Tabelle 1.1.

Selbsttest:

1. Wählen Sie eine andere Belegung für die Terme aus dem Beispiel oben und werten Sie den Term bezüglich dieser Belegung aus.
2. Füllen Sie die Wahrheitstabelle für den Term $a \wedge (b \vee a)$ aus.



1.1.3 Eigenschaften von Termen und Boolesche Gesetze

Definition 1.5 (Erfüllbarkeit, Tautologie, Kontradiktion, Äquivalenz).

Erfüllbarkeit Ein Term, für den es (mindestens) eine Belegung gibt, die zu 1 auswertet ist **erfüllbar**.

Widerlegbar Ein Term, für den es (mindestens) eine Belegung gibt, die zu 0 auswertet ist **widerlegbar**.

Tautologie Ein Term bei dem alle möglichen Belegungen zu 1 auswerten, nennt man **Tautologie**.

Kontradiktion Eine Term, bei dem alle möglichen Belegungen zu 0 auswerten, nennt man **Kontradiktion**.

Äquivalenz Zwei Terme sind **äquivalent**, wenn alle möglichen Belegungen zum gleichen Wahrheitswert ausgewertet werden. Wenn α und β äquivalent sind, schreiben wir auch $\alpha \equiv \beta$.


Wir gehen hier kurz einmal auf den Unterschied zwischen \equiv und \leftrightarrow ein. Das Zeichen \equiv bezeichnet keinen logischen Junktor, sondern ist ähnlich zu lesen, wie das Gleichheitszeichen beim Umformen von arithmetischen Ausdrücken.⁴

Beispiel. Wir betrachten die folgenden Terme:

$$\begin{aligned} t_1 &:= (a \wedge b) \\ t_2 &:= (a \wedge (b \wedge \neg b)) \\ t_3 &:= (a \wedge b) \wedge (b \vee a) \\ t_4 &:= (a \vee b) \vee (\neg a) \end{aligned}$$

Wir betrachten die Wahrheitstabellen für diese drei Terme, um dann zu entscheiden, welche Terme erfüllbar sind, welche Kontradiktionen oder Tautologien sind, und ob es Terme gibt, die äquivalent sind.

a	b	t_1	t_2	t_3	t_4
0	0	0	0	0	1
0	1	0	0	0	1
1	0	0	0	0	1
1	1	1	0	1	1

Ein Blick auf die Tabelle verrät uns, dass t_1, t_3 und t_4 jeweils bei mindestens einer Belegung zu 1 auswerten, diese Terme sind also erfüllbar. Bei t_2 werten sogar alle Belegungen zu 1 aus, dieser Term ist also eine Tautologie. Auf der anderen Seite gibt es für t_2 keine Belegung, die zu 1 auswertet, dieser Term ist also eine Kontradiktion. Die Terme t_1 und t_3 werden für alle möglichen Belegungen gleich ausgewertet, sind also äquivalent. 

Im Beispiel haben wir gesehen, dass das Ausfüllen einer Wahrheitstabelle eine Möglichkeit ist, um zu zeigen, dass zwei Terme äquivalent sind. Dies ist allerdings für Terme mit vielen Variablen sehr Zeitaufwändig, warum zeigt die folgende Beobachtung:

Beobachtung 1.1. Die Anzahl der möglichen Belegung von n Variablen mit Wahrheitswerten ist 2^n .

Beweis. Wir nennen die Variablen x_1, \dots, x_n . Für x_1 gibt es zwei mögliche Belegungen. Für jede dieser beiden Belegungen gibt es dann wieder zwei mögliche Belegungen für x_2 und so weiter. Insgesamt gibt es also $\underbrace{2 \cdot 2 \cdot \dots \cdot 2}_{\times n} = 2^n$ mögliche Belegungen der

Variablen. □

⁴Ein arithmetischer Ausdruck ist ein Ausdruck, der Zahlen und Rechenoperationen verwendet z.B. $4 + 5 \cdot 10$.

Um die logische Äquivalenz auch von komplexeren Termen zu zeigen, gibt es eine Reihe von äquivalenten Termen, die besonders wichtig sind.

Satz 1.2. Für beliebige Terme α, β, γ gelten die folgenden Äquivalenzen:

$$\begin{aligned}
 (\alpha \wedge \beta) \wedge \gamma &\equiv \alpha \wedge (\beta \wedge \gamma) \\
 (\alpha \vee \beta) \vee \gamma &\equiv \alpha \vee (\beta \vee \gamma) && \text{(Assoziativität)} \\
 \alpha \wedge \beta &\equiv \beta \wedge \alpha \\
 \alpha \vee \beta &\equiv \beta \vee \alpha && \text{(Kommutativität)} \\
 \alpha \wedge (\beta \vee \gamma) &\equiv (\alpha \wedge \beta) \vee (\alpha \wedge \gamma) \\
 \alpha \vee (\beta \wedge \gamma) &\equiv (\alpha \vee \beta) \wedge (\alpha \vee \gamma) && \text{(Distributivität)} \\
 \alpha \wedge \alpha &\equiv \alpha \\
 \alpha \vee \alpha &\equiv \alpha && \text{(Idempotenz)} \\
 \alpha \wedge (\alpha \vee \beta) &\equiv \alpha \\
 \alpha \vee (\alpha \wedge \beta) &\equiv \alpha && \text{(Absorption)} \\
 \neg(\alpha \wedge \beta) &\equiv \neg\alpha \vee \neg\beta \\
 \neg(\alpha \vee \beta) &\equiv \neg\alpha \wedge \neg\beta && \text{(deMorgansche Regel)} \\
 \neg\neg\alpha &\equiv \alpha && \text{(doppelte Negation)} \\
 \alpha \rightarrow \beta &\equiv \neg\alpha \vee \beta \\
 \alpha \rightarrow \beta &\equiv \neg\beta \rightarrow \neg\alpha && \text{(Kontraposition)} \\
 \alpha \leftrightarrow \beta &\equiv (\alpha \wedge \beta) \vee (\neg\alpha \wedge \neg\beta) \\
 \alpha \rightarrow (\beta \wedge \gamma) &\equiv (\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma) \\
 \alpha \rightarrow (\beta \vee \gamma) &\equiv (\alpha \rightarrow \beta) \vee (\alpha \rightarrow \gamma) \\
 (\alpha \wedge \beta) \rightarrow \gamma &\equiv (\alpha \rightarrow \gamma) \vee (\beta \rightarrow \gamma) \\
 (\alpha \vee \beta) \rightarrow \gamma &\equiv (\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)
 \end{aligned}$$

Beweis. Die Beweise können alle mit Hilfe von Wahrheitstabellen geführt werden. Als Beispiel zeigen wir hier die erste Version der Assoziativität:

α	β	γ	$\alpha \wedge \beta$	$\beta \wedge \gamma$	$(\alpha \wedge \beta) \wedge \gamma$	$\alpha \wedge (\beta \wedge \gamma)$
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	1	0	0	0	0	0
0	1	1	0	1	0	0
1	0	0	0	0	0	0
1	0	1	0	0	0	0
1	1	0	1	0	0	0
1	1	1	1	1	1	1

□

Selbsttest: Zeigen Sie für zwei weitere Äquivalenzen deren Äquivalenz mit Hilfe einer Wahrheitstabelle.



Mit Hilfe von Satz 1.2 können wir nun für größere Terme einfach zeigen, dass diese äquivalent sind. Hierfür betrachten wir Teile des Terms wie eine Variable um die Regeln anwenden zu können.

Beispiel. Die für die Anwendung der Regel relevanten Teile sind farblich markiert.

$$\begin{aligned} & a \vee ((b \vee c) \wedge \neg(\neg a \wedge (\neg a \vee d))) \\ \text{Absorption} & \equiv (a \vee ((b \vee c) \wedge \neg(\neg a))) \\ \text{doppelte Negation} & \equiv a \vee ((b \vee c) \wedge a) \\ \text{Kommutativität} & \equiv a \vee (a \wedge (b \vee c)) \\ \text{Absorption} & \equiv a \end{aligned}$$



Am Anfang wird es einige Zeit dauern, bis die richtigen Regeln zum Umformen gefunden werden. Dies ist ähnlich zum Umformungen von (arithmetischen) Gleichungen in der Schule reine Übungssache.

Um zu überprüfen, ob ein Term erfüllbar, eine Tautologie oder eine Kontradiktion ist, gibt es kein allgemeines effizientes Verfahren. In ?? werden wir uns mit dem Resolutionskalkül beschäftigen. Dies ist ein Verfahren, mit dem ein Term darauf getestet werden kann, ob er nicht erfüllbar ist. In ?? werden wir jedoch auch die Grenzen dieses Verfahrens kennenlernen.

Klost: Der Teil zum Resolutionskalkül ist noch nicht im Skript und wird voraussichtlich auch nicht in der Vorlesung behandelt werden.

1.2 Prädikatenlogik

Voraussetzung:

- Grundlagen der Aussagenlogik (Abschnitt 1.1)
- natürliche, ganzen, rationalen und reellen Zahlen ($\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$)



Lernziele

Die Studierenden ...

- ... können für ein gegebenes Konstrukt entscheiden, ob dieses eine Aussage, ein Prädikat oder etwas anderes ist
- ... können quantifizierte Aussagen lesen und selber schreiben



- ... können logische Äquivalenzen für prädikatenlogische Aussagen anwenden
- ... können Aussagen in Pränexform und Negationsnormalform umwandeln
- ... führen Beweise mithilfe des „Gegenspieler-Beweiser Prinzips“

Eines der ersten Beispiele für Aussagen, welches wir uns angeschaut haben, war *Die Zahl 7 ist eine Primzahl*. In diesem Abschnitt schauen wir uns an was passiert, wenn wir Variablen in die Elementaraussagen hinzufügen. Ein Beispiel wäre *Die Zahl x ist eine Primzahl*. Dieser Satz ist nun keine Aussage mehr, da ohne zu wissen welcher Wert für x eingesetzt wird, der Wahrheitswert nicht entschieden werden kann. Wird nun aber 7 für x eingesetzt, erhalten wir eine (wahre) Aussage, wird 8 eingesetzt, ergibt sich eine falsche Aussage. Diese Art von Konstrukten mit Platzhaltern wird uns sehr häufig begegnen.

Definition 1.6 (Prädikat). Ein **Prädikat** $P(x_1, \dots, x_n)$ über den Universen^a U_1, \dots, U_n ist ein Satz mit freien Variablen x_1, \dots, x_n der zur Aussage wird, wenn jedes x_i durch einen konkreten Wert aus U_i ersetzt wird.

^aEin Universum kann man sich einfach als alle möglichen Werte, die eine Variable annehmen kann vorstellen. Häufig verwendete Universen sind $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$

Beispiel. $P(x) : x > 1$ und $Q(x) : x + 0 = x$ sind Prädikate über den ganzen Zahlen. Sie haben noch keinen Wahrheitswert, solange x kein konkreter Wert zugewiesen wurde. $P(0)$ ist eine falsche Aussage, $Q(10)$ ist eine wahre Aussage. $R(x, y) : x + y = x$ ist ebenfalls ein Prädikat über den ganzen Zahlen. Auch diese hat keinen Wahrheitswert, solange x und y kein Wert aus den ganzen Zahlen zugewiesen wurde. $R(42, y)$ ist keine Aussage, da kein Wert für y eingesetzt wurde. $R(42, 0)$ ist eine wahre Aussage. ◀

Selbsttest: Welche der folgenden Konstrukte sind Prädikate, welche Aussagen, welche nicht gültig? Wenn die Konstrukte Aussagen sind, was ist der Wahrheitswert?

1. $P(x, y) : x > y$ mit $x \in \mathbb{N}$ und $y \in \mathbb{Z}$.
2. $P(x, 8)$
3. $P(0, -3)$
4. $P(-3, 0)$

Oft interessiert uns bei einem Prädikat nicht der Wahrheitswert, wenn ein konkreter Wert eingesetzt wird, sondern wir verwenden Prädikate, um allgemeine Aussagen zu formulieren. Dies geschieht mithilfe von sogenannten **Quantoren**. Wir betrachten hier die beiden am meisten verwendeten Quantoren.

Definition 1.7 (Allquantor und Existenzquantor).

Allquantor Ein Satz der Form Für jedes konkrete $x \in U$ gilt $P(x)^a$ ist eine Aussage. Formell wird der sogenannte Allquantor \forall verwendet, um dies darzustellen. In der mathematischen Notation können wir den Satz also schreiben als:

$$\forall x \in U : P(x)$$

Existenzquantor Ein Satz der Form Es gibt ein $x \in U$, sodass $P(x)$ gilt ist eine Aussage. Formell wird der sogenannte Existenzquantor \exists verwendet, um dies darzustellen. In der mathematischen Notation können wir den Satz also schreiben als:

$$\exists x \in U : P(x)$$

^a $x \in U$ bedeutet, dass x ein Wert aus U zugewiesen wird

Wir sagen auch, dass die Variable x vom Quantor **gebunden** wird. Eine Variable, die nicht gebunden ist, nennt man eine **freie** Variable. In $P(x)$ ist x also eine freie Variable, in $\exists x : P(x)$ ist x gebunden. Ein erstes Beispiel eines Prädikats, das mithilfe eines Quantors zu einer Aussage wurde, haben wir schon in Abschnitt 1.1 gesehen. Dort hatten wir die Goldbach-Vermutung (Jede gerade natürliche Zahl größer als zwei lässt sich als Summe von Primzahlen schreiben) kennengelernt. Diese kann man auch als „Für alle natürlichen Zahlen n gilt: $(n > 2 \wedge n \text{ gerade}) \rightarrow n$ ist Summe von Primzahlen“ schreiben.

Beispiel. Wir schauen uns einige weitere Beispiele für Aussagen mit Quantoren an.

- $\forall x \in \mathbb{N} : x + 0 = x$ ist eine wahre Aussage.
- $\exists x \in \mathbb{N} : x^2 = x$ ist eine wahre Aussage.
- $\exists x \in \mathbb{N} : x + 1 = x$ ist eine falsche Aussage.
- $\forall x \in \mathbb{N} : x^2 = x$ ist eine falsche Aussage.

Selbsttest: Welche der folgenden Aussagen sind wahr, welche sind falsch?

1. $\forall x \in \mathbb{Z} : x + (-x) = 0$
2. $\exists x \in \mathbb{N} : x + 1 \geq 10$
3. $\forall x \in \mathbb{N} : x + 1 \geq 10$

Ähnlich wie für logische Terme, können wir für Aussagen mit Quantoren Äquivalenzen definieren:

Satz 1.3. Für beliebige Prädikate $P(x)$ und $Q(x)$ und $R(x, y)$ gelten die folgenden Äquivalenzen.

$$\neg(\forall x : P(x)) \simeq \exists x : \neg P(x)$$

$$\neg(\exists x : P(x)) \simeq \forall x : \neg P(x)$$

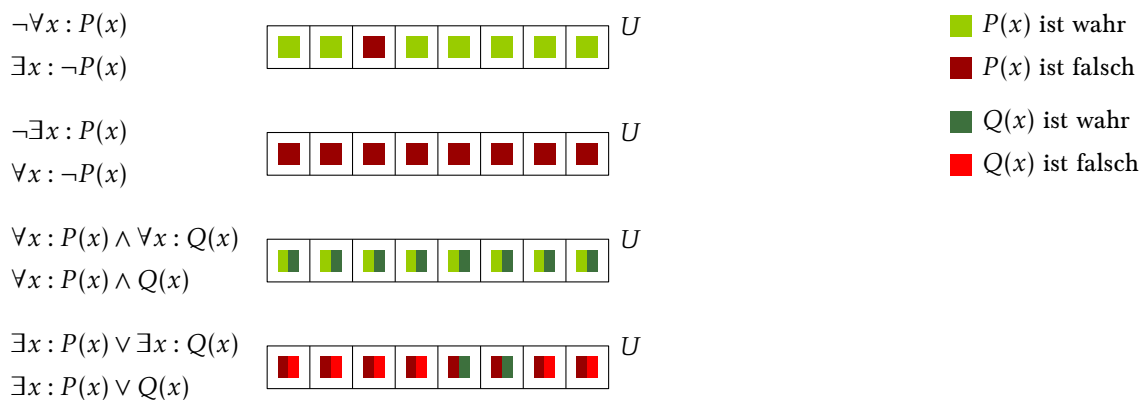
$$\forall x : P(x) \wedge \forall x : Q(x) \simeq \forall x : (P(x) \wedge Q(x))$$

$$\exists x : P(x) \vee \exists x : Q(x) \simeq \exists x : (P(x) \vee Q(x))$$

$$\forall x \forall y : R(x, y) \simeq \forall y \forall x : R(x, y)$$

$$\exists x \exists y : R(x, y) \simeq \exists y \exists x : R(x, y)$$

Beweis. Wir verzichten auf einen formellen Beweis der Aussage. Stattdessen stellen wir die ersten vier Aussagen grafisch dar. Die Darstellung der verbleibenden Aussagen ist eine einfache Übung.



□

Achtung es gibt auch einige Formelpaare, die nicht äquivalent sind:

Satz 1.4. Für beliebige Prädikate $P(x)$ und $Q(x)$ sind die folgenden Paare von Aussagen im Allgemeinen nicht äquivalent:

$$\forall x : P(x) \vee \forall x : Q(x) \not\simeq \forall x : (P(x) \vee Q(x))$$

$$\exists x : P(x) \wedge \exists x : Q(x) \not\simeq \exists x : (P(x) \wedge Q(x))$$

Beweis. Da wir zeigen wollen, dass die Paare im Allgemeinen nicht äquivalent sind, reicht es ein zwei Prädikate anzugeben, für die die Aussage nicht gilt.⁵ Wir wählen

⁵Dies nennt man auch ein Gegenbeispiel

Aussage	Wann ist die Aussage wahr?	Wann ist die Aussage falsch?
$\forall x \forall y P(x, y)$	Wenn $P(x, y)$ für jedes Paar x, y wahr ist	Wenn es mindestens eine Paar x, y gibt, für das $P(x, y)$ falsch ist
$\forall y \forall x P(x, y)$		
$\forall x \exists y P(x, y)$	Für jedes x gibt es ein y , sodass $P(x, y)$ wahr ist.	Es gibt ein x , sodass $P(x, y)$ für jedes y falsch ist.
$\exists x \forall y P(x, y)$	Es gibt ein x , sodass $P(x, y)$ für jedes y wahr ist.	Für jedes x gibt es ein y , sodass $P(x, y)$ falsch ist.
$\exists x \exists y P(x, y)$	Wenn es ein Paar x, y gibt, für das $P(x, y)$ wahr ist	Wenn $P(x, y)$ für alle Paare x, y falsch ist
$\exists y \exists x P(x, y)$		

Tabelle 1.2: Beschreibung für verschachtelte Quantoren

$P(x)$: x ist ein Werktag⁶ und $Q(x)$: x ist ein Tag am Wochenende über dem Universum aller möglichen Wochentage.

Dann ist $\forall x : P(x) \vee \forall x : Q(x)$ eine falsche Aussage, da nicht jeder Tag am Wochenende ist und nicht jeder Tag ein Werktag. Die Aussage $\forall x : (P(x) \vee Q(x))$ ist aber wahr, da jeder Tag ein Werktag oder ein Tag am Wochenende ist.

Ähnlich ist $\exists x : P(x) \wedge \exists x : Q(x)$ eine wahre Aussage, da es mindestens einen Werktag und einen Tag am Wochenende gibt. Umgekehrt ist $\exists x : (P(x) \wedge Q(x))$ eine falsche Aussage, da es keinen Tag gibt, der Werktag und Wochenende ist. \square

Quantoren können auch geschachtelt werden. Wenn $P(x, y)$ ein Prädikat über zwei Variablen ist, dann ist $Q(x) := \forall y : P(x, y)$ wieder ein Prädikat, allerdings eines über einer Variablen. $\forall x : Q(x)$ ist dann eine Aussage, die auch als $\forall x \forall y P(x, y)$ geschrieben werden kann. Tabelle 1.2 gibt einen Überblick darüber, wie verschiedene Aussagen mit verschachtelten Quantoren zu lesen sind. In Abbildung 1.1 findet sich eine grafische Darstellung der Aussagen.

Achtung: Die Aussagen $\forall x \forall y : P(x, y)$ und $\forall y \forall x : P(x, y)$ sind äquivalent. Dies gibt jedoch nicht für die Aussagen $\forall x \exists y : P(x, y)$ und $\exists y \forall x : P(x, y)$. Ähnlich sind $\exists x \exists y : P(x, y)$ und $\exists y \exists x : P(x, y)$ äquivalent, nicht aber $\exists x \forall y : P(x, y)$ und $\forall y \exists x : P(x, y)$.

Beispiel.

- $\forall x \in \mathbb{N} \forall y \in \mathbb{N} : x \geq -y$ ist wahr, da die Aussage für alle Paare von natürlichen Zahlen gilt.
- $\forall x \in \mathbb{N} \forall y \in \mathbb{N} : x \geq y$ ist falsch, da die Aussage für $x = 1$ und $y = 2$ nicht gilt.
- $\forall x \in \mathbb{N} \exists y \in \mathbb{N} : x \geq y$ ist wahr, da für jedes feste $x \in \mathbb{N}$ der Wert $y = x$ gewählt werden kann.
- $\forall x \in \mathbb{N} \exists y \in \mathbb{N} : x > y$ ist falsch, da es für $x = 0$ kein $y \in \mathbb{N}$ gibt, dass echt kleiner als x ist.

⁶Mit Werktag meinen wir hier einen Tag von Montag bis Freitag.

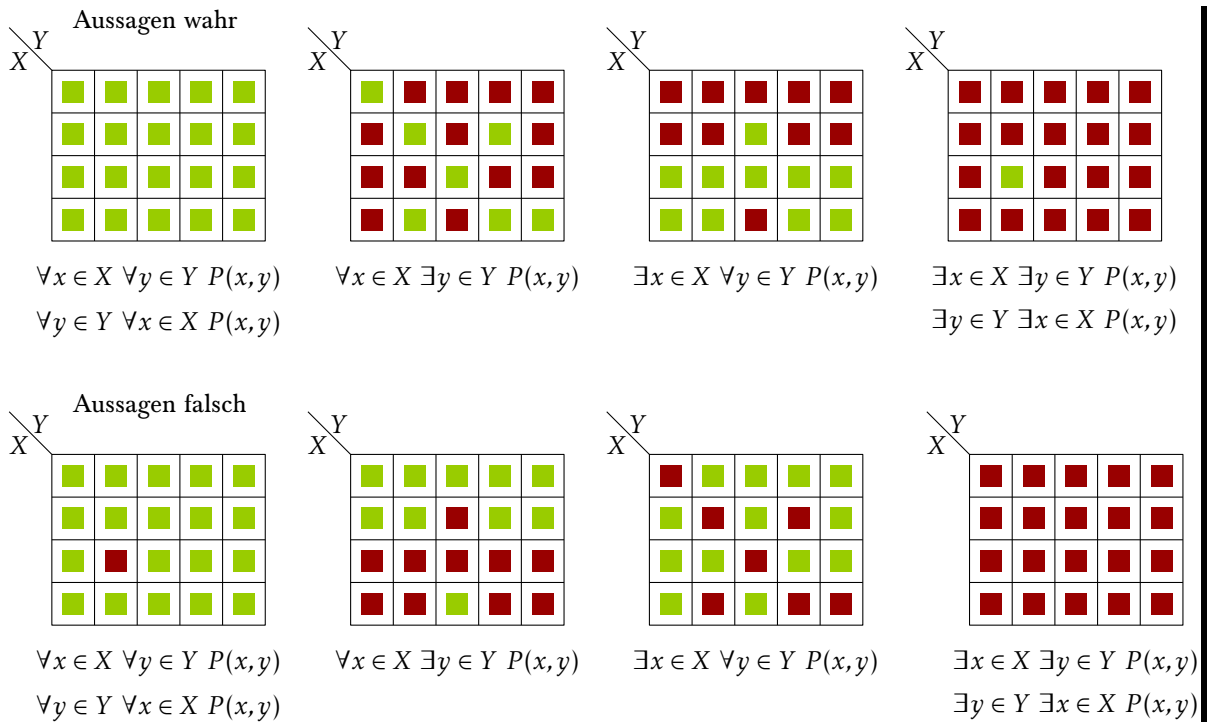


Abbildung 1.1: Grafische Darstellung von verschachtelten Quantoren

- $\exists x \in \mathbb{N} \forall y \in \mathbb{N} : x \leq y$ ist wahr, da für $x = 0$ alle anderen Zahlen in \mathbb{N} mindestens genauso groß sind.
- $\exists x \in \mathbb{N} \forall y \in \mathbb{N} : x \geq y$ ist falsch, da für alle $x \in \mathbb{N}$ die Zahl $y = x + 1$ immer größer ist.
- $\exists x \in \mathbb{N} \exists y \in \mathbb{N} : x \geq y$ ist wahr, ein Beispiel ist $x = 2$ und $y = 1$.
- $\exists x \in \mathbb{N} \exists y \in \mathbb{N} : x < -y$ ist falsch, da für jedes Paar $x \in \mathbb{N}, y \in \mathbb{N}$ gilt, dass $x \geq 0$ aber $-y < 0$ ist. ◀

Die Regeln für die Negation von quantifizierten Aussagen können auch auf verschachtelten Quantoren angewendet werden.

Korollar 1.5. Es gelten die folgenden logischen Äquivalenzen

$$\begin{aligned} \neg(\forall x \forall y P(x, y)) &\simeq \exists x \exists y \neg P(x, y) \\ \neg(\forall x \exists y P(x, y)) &\simeq \exists x \forall y \neg P(x, y) \\ \neg(\exists x \forall y P(x, y)) &\simeq \forall x \exists y \neg P(x, y) \\ \neg(\exists x \exists y P(x, y)) &\simeq \forall x \forall y \neg P(x, y) \end{aligned}$$

Beweis. Ein Blick auf Abbildung 1.1 gibt uns schon ein Indiz dafür, dass die Aussage stimmt. Wir führen den formellen Beweis für die erste Aussage.

$$\neg(\forall x \forall y P(x, y)) \simeq \exists x \neg(\forall y P(x, y)) \simeq \exists x \exists y \neg P(x, y)$$

Die Beweise für die anderen Aussagen werden genauso geführt. □

Längere verschachtelte, quantifizierte Aussagen lassen sich teilweise schwer lesen, gerade wenn Negationen weit außen in der Aussage stehen. Quantifizierte Aussagen, bei denen sich die Negation nur auf Elementaraussagen bezieht, haben einen speziellen Namen:

Definition 1.8. Eine quantifizierte Aussage ist in **Negationsnormalform**, wenn sich die Negationsjunktoren nur auf Elementaraussagen beziehen und nur die Junktoren \neg, \wedge, \vee verwendet werden.

Beispiel.

- $\exists x \in \mathbb{N} \neg(\forall y \in \mathbb{Z} \exists z \in \mathbb{Z} : x \leq y \wedge z > x)$ ist nicht in Negationsnormalform, da sich das \neg auf eine quantifizierte Formel bezieht.
- $\exists x \in \mathbb{N} \forall y \in \mathbb{Z} \exists z \in \mathbb{Z} : x \leq y \rightarrow z > x$ ist nicht in Negationsnormalform, da \rightarrow als Junktor verwendet wird.
- $\exists x \in \mathbb{N} \forall y \in \mathbb{Z} \exists z \in \mathbb{Z} : \neg(x \leq y \wedge z > x)$ ist nicht in Negationsnormalform, da sich das \neg nicht auf eine atomare Aussage bezieht.
- $\exists x \in \mathbb{N} \forall y \in \mathbb{Z} \exists z \in \mathbb{Z} : \neg(x \leq y) \vee \neg(z > x)$ ist in Negationsnormalform.

Eine beliebige Formel kann in Negationsnormalform umgewandelt werden, in dem die Regeln aus Satz 1.3 und Satz 1.2 verwendet werden, um sicherzustellen, dass sich die Negation nur auf elementare Aussagen bezieht und dass nur die geforderten Junktoren verwendet werden.

Beispiel. Wir zeigen dieses Vorgehen anhand der Negation der Aussage, dass es für alle $x < y$ immer ein z gibt mit $x < z < y$.

$$\begin{aligned} & \neg(\forall x \forall y (x < y \rightarrow (\exists z (x < z \wedge z < y)))) \\ & \simeq \exists x \exists y (\neg(x < y \rightarrow (\exists z (x < z \wedge z < y)))) \\ & \simeq \exists x \exists y (\neg(\neg(x < y) \vee (\exists z (x < z \wedge z < y)))) \\ & \simeq \exists x \exists y ((x < y) \wedge \neg(\exists z (x < z \wedge z < y))) \\ & \simeq \exists x \exists y ((x < y) \wedge (\forall z \neg(x < z \wedge z < y))) \\ & \simeq \exists x \exists y ((x < y) \wedge (\forall z ((x \geq z) \vee (z \geq y)))) \end{aligned}$$

Definition 1.9. Eine quantifizierte Aussage ist in **Pränexform**, wenn sie in Negationsnormalform ist und alle auftretenden Quantoren am Anfang der Aussage stehen.

Beispiel. $(\forall x \forall y (x < y \rightarrow (\exists z (x < z \wedge z < y))))$ ist nicht in Pränexform. Wir verwenden logische Äquivalenzen, um die Aussage umzuformen.

$$\begin{aligned} & \forall x \forall y (x < y \rightarrow (\exists z (x < z \wedge z < y))) \\ \simeq & \forall x \forall y (\neg(x < y) \vee \exists z (x < z \wedge z < y)) \\ \simeq & \forall x \forall y ((x \geq y) \vee \exists z (x < z \wedge z < y)) \\ \simeq & \forall x \forall y (\exists z (x \geq y) \vee \exists z (x < z \wedge z < y)) \\ \simeq & \forall x \forall y \exists z ((x \geq y) \vee (x < z \wedge z < y)) \end{aligned}$$

Selbsttest: Bringen Sie die folgende Aussage in Pränexform.

$$\neg(\forall x \exists y : (x > y) \implies \forall z : (z < y))$$

Anders als bei Termen, kann gezeigt werden, dass es keine Berechnungsvorschrift gibt, um für eine gegebene quantifizierte Aussage immer festzustellen, ob diese wahr oder falsch ist. Formal kann man sagen, dass dieses Problem *nicht entscheidbar* ist.⁷

Für viele konkrete Aussagen kann jedoch gezeigt werden, dass diese wahr oder falsch sind. Eine Möglichkeit dies mit einem allgemeinen Vorgehen anzugehen ist der Beweis durch das *Gegenspieler-Beweiser* Prinzip, welches wir uns im Folgenden anschauen werden. Dieses Prinzip funktioniert in drei Schritten, auf die wir nach und nach eingehen werden.

1. Bringe die Aussage in Pränexform.
2. Belege die Variablen nach dem Gegenspieler-Beweiser Muster
3. Verifiziere die Formel.

Den ersten Schritt haben wir oben schon betrachtet. Für den zweiten Schritt gehen wir davon aus, dass wir (als *Beweiser*) zeigen wollen, dass die gegebene Aussage wahr ist. Allerdings gibt es einen imaginären *Gegenspieler*, der die Werte von allen Variablen angeben kann, die mit Allquantoren gebunden sind. Als Beweiser müssen wir nun Belegungen der Variablen, die mit einem Existenzquantor gebunden sind, finden, die für jeden möglichen Wert, den der Gegenspieler gesetzt hat, zu einer wahren Aussage führt. Im letzten Schritt wird dann argumentiert, dass die Aussage mit dieser Belegung korrekt ist.

Wir betrachten dieses Vorgehen an zwei Beispielen:

Beispiel. Wir wollen zeigen, dass die Aussage $\forall x \in \mathbb{Q} \forall y \in \mathbb{Q} (x < y) \rightarrow \exists z (x < z \wedge z < y)$ wahr ist. Dazu verwenden wir das Gegenspieler-Beweiser Prinzip.

⁷Details zu diesen Konzepten werden in der Vorlesung *Grundlagen der theoretischen Informatik* besprochen.

1. Dies ist genau die Aussage, an der wir die Umformung in Pränexform im Beispiel oben geübt haben, die Aussage in Pränexform ist also

$$\forall x \in \mathbb{Q} \forall y \in \mathbb{Q} \exists z \in \mathbb{Q} ((x \geq y) \vee (x < z \wedge z < y))$$

2. Nun kommen wir zum Herzstück des Beweises. Die ersten beiden Quantoren sind Allquantoren, also wählt der Gegenspieler Zahlen $a \in \mathbb{Q}$ und $b \in \mathbb{Q}$ als Belegung für x und y . Im Gegensatz zu den gebundenen Variablen x und y sind a und b konkrete Werte. Da der Beweiser aber auf alle möglichen Werte vorbereitet sein muss, werden diese wieder mit Variablen benannt. Der Beweiser muss nun eine Zahl $c \in \mathbb{Q}$ als Belegung für z wählen, die zwischen a und b liegt. Der Beweiser wählt $c = \frac{a+b}{2}$.
3. Nun zeigen wir, dass $(a \geq b) \vee (a < c \wedge c < b)$ eine wahre Aussage ist. Hat der Gegenspieler a und b so gewählt, dass $a \geq b$ gilt, ist die Aussage direkt wahr. Im anderen Fall gilt $a < b$. Mit der Wahl von c muss nun also gezeigt werden, dass $a < \frac{a+b}{2}$ und $\frac{a+b}{2} < b$ gilt. Es gilt:

$$a = \frac{2a}{2} < \frac{a+b}{2} < \frac{2b}{2} = b$$

und die Aussage ist gezeigt.

Als zweites Beispiel zeigen wir, dass die Aussage nicht mehr gilt, wenn die Zahlen statt aus \mathbb{Q} aus \mathbb{N} kommen. Intuitiv geht der oben gezeigte Beweis an der Stelle kaputt, an der $c = \frac{a+b}{2}$ gewählt wird, da dieser Wert keine ganze Zahl (und damit auch keine natürliche Zahl) sein muss. Aber nur zu argumentieren, dass der Beweis sich so nicht auf die natürlichen Zahlen übertragen lässt ist noch nicht ausreichend um zu schlussfolgern, dass die Aussage nicht gilt, es könnte ja einen alternativen Beweis geben. Um dies auszuschließen, zeigen wir, dass die Negation der Aussage eine wahre Aussage ist, damit ist dann auch gezeigt, dass die ursprüngliche Aussage über den natürlichen Zahlen eine falsche Aussage ist.

1. Wir negieren die Aussage $\forall x \in \mathbb{N} \forall y \in \mathbb{N} \exists z \in \mathbb{N} ((x \geq y) \vee (x < z \wedge z < y))$

$$\begin{aligned} & \neg(\forall x \in \mathbb{N} \forall y \in \mathbb{N} \exists z \in \mathbb{N} ((x \geq y) \vee (x < z \wedge z < y))) \\ & \simeq \exists x \in \mathbb{N} \exists y \in \mathbb{N} \forall z \in \mathbb{N} \neg((x \geq y) \vee (x < z \wedge z < y)) \\ & \simeq \exists x \in \mathbb{N} \exists y \in \mathbb{N} \forall z \in \mathbb{N} \neg(x \geq y) \wedge (\neg(x < z \wedge z < y)) \\ & \simeq \exists x \in \mathbb{N} \exists y \in \mathbb{N} \forall z \in \mathbb{N} \neg(x \geq y) \wedge (\neg(x < z) \vee \neg(z < y)) \\ & \simeq \exists x \in \mathbb{N} \exists y \in \mathbb{N} \forall z \in \mathbb{N} (x < y) \wedge ((x \geq z) \vee (z \geq y)) \end{aligned}$$

2. Die Aussage beginnt nun mit zwei Existenzquantoren, als Beweiser können also konkrete Zahlenwerte gewählt werden. Wir wählen $a = 1$ und $b = 2$. Der Gegenspieler wählt nun ein beliebiges $c \in \mathbb{N}$.

3. Wir betrachten den Wahrheitswert der Aussage $(1 < 2) \wedge ((1 \geq c) \vee (c \geq 2))$. Die Aussage $(1 < 2)$ ist offensichtlich wahr, es reicht also zu zeigen, dass $(1 \geq c) \vee (c \geq 2)$ eine wahre Aussage ist. Wenn $c \leq 1$ ist, ist die Aussage offensichtlich wahr. Also nehmen wir an, dass $c > 1$ gilt. In diesem Fall ist c aber mindestens 2, da c eine natürliche Zahl ist, und der Teil $c \geq 2$ der Aussage ist wahr. ◀

1.3 Syntax der Aussagenlogik, Boolesche Algebra und Boolesche Funktionen

Anmerkung: Dieser Abschnitt wird in der Vorlesung häufig zunächst übersprungen und dann nach dem Themenbereich Funktionen behandelt. Der Grund dafür ist, dass in diesem Abschnitt Konzepte aus dem Bereich Funktionen verwendet werden.



Voraussetzung:

- Kartesisches Produkt (Definition 3.8)
- Begriff der Funktion (Abschnitt 3.3)



Lernziele

Die Studierenden ...

- ... unterscheiden zwischen der Syntax und der Semantik eines Booleschen Terms
- ... verwenden die Baumdarstellung von Booleschen Termen
- ... unterscheiden zwischen einem Booleschen Term und einer Booleschen Funktion
- ... finden boolesche Terme in Normalformen zu gegebenen Booleschen Funktionen



1.3.1 Syntax der Aussagenlogik

In diesem Abschnitt werden den Begriff des *Booleschen Terms* formal einführen, sowie auf den Unterschied zwischen Syntax und Semantik kurz eingehen. Einfach gesagt, beschreibt die *Syntax* eine Art Grammatik, die beschreibt, wie ein korrekter Satz aufgebaut ist. Die Semantik weist einem solchen Satz dann eine Bedeutung zu. Der Satz „Das Pferd fährt auf dem Fahrrad das Meer hinauf.“ ist ein grammatikalisch korrekter deutscher Satz, er erfüllt die Regeln für die Syntax. Dies ist unabhängig davon, welche Bedeutung wir dem Satz zuweisen.

Definition 1.10 (Syntax der Aussagenlogik). Die Menge der Booleschen Formeln (Booleschen Terme) der Aussagenlogik über der Variablenmenge V und der Aussagenmenge $A = \{\text{true}, \text{false}\}$ ist induktiv definiert:

1. Jedes $a \in A$ und jedes $v \in V$ sind Boolesche Terme.
2. Wenn t ein Boolescher Term ist, so ist auch $(\neg t)$ ein Boolescher Term.
3. Wenn t_1 und t_2 Boolesche Terme sind, so sind auch $(t_1 \wedge t_2)$ und $(t_1 \vee t_2)$ Boolesche Terme.
4. (Minimalitätsprinzip) Nur Konstrukte, die sich durch endlich oft wiederholtes Anwenden der Regeln 1, 2 und 3 erzeugen lassen sind Boolesche Terme.

Die ersten drei Regeln definieren dabei wie wir uns vorstellen, dass ein Boolescher Term aussehen soll, die vierte Regel stellt sicher, dass keine anderen Konstrukte als Boolesche Terme gelesen werden. Wir haben nicht alle der in Abschnitt 1.1.2 eingeführten Junktoren verwendet, sondern nur zusätzlich zu true und false nur \neg , \wedge und \vee . Die Menge $\{\neg, \vee, \wedge\}$ nennt man die **Standardsignatur** für Boolesche Terme. **Wichtig:** Definition 1.10 weist den Bestandteilen der Aussagen bis jetzt noch keine Interpretation bezüglich des Wahrheitswerts zu.

Beispiel. Wir betrachten für die folgenden Beispiele die Variablenmenge $\{a, b, c\}$.

- $(a \wedge b)$, $(a \vee b)$, $\neg(a \vee b)$ und $((a \vee b) \wedge c)$ sind Boolesche Terme.
- $a \wedge b$, $a \wedge \vee b$, $a \neg b$ sind keine Booleschen Terme, da die Junktoren unerlaubt kombiniert werden, oder die Klammerung nicht stimmt.
- $a \wedge b \wedge c$ und $\neg a \wedge b$ sind keine Booleschen Terme nach Definition 1.10, da dort um jeden Schritt Klammern gefordert werden. ◀

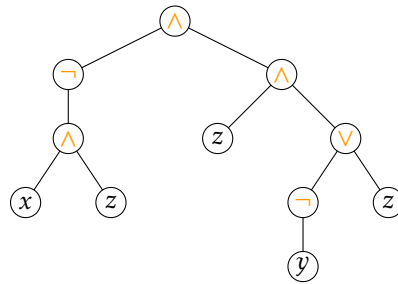
Selbsttest: Finden Sie drei Konstrukte, welche gültige Terme sind. Verwenden Sie dafür wenn möglich alle Regeln. Finden Sie drei Konstrukte, welche keine gültigen Terme sind.



Wir betrachten nun die **Baumdarstellung eines Booleschen Terms (Syntaxbaum)**⁸. Wir werden die Darstellung nicht formal einführen, sondern an einem Beispiel illustrieren.

Beispiel. Der Syntaxbaum zum Term $t = ((\neg(x \wedge z)) \wedge (z \wedge ((\neg y) \vee z)))$ ist

⁸Warum diese Darstellung Baumdarstellung heißt werden wir hier nicht genauer besprechen, der Grund wird in Abschnitt 6.2 klarer.



Diese Baumdarstellung stellt die Reihenfolge, in der der Term nach den Regeln in Definition 1.10 erzeugt wurde, dar. Als letzte Operation wurden zwei Terme mit $(t_1 \wedge t_2)$ verbunden. Der linke Term wiederum wurde mit $(\neg t_3)$ erzeugt, der rechte mit $(t_4 \wedge t_5)$ und so weiter.

Ein vollständig geklammerter Term definiert eindeutig den zugehörigen Syntaxbaum. Umgekehrt kann aus einem Syntaxbaum der vollständig geklammerte Term abgelesen werden. Diese Verbindung zeigt sich auch in der Definition des Rangs $\text{rg}(t)$ eines Term. Intuitiv ist der Rang eines Terms die maximale Schachtelungstiefe der Klammern. In der Verbindung zur Baumdarstellung, kann der Rang auch als die maximale Anzahl an Schritten gesehen werden, die gegangen werden muss, um von der obersten Ebene nach unten zu Laufen und bei einer Variable zu enden. Der Rang des Beispielterms ist 4. Wir betrachten nun die formale Definition des Rangs. Der Rang ist induktiv definiert und die Definition spiegelt die Struktur von Definition 1.10 wider.

Definition 1.11 (Rang eines Booleschen Terms). Sei t ein Boolescher Term über der Aussagenmenge A und der Variablenmenge V .

1. Für jedes $a \in A$ und $v \in V$ ist $\text{rg}(a) = \text{rg}(v) = 0$.
2. Wenn $t = (\neg t_1)$, dann ist $\text{rg}(t) = \text{rg}(t_1) + 1$.
3. Wenn $t = (t_1 \wedge t_2)$ oder $t = (t_1 \vee t_2)$, dann ist $\text{rg}(t) = \max\{\text{rg}(t_1), \text{rg}(t_2)\} + 1$.

Selbsttest: Schreiben Sie zwei verschiedene Boolesche Terme mit jeweils mindestens 3 Variablen auf. Finden Sie den Syntaxbaum zu den Termen. Was ist der Rang der Terme?

1.3.2 Boolesche Algebra

Wir haben in Abschnitt 1.1 schon gesehen, wie zusammengesetzten Aussagen ein Wahrheitswert zugewiesen werden kann. Hier werden wir uns die Junktoren aus dem Blickwinkel von Funktionen anschauen.

Sei $\mathbb{B} = \{0, 1\}$ die Menge der Booleschen Wahrheitswerte. Mit $\mathbb{B} \times \mathbb{B}$ bezeichnen wir die Menge der geordneten Paare von Wahrheitswerten (das sogenannte kartesische Produkt⁹). Das bedeutet also $\mathbb{B} \times \mathbb{B} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Die Negation kann jetzt als Funktion $\neg : \mathbb{B} \rightarrow \mathbb{B}$, die als $\neg 0 = 1$ und $\neg 1 = 0$ definiert ist, betrachtet werden. Ebenso können die zweistelligen Junktoren als zweistellige Boolesche Funktionen über die Wertetabellen definiert werden:

a	b	$a \wedge b$	$a \vee b$	$a \oplus b$	$a \rightarrow b$	$a \leftrightarrow b$
0	0	0	0	0	1	1
0	1	0	1	1	1	0
1	0	0	1	1	0	0
1	1	1	1	0	1	1

Zur Unterscheidung der Syntax und der Semantik werden hier verschiedene Farben für die Verknüpfungen auf der Ebene der Syntax und der Ebene der Semantik verwendet. Später werden wir die Zeichen auf beiden Ebenen wieder schwarz darstellen, da im Allgemeinen aus dem Kontext zu erkennen ist, welche Ebene aktuell betrachtet wird.

1.3.3 Interpretation eines Booleschen Terms

Informell haben wir die Begriffe der Belegung und der Auswertung (Interpretation) eines Booleschen Terms schon in Definition 1.4 eingeführt. Nun werden wir den Begriff formell definieren. Die **Belegung** eines Terms, ist eine Funktion $\beta : V \rightarrow \mathbb{B}$, die jeder Variable einen Wahrheitswert zuordnet.

Definition 1.12 (Interpretation eines Booleschen Terms). Sei t ein Boolescher Term über der Variablenmenge V und Aussagenmenge A und sei β eine Belegung der Variablen.

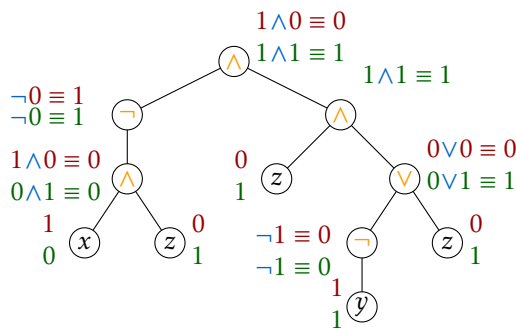
Dann ist die **Interpretation** $I_\beta(t)$ gegeben durch:

1. Falls $t = a$ für ein $a \in A$, dann ist $I_\beta(a)$ bekannt, da wir davon ausgehen die Wahrheitswerte der Elementaraussagen zu kennen. Es gilt $I_\beta(\text{true}) = 1$ und $I_\beta(\text{false}) = 0$.
2. Falls $t = v$ für $v \in V$, ist $I_\beta(t) = \beta(v)$.
3. Falls $t = (\neg t_1)$, dann ist $I_\beta(t) = \neg I_\beta(t_1)$.
4. Falls $t = (t_1 \vee t_2)$, dann ist $I_\beta(t) = I_\beta(t_1) \vee I_\beta(t_2)$. Falls $t = (t_1 \wedge t_2)$, dann ist $I_\beta(t) = I_\beta(t_1) \wedge I_\beta(t_2)$.

Anmerkungen:

⁹mehr dazu in Kapitel 3

1. Wenn eine feste Belegung β gegeben ist, kann die Interpretation durch eine **bottom-up Auswertung** entlang des Syntaxbaums gefunden werden. Die Variablen werden durch die Wahrheitswerte ersetzt und die Operatoren aus der Booleschen Algebra werden verwendet.



v	$\beta_1(v)$	$\beta_2(v)$
x	1	0
y	1	1
z	0	1

$$I_{\beta_1}(\neg(x \wedge z) \wedge (z \wedge (\neg y \vee z))) = 0$$

$$I_{\beta_2}(\neg(x \wedge z) \wedge (z \wedge (\neg y \vee z))) = 1$$

2. Zur Erinnerung: Die farbliche Unterscheidung von \neg, \vee, \wedge und \neg, \vee, \wedge ist sehr bewusst gewählt. Die ersten drei Zeichen sind für die syntaktische Ebene, die letzten drei operieren auf der Seite der Wahrheitswerte.

Selbsttest: Betrachten Sie Ihre Terme und Syntaxbäume aus dem vorherigen Selbsttest. Wählen Sie jeweils eine Belegung der Variablen für jeden Term und finden Sie die Interpretation der Terme bezüglich der Belegungen.



1.3.4 Boolesche Funktionen

Für eine Variablenmenge $V = \{x_1, \dots, x_n\}$ kann eine konkrete Belegung β als n -Tupel (b_1, \dots, b_n) von Wahrheitswerten geschrieben werden, mit $b_i = \beta(x_i)$. Die Menge aller 2^n solcher verschiedener n -Tupel bezeichnet man als $B \times B \times \dots \times B$. Wir haben schon \wedge und \vee als Funktion von $B \times B$ nach B interpretiert. Im Folgenden verallgemeinern wir diese Definition.

Definition 1.13 (Boolesche Funktion). Eine n -stellige Boolesche Funktion f ist eine Funktion

$$f : \underbrace{B \times B \times \dots \times B}_{n \text{ mal}} \rightarrow B$$

Eine Boolesche Funktion kann, wie oben gesehen, durch das explizite Angeben einer Wertetabelle definiert werden. Häufiger wird jedoch ein Term angegeben, der die Funktion definiert. Jeder Term t über einer n -elementigen Variablenmenge definiert eine n -stellige Boolesche Funktion:

$$f_t : \underbrace{B \times B \times \dots \times B}_{n \text{ mal}} \rightarrow B$$

$$f_t(b_1, \dots, b_n) = I_\beta(t)$$

Konventionen Zur vereinfachten Darstellung von Booleschen Termen werden folgende Vereinbarungen getroffen:

1. Das äußerste Klammerpaar kann weggelassen werden.
2. Die Bindungskraft der logischen Operatoren $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ nimmt in dieser Reihenfolge von links nach rechts ab.
Beispiel: Der Term $t = (((\neg x) \vee (\neg y)) \wedge z)$ lässt sich somit schreiben als $t = (\neg x \vee \neg y) \wedge z$.
3. Im Zweifelsfall immer die Klammern stehen lassen!
4. Für einfachere Lesbarkeit werden wir auch auf Termebene 0 und 1 statt `true` und `false` verwenden.
5. Wegen der Assoziativität von \wedge und \vee können wir $\alpha \wedge \beta \wedge \gamma$ und $\alpha \vee \beta \vee \gamma$ ohne Klammer schreiben. Die Funktionen \wedge und \vee bleiben aber zweistellige Boolesche Funktionen.

Wir haben gesehen, dass wir jedem Booleschen Term eine eindeutige Boolesche Funktion zuweisen können, aber geht das auch umgekehrt?

Beobachtung 1.6. *Es gibt unendlich viele syntaktisch verschiedene Boolesche Terme, die ein und dieselbe Boolesche Funktion repräsentieren.*

Beweis. Es gilt $t \equiv t \vee t \equiv t \vee t \vee t \equiv \dots$. All diese syntaktisch verschiedenen Terme haben dieselbe semantische Interpretation f_t . \square

Beobachtung 1.7. *Es gibt genau 2^{2^n} viele verschiedene n -stellige Boolesche Funktionen.*

Beweis. Eine Funktion ist eindeutig durch ihre Wertetabelle bestimmt. Wir zählen also die Anzahl an möglichen verschiedenen Wertetabellen. Bei einer n -stelligen Booleschen Funktion hat die Wertetabelle 2^n Zeilen, eine für jede mögliche Belegung. Jeder Zeile wird einer von zwei möglichen Werten aus \mathbb{B} zugeordnet. Also gibt es insgesamt

$$\underbrace{2 \cdot 2 \cdot 2 \cdot \dots \cdot 2}_{2^n \text{ Faktoren}} = 2^{2^n}$$


Möglichkeiten, die Wertetabelle auszufüllen. \square

Beobachtung 1.6 sagt, dass wenn es einen Term gibt, der eine Boolesche Funktion beschreibt, dann gibt es unendlich viele syntaktisch verschiedene Terme. Aber gibt es für jede Boolesche Funktion auch immer mindestens einen Term, der sie repräsentiert? Wir werden feststellen, dass die Antwort auf diese Frage *ja* ist und auch ein allgemeines Verfahren angeben, um einen solchen Term zu finden. Wir wollen die folgende Aufgabe lösen

Aufgabe Gegeben sei eine n -stellige Boolesche Funktion g , finde einen Booleschen Term t , sodass $f_t = g$

Um diese Aufgabe zu lösen, betrachten wir zunächst eine einfachere Teilaufgabe:


Teilaufgabe Sei g_b eine n -stellige Boolesche Funktion für $n \geq 1$, die für das konkrete Element $b = (b_1, \dots, b_n) \in \mathbb{B} \times \dots \times \mathbb{B}$ den Wert 1 annimmt und für alle anderen Argumente den Wert 0. Finde einen Term über der Variablenmenge $V = \{x_1, x_2, \dots, x_n\}$, der diese Funktion repräsentiert.

Beispiel. Sei $n = 3$ und $b = (1, 1, 0)$. Der Term $t = x_1 \wedge x_2 \wedge \neg x_3$ leistet das Gewünschte. Zunächst ist einfach zu sehen, dass $(1, 1, 0)$ eine erfüllende Belegung für den Term t ist. Auf der anderen Seite ist bei einer anderen Belegung mindestens einer der drei Ausdrücke x_1 , x_2 oder $\neg x_3$ immer falsch. 

Um die Teilaufgabe allgemein zu lösen, führen wir zunächst einiges an Notation ein und Begriffen ein. Wir definieren für jedes i die Teilformel $x_i^{b_i}$. Wenn $b_i = 1$ ist, ist die Teilformel x_i , sonst $\neg x_i$. Formal kann dies so aufgeschrieben werden:¹⁰

$$x_i^{b_i} = \begin{cases} x_i & b_i = 1 \\ \neg x_i & b_i = 0 \end{cases}$$

Ein allgemeiner Term der bei Belegung $b = (b_1, \dots, b_n)$ zu 1 und sonst zu 0 auswerten soll, kann jetzt als $t = x_1^{b_1} \wedge \dots \wedge x_n^{b_n}$ geschrieben werden. Umgekehrt ist der Term $\neg t := \neg(x_1^{b_1} \wedge \dots \wedge x_n^{b_n}) \equiv \neg x_1^{b_1} \vee \dots \vee \neg x_n^{b_n}$ ein Term der bei Belegung b zu 0 auswertet und sonst zu 1.

Selbsttest: Wählen Sie eine beliebige Belegung für die Variablen x_1, \dots, x_4 . Was ist die Formel $t = x_1^{b_1} \wedge \dots \wedge x_n^{b_n}$? 

Diese Erkenntnis hilft uns jetzt beim Lösen der eigentlichen Aufgabe. Wir sammeln alle Belegungen bei denen die Boolesche Funktion g zu 1 auswertet in der Menge $g^{-1}(1) = \{(b_1, \dots, b_n) \mid g(b_1, \dots, b_n) = 1\}$ und alle Belegungen bei denen die Boolesche Funktion g zu 0 auswertet in der Menge $g^{-1}(0) = \{(b_1, \dots, b_n) \mid g(b_1, \dots, b_n) = 0\}$. Wie im Abschnitt über Funktionen (Abschnitt 3.3) gelernt, nennt man diese Mengen auch das Urbild der 1 beziehungsweise der 0 bei der Funktion g .

Um die Aufgabe zu lösen, gibt es nun zwei Möglichkeiten:

Variante 1 Finde den Booleschen Term t , der genau dann zu 1 ausgewertet wird, wenn man die erste **oder** die zweite **oder** ... **oder** die letzte Belegung aus $g^{-1}(1)$ wählt.

¹⁰Diese Notation unterscheidet zwei verschiedene Fälle, wenn die Bedingung hinten die in der ersten Zeile steht wahr ist, wird der Wert vorne in der ersten Zeile zurückgegeben, sonst wird die nächste Zeile angeschaut, bis eine Bedingung gilt.

Variante 2 Finde den Booleschen Term t , der genau dann zu 1 ausgewertet wird, wenn man **nicht** die erste **und nicht** die zweite **und nicht** ... **und nicht** die letzte Belegung aus $g^{-1}(0)$ wählt.

Um diese Ideen formal aufzuschreiben brauchen wir noch einige Definitionen:

Definition 1.14.

1. Ein **Literal** ist ein Variable oder deren Negation.
2. Ein **Maxterm** ist eine Disjunktion von Literalen und ein **Minterm** ist eine Konjunktion von Literalen. In beiden Fällen gilt ein einzelnes Literal als Maxterm bzw. Minterm.
3. Wenn V die endliche Menge aller Literale ist, dann ist ein Maxterm/Minterm **vollständig**, wenn er für jedes $1 \leq i \leq n$ entweder das Literal x_i oder das Literal $\neg x_i$ enthält.
4. Ein Boolescher Term ist in **disjunktiver Normalform (DNF)**, falls er eine Disjunktion von Mintermen ist.
5. Ein Boolescher Term ist in **konjunktiver Normalform (KNF)**, falls er eine Konjunktion von Maxtermen ist.

Beispiel. x_1 ist ein Literal, $\neg x_3$ ist ein Literal.

Für die Variablenmenge $V = \{x_1, x_2, x_3\}$ ist $\neg x_1 \wedge x_2 \wedge \neg x_3$ ein vollständiger Minterm, $\neg x_1 \vee \neg x_2 \vee x_3$ ist ein vollständiger Maxterm.

$(x_1 \wedge x_2) \vee (\neg x_1 \wedge x_3) \vee x_2$ ist eine DNF aber keine KNF. Die Formeln $x_1 \vee x_2$ und $\neg x_1 \wedge x_4 \wedge \neg x_6$ sind sowohl in DNF als auch in KNF. ◀

Selbsttest: Können Sie für alle Teile im Beispiel begründen, warum die Eigenschaften jeweils gelten bzw. nicht gelten? ?

Nun können wir die Konstruktion formal definieren. Dies geschieht im Rahmen eines Satzes mit dazugehörendem Beweis. Für das bessere Verständnis des Satzes und Beweises kann es hilfreich sein, sich vor dem Beweis, das nach dem Satz zu findende Beispiel anzuschauen.

Satz 1.8. Jede n -stellige Boolesche Funktion g ist durch die sogenannte **kanonische DNF** $\text{dnf}(g)$ und durch die **kanonische KNF** $\text{knt}(g)$ über der Variablenmenge $V = \{x_1, \dots, x_n\}$ repräsentierbar, wobei

$$\begin{aligned}\text{dnf}(g) &= \bigvee_{(b_1, \dots, b_n) \in g^{-1}(1)} (x_1^{b_1} \wedge \dots \wedge x_n^{b_n}) \\ \text{knt}(g) &= \bigwedge_{(b_1, \dots, b_n) \in g^{-1}(0)} (\neg x_1^{b_1} \vee \dots \vee \neg x_n^{b_n})\end{aligned}$$

Im Spezialfall $g^{-1}(1) = \emptyset$ (also es gibt keine Belegung die zu 1 auswertet), setzen wir $\text{dnf}(g) = \text{false}$ und im Spezialfall $g^{-1}(0) = \emptyset$ (also es gibt keine Belegung, die zu 0 auswertet), setzen wir $\text{knf}(g) = \text{true}$.

Beweis. $\text{dnf}(g)$ setzt genau Variante 1 und $\text{knf}(g)$ Variante 2 um. Dies ist natürlich kein Beweis, daher schauen wir uns jetzt die formalen Argumente an.

Wie in der Teilaufgabe gesehen, ist $x_1^{b_1} \wedge \dots \wedge x_n^{b_n}$ für die Belegung (b_1, \dots, b_n) wahr und für alle anderen Belegungen falsch. Genauer wird die Konjunktion¹¹ von Literalen nur wahr, wenn alle Literale wahr sind, und dies geschieht genau durch (b_1, \dots, b_n) . Da wir für alle Belegungen, die mit g zu wahr auswerten die Disjunktion¹² dieser vollständigen Minterme bilden, wird für jede solche Belegung genau ein Minterm und damit die ganze Disjunktion wahr. Alle Belegungen aus $g^{-1}(0)$ haben die gemeinsame Eigenschaft, dass sie keinen der Minterme wahr machen, damit ist die ganze Disjunktion falsch.

Der Beweis für $\text{knf}(g)$ ist analog¹³. Jede Disjunktion $\neg x_1^{b_1} \vee \dots \vee \neg x_n^{b_n}$ wird für die konkrete Belegung (b_1, \dots, b_n) falsch und für alle anderen wahr. Durch die Konjunktion über alle (b_1, \dots, b_n) aus $g^{-1}(0)$ ergibt sich eine Formel, die für alle solche Belegungen falsch ist. Für alle Belegungen aus $g^{-1}(1)$ ergibt sich dann eine wahre Belegung. \square

Korollar 1.9. Jeder Boolesche Term t ist semantisch äquivalent zu einem Term in DNF und zu einem Term in KNF.

Beweis. Wir haben oben schon gesehen, dass jedem Term t mit f_t eine Boolesche Funktion zugewiesen werden kann. Aus dieser kann dann nach dem obigen Schema die kanonisch KNF und die kanonische DNF gebildet werden. \square

Beispiel. Die folgende Tabelle zeigt eine 3-stellige Boolesche Funktion f zusammen mit dem vollständigen Mintermen für die Belegungen in $f^{-1}(1)$ und den vollständigen Maxtermen zu den Belegungen aus $f^{-1}(0)$.

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	Minterm	Maxterm
0	0	0	0	–	$x_1 \vee x_2 \vee x_3$
0	0	1	1	$\neg x_1 \wedge \neg x_2 \wedge x_3$	–
0	1	0	0	–	$x_1 \vee \neg x_2 \vee x_3$
0	1	1	1	$\neg x_1 \wedge x_2 \wedge x_3$	–
1	0	0	1	$x_1 \wedge \neg x_2 \wedge \neg x_3$	–
1	0	1	1	$x_1 \wedge \neg x_2 \wedge x_3$	–
1	1	0	0	–	$\neg x_1 \vee \neg x_2 \vee x_3$
1	1	1	1	$x_1 \wedge x_2 \wedge x_3$	–

¹¹„Verundung“

¹²„Veroderung“

¹³Die Beweisstruktur ist also die gleiche, es müssen nur ab und zu Wörter ersetzt werden

Dies ergibt die folgenden kanonische DNF und kanonische KNF:

$$\begin{aligned}\text{dnf}(f) &= (\neg x_1 \wedge \neg x_2 \wedge x_3) \vee (\neg x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge \neg x_2 \wedge \neg x_3) \\ &\quad \vee (x_1 \wedge \neg x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge x_3) \\ \text{knf}(f) &= (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3)\end{aligned}$$

Selbsttest: Schreiben Sie eine Wertetabelle zu einer Booleschen Funktion auf. Finden Sie dann die kanonische KNF und die kanonische DNF.



1.3.5 Vollständige logische Signaturen

In diesem Abschnitt beschäftigen wir uns damit, welche Junktoren ausreichend sind um alle möglichen Booleschen Funktionen darzustellen. Dazu betrachten wir die folgende Definition:

Definition 1.15 (Logische Signatur). Die Menge der Junktoren, die zur Definition von Booleschen Termen benutzt wird, nennt man **logische Signatur**. Die Signatur $\Sigma_0 = \{\neg, \wedge, \vee\}$ heißt **Boolesche Standardsignatur**.

Definition 1.16 (Funktionale Vollständigkeit). Eine logische Signatur ist **funktional vollständig**, wenn jede Boolesche Funktion von durch die Signatur erzeugten Termen dargestellt werden kann.

Beobachtung 1.10. Die Boolesche Standardsignatur ist funktional vollständig.

Beweis. Wir führen einen Beweis durch Fallunterscheidung. Sei f eine beliebige Boolesche Funktion. Wir zeigen nun, dass jede Funktion f mit Σ_0 dargestellt werden kann.

Fall 1: f ist konstant. Die konstante Funktion mit Wert 0 kann durch den Term $x \wedge \neg x$ dargestellt werden. Die konstante Funktion mit Wert 1 durch den Term $x \vee \neg x$.

Fall 2: f ist nicht konstant. Dann ist nach Satz 1.8 die kanonische KNF oder die kanonische DNF ein Term, welcher die Funktion repräsentiert. \square

Nun wenden wir uns zwei Fragen zu: 1. Gibt es weitere funktional vollständige Signaturen und 2. wie können wir zeigen, dass eine Signatur funktional vollständig bzw. nicht vollständig ist.

Wir betrachten zunächst die zweite Frage, die Antwort auf die erste Frage ergibt sich dann aus den Beispielen. Um die zweite Frage zu beantworten, brauchen wir zunächst noch eine weitere Definition.

Definition 1.17. Eine Signatur Γ simuliert eine Signatur Σ , wenn man zu jedem Term über Σ einen semantisch äquivalenten Term über Γ finden kann.

Insbesondere reicht es aus, wenn alle Junktoren aus Σ mit den Junktoren aus Γ dargestellt werden können.

Nun können wir den folgenden Satz aufstellen, mit dessen Hilfe wir dann zeigen können, dass eine Signatur funktional vollständig oder eben auch nicht funktional vollständig ist.

Satz 1.11. Angenommen Γ kann Σ simulieren. Dann gilt

- (i) Wenn Σ funktional vollständig ist, dann ist auch Γ funktional vollständig.
- (ii) Wenn Γ nicht funktional vollständig ist, dann ist auch Σ nicht funktional vollständig.

Beweis. (i) Sei f eine Funktion und t_Σ der Term über Σ , der f repräsentiert. Dann gibt es nach Voraussetzung einen Term t_Γ^* über Γ , der f repräsentiert.

(ii) Die Aussage ist die Kontraposition von (i). □

Wir betrachten nun eine Reihe von Beispielen von funktional vollständigen und nicht funktional vollständigen Signaturen.

Beispiel.

1. $\Sigma_1 = \{\neg, \wedge\}$ und $\Sigma_2 = \{\neg, \vee\}$ sind funktional vollständig. Nach Satz 1.11 reicht es aus, zu zeigen, dass Σ_0 durch Σ_1 bzw. durch Σ_2 simuliert werden kann.

Wir betrachten zunächst Σ_1 . Dass \neg und \wedge dargestellt werden können ist klar, da diese sowohl in Σ_0 als auch in Σ_1 enthalten sind. Nun muss noch \vee simuliert werden. Dazu stellen wir fest, dass folgendes gilt:

$$\begin{aligned} a \vee b &\equiv \neg(\neg(a \vee b)) \\ &\equiv \neg(\neg a \wedge \neg b) \end{aligned}$$

Wobei die letzte Zeile aus dem deMorganschen Gesetz folgt. Analog kann Σ_0 auch von Σ_2 simuliert werden. \neg und \vee sind direkt enthalten und es gilt $a \wedge b \equiv \neg(\neg a \vee \neg b)$.

Selbsttest: Denken Sie sich einen beliebigen Term aus, der alle Junktoren aus der Standardsignatur verwendet. Wandeln Sie diesen Term in einen Term über Σ_1 bzw. über Σ_2 um. ?

2. $\Sigma_3 = \{\wedge, \vee\}$ ist nicht funktional vollständig. Betrachte die Belegung $\beta = (0, \dots, 0)$. Dann gilt $f_t(\beta) = 0$ für alle Terme t über Σ_3 . Damit kann keine Funktion f mit $f(\beta) = 1$ dargestellt werden.

3. $\Sigma_4 = \{\bar{\wedge}\}$ ist funktional vollständig. Dabei gilt $a \bar{\wedge} b \equiv \neg(a \wedge b)$. Wir simulieren die vollständige Signatur $\Sigma_1 = \{\neg, \wedge\}$ mithilfe von Σ_4 . Es gelten die folgenden Äquivalenzen:

$$\begin{aligned}\neg a &\equiv a \bar{\wedge} a \\ (a \wedge b) &\equiv \neg(\neg(a \wedge b)) \\ &\equiv \neg(a \bar{\wedge} b) \\ &\equiv (a \bar{\wedge} b) \bar{\wedge} (a \bar{\wedge} b)\end{aligned}$$

Analog kann gezeigt werden, dass $\Sigma_5 = \{\bar{\vee}\}$ funktional vollständig ist.

Selbsttest: Zeigen Sie, dass $\{\bar{\vee}\}$ funktional vollständig ist.



4. $\Sigma_6 = \{1, \oplus\}$ ist nicht funktional vollständig. Um dies zu zeigen, argumentieren wir, dass $a \wedge b$ nicht mit Σ_6 darstellbar ist. Dazu verwenden wir die folgenden Fakten:

$$t \oplus t \equiv 0 \quad (1.1)$$

$$t \oplus 0 \equiv t \quad (1.2)$$

$$t_1 \oplus t_2 \equiv t_2 \oplus t_1 \quad (1.3)$$

$$t_1 \oplus (t_2 \oplus t_3) \equiv (t_1 \oplus t_2) \oplus t_3 \quad (1.4)$$

Wir betrachten einen beliebigen Term s über der Signatur Σ_6 und formen diesen in mehreren Schritten zu einem äquivalenten Term um, über dessen Semantik wir einfacher argumentieren können.


Wir beobachten zunächst, dass wir nur einen Junktor haben und wegen (1.4) alle Klammern im Term einfach weglassen können.

- Sortiere die einzelnen Elemente des Terms so um, dass alle Vorkommen einer Variable und alle Vorkommen von 1 jeweils gruppiert sind. Dies ist wegen (1.3) erlaubt.
- Fasse alle doppelten Variablen und Konstanten zu 0 zusammen.
- Sortiere wieder um, so dann alle Vorkommen von Konstanten gruppiert sind.
- Wiederhole b) und c) so lange, bis jede Variable nur noch einmal vorkommt und pro Konstante nur noch ein Vorkommen vorhanden ist.
- Wenn zwei Konstanten vorhanden sind, ersetze das $0 \oplus 1$ durch 1.

Nun haben wir einen Term der Form $s' = 0 \oplus v_1 \oplus \dots \oplus v_k$ oder $s'' = 1 \oplus v_1 \oplus \dots \oplus v_k$. In diesem Term sind alle Variablen übrig geblieben, die im ursprünglichen Term s in ungerader Anzahl aufgetreten sind.

Der Term s' wird zu 0 ausgewertet, wenn eine gerade Anzahl von Variablen mit 1 belegt ist, und sonst zu 1. Der Term s'' wird zu 0 ausgewertet, wenn eine ungerade Anzahl an Variablen mit 1 belegt ist, und sonst zu 1.

In beiden Fällen ist die Anzahl der Belegungen die zu 0 bzw. 1 auswerten jeweils die Hälfte aller Belegungen. Damit kann unter anderem die Konjunktion nicht mit einem solchen Term dargestellt werden.

5. Die Signatur $\Sigma_7 = \{1, \oplus, \neg\}$ ist nicht funktional vollständig. Es gilt $\neg a \equiv a \oplus 1$, damit simuliert Σ_6 die Signatur Σ_7 und nach Satz 1.11 kann Σ_7 dann nicht funktional vollständig sein. 

KAPITEL 2

Beweistechniken

Voraussetzung:

- Grundlagen der Aussagenlogik (Abschnitt 1.1)
- Primzahlen, Teilbarkeit



Lernziele

Die Studierenden ...

- ... unterscheiden verschiedene Beweistechniken
- ... führen Beweise unter Verwendung der Beweistechniken



Dieses Kapitel beschäftigt sich mit einigen grundlegenden Beweistechniken. Warum es mit diesen möglich ist zu zeigen, dass eine gegebene (Elementar)aussage wahr ist, kann mithilfe der logischen Äquivalenzen, die wir in Abschnitt 1.1 kennen gelernt haben argumentiert werden. Bevor wir uns jedoch den technischen Details zuwenden, betrachten wir einige allgemeine Gedanken zum Aufschreiben von mathematischen Texten, insbesondere von Beweisen. Für einen detaillierteren Einstieg ist hier das Buch „Das ist o.B.d.A trivial“ von Albrecht Beutelspacher empfohlen.¹

2.1 Allgemeines zu mathematischen Beweisen

Die folgenden Gedanken sind nicht als absolute Wahrheit zu sehen, bieten aber gerade am Anfang eine gute Grundlage.

1. *Überlege, was du aufschreiben willst, bevor du es aufschreibst.* Bevor die Lösung der Übungsaufgabe in Reinform aufgeschrieben wird, sollte die Argumentationslinie schon klar sein. Hierfür bietet es sich an die Gedanken auf einem Schmierzettel, einer Tafel, ... zu sammeln.

¹<https://link.springer.com/book/10.1007/978-3-8348-9075-7> kostenlos aus dem Uninetz

2. *Auch ein mathematischer Text ist ein deutscher Text.* Das bedeutet zum einen, dass nicht unnötig viele Formeln und viel Formalismus verwendet werden sollte. Zum anderen sollte auf vollständige Sätze und korrekte Grammatik geachtet werden.
3. *Niemand mag Schachtelsätze.* Auch wenn die deutsche Sprache viele Möglichkeiten für verschachtelte Sätze gibt, sollten diese Möglichkeiten nicht unbedingt ausgenutzt werden. Ein Gedankengang ist in kurzen, klar getrennten Sätzen oft leichter zu verstehen.
4. *Finde eine nachvollziehbare Struktur.* Der Text soll nicht nur geschrieben, sondern auch gelesen werden. Lassen Sie den Text gerne etwas ruhen und lesen Sie ihn dann selber noch einmal. Sie sollten dann immer noch nachvollziehen können, was Sie mit dem Text aussagen wollten. Hier bietet es sich auch an, den Leser an die Hand zu nehmen. „Jetzt da wir A gezeigt haben, können wir B zeigen.“ oder „Wir zeigen die Aussage mithilfe von vollständiger Induktion.“
5. *Keine unbewiesenen Fakten verwenden.* Dieser Punkt ist weitestgehend selbsterklärend. Im Text sollte so weit wie möglich eine Struktur gefunden werden, sodass alle Bezeichner und Fakten die verwendet werden, vor der ersten Verwendung eingeführt werden.
6. *Wähle sinnvolle Bezeichner.* Ähnliche Dinge sollten ähnlich bezeichnet werden, zum Beispiel mit Buchstaben, die im Alphabet nah aneinander sind (X, Y, Z) oder mit Indizes X_1, X_2, X_3 . Es gibt einige Bezeichner, die für bestimmte Arten von Objekten reserviert sind. Zum Beispiel f, g, h für Funktionen, n, m für natürliche Zahlen, i, j, k für Indizes oder ε für kleine positive reelle Zahlen.
7. *Keine Abkürzungen!* Gerade am Anfang sollten Formulierungen wie „Offensichtlich gilt“ vermieden werden. Genauso sollte der Beweis zu Ende geführt werden.

2.2 Beweise von Implikationen

Viele mathematischen Aussagen haben die Form einer Implikation, also „Wenn p gilt, dann gilt auch q “. Um diese Art von Aussagen zu beweisen, gibt es verschiedene Beweistechniken, die wir uns im Folgenden anschauen. All diese Techniken können im Prinzip auch genutzt werden, um Aussagen ohne Implikation, wie zum Beispiel „ $\sqrt{2}$ ist keine rationale Zahl“ zu zeigen. Dann wird statt $p \rightarrow q$ die Aussage $1 \rightarrow q$, wobei q die zu zeigende Aussage ist, gezeigt.

Soll keine Implikation, sondern eine Äquivalenz ($p \leftrightarrow q$) gezeigt werden, kann diese wegen $p \leftrightarrow q \equiv p \rightarrow q \wedge q \rightarrow p$ in zwei Implikationen zerlegt werden. Werden beide Richtungen getrennt gezeigt, wurde auch die Äquivalenz gezeigt.

2.2.1 Direkter Beweis

Statt die Implikation $p \rightarrow q$ direkt zu zeigen, zerlegen wir sie in kleine Schritte. Wir wählen also eine Aussage r als Zwischenschritt und zeigen erst $p \rightarrow r$ und dann $r \rightarrow q$. Wenn nötig, können noch weitere Zwischenschritte eingefügt werden. Dass wir damit dann $p \rightarrow q$ gezeigt haben, ergibt sich daraus, dass $((p \rightarrow r) \wedge (r \rightarrow q)) \rightarrow (p \rightarrow q)$ eine Tautologie, also eine immer wahre Aussage ist.

Um die einzelnen Teilschritte zu zeigen, bleibt man in einem System, dass auf einigen Grundannahmen (Axiomen) aufbaut, und diese Grundannahmen als bereits bewiesene Tatsachen betrachtet.

Beispiel. Wir wollen die folgende Aussage zeigen.

Ist eine natürliche Zahl n durch 6 teilbar, so ist ihr Quadrat durch 9 teilbar.

Zunächst beschreiben wir die Beweisidee verbal, bevor wir den Beweis formal führen.² Die Idee, die wir verwenden, ist, dass wenn n durch 6 teilbar ist, n natürlich auch durch 3 teilbar ist. Damit ist die 3 in n^2 zweimal enthalten und damit ist n^2 durch 9 teilbar.

Für den formalen Beweis brauchen wir zunächst die Definition der Teilbarkeit. Eine natürliche Zahl $n \in \mathbb{N}$ ist durch $k \in \mathbb{N}$ **teilbar**, falls ein $l \in \mathbb{N}$ existiert, sodass $n = k \cdot l$. Wir schreiben dann auch kurz $k \mid n$.

Nun können wir die Implikation Schritt für Schritt zeigen.

	$6 \mid n$	Hypothese
$\exists l \in \mathbb{N}$	$n = 6 \cdot l$	Definition der Teilbarkeit
$\exists l \in \mathbb{N}$	$n = (3 \cdot 2) \cdot l$	$6 = 3 \cdot 2$
$\exists l \in \mathbb{N}$	$n^2 = ((3 \cdot 2) \cdot l) \cdot ((3 \cdot 2) \cdot l)$	Quadrieren
$\exists l \in \mathbb{N}$	$n^2 = (3 \cdot 3) \cdot (2 \cdot 2) \cdot (l \cdot l)$	\cdot ist assoziativ und kommutativ

Beispiel. Wir wollen die folgende Aussage zeigen.

Sei $n \in \mathbb{N}$. Wenn n gerade ist, dann ist auch n^2 gerade.

Wir führen einen direkten Beweis. Da n eine gerade Zahl ist, gibt es ein $k \in \mathbb{N}$ mit $n = 2 \cdot k$. Also gilt:

$$\begin{aligned} n^2 &= (2 \cdot k)^2 \\ &= 2 \cdot 2k^2 \end{aligned}$$

Also ist auch n^2 gerade.

²Dies ist auch insgesamt eine gute Idee, einen Gedanken zunächst verbal zu formulieren, bevor der formale Beweis geführt wird.

Selbsttest: Zeigen Sie mit einem direkten Beweis, dass wenn n ungerade ist, dann auch n^2 ungerade ist.



2.2.2 Indirekte Beweise

Oft ist es nicht ohne weiteres möglich eine Implikation direkt zu zeigen. Hier kommen dann die sogenannten *indirekten Beweise* ins Spiel.

Beweis durch Kontraposition Statt direkt $p \rightarrow q$ zu zeigen, zeigen wir die logisch äquivalente Aussage $\neg q \rightarrow \neg p$. Der Beweis der Kontraposition verwendet dann meistens einen direkten Beweis oder kombiniert andere Beweistechniken, die wir noch kennenlernen werden.

Beispiel. Wir zeigen die Aussage:

Wenn n^2 ungerade ist, dann ist auch n ungerade.

Wir führen einen Beweis durch Kontraposition. In diesem Beispiel ist p die Aussage, n^2 ist ungerade und q die Aussage n ist ungerade.

Die Kontraposition ist nun also *Wenn n gerade ist, dann ist auch n^2 gerade*. Das ist genau die Aussage, die wir im zweiten Beispiel zum direkten Beweis schon gezeigt haben und wir sind fertig. ◀

Beweis durch Widerspruch Eng verwandt zum Beweis durch Kontraposition ist der Beweis durch Widerspruch. Auch hier wird eine logisch äquivalente Formulierung zu $p \rightarrow q$ gezeigt, nämlich $(p \wedge \neg q) \rightarrow 0$.

Widerspruchsbeweise werden auch häufig für Aussagen ohne Implikation verwendet, in diesem Fall wird gezeigt $\neg q \rightarrow 0$, also die Annahme, dass die zu zeigende Aussage nicht gilt, führt zu einem Widerspruch.

Wir betrachten Beispiele für beide Varianten:

Beispiel. Wir zeigen die Aussage:

Wenn n^2 ungerade ist, dann ist auch n ungerade.

Wir führen einen Beweis durch Widerspruch. Also angenommen n^2 ist ungerade und n ist gerade. Dann ist n^2 gerade (siehe Argumente oben). Damit wäre n^2 sowohl gerade als auch ungerade, ein Widerspruch. ◀

Beispiel. Wir zeigen die Aussage:

$\sqrt{2}$ ist keine rationale Zahl.

Wir führen den Beweis durch Widerspruch. Angenommen, $\sqrt{2}$ wäre eine rationale Zahl. Dann können wir diese schreiben als $\sqrt{2} = \frac{p}{q}$, wobei p und q teilerfremde³ ganze


³Das bedeutet, dass der Bruch $\frac{p}{q}$ nicht gekürzt werden kann

Zahlen sind. Es gilt also:

$$\begin{aligned}\sqrt{2} &= \frac{p}{q} && |^2 \\ 2 &= \frac{p^2}{q^2} && | \cdot q^2 \\ p^2 &= 2q^2\end{aligned}$$

Also ist p^2 durch zwei teilbar und damit eine gerade Zahl. Mit einem ähnlichen Beweis wie den Beispielen oben kann gezeigt werden, dass damit auch p eine gerade Zahl ist.⁴ Daher gibt es ein k , sodass $p = 2 \cdot k$ ist. Dies können wir dann in die Gleichung $2 = \frac{p^2}{q^2}$ einsetzen und erhalten:

$$\begin{aligned}2 &= \frac{(2k)^2}{q^2} && | \cdot q^2 \\ 2q^2 &= 4k^2 && | \div 2 \\ q^2 &= 2k^2\end{aligned}$$

Also ist q^2 und damit auch q durch zwei teilbar. Dies ist ein Widerspruch zur Annahme, dass p und q teilerfremd sind. 

Beweis durch Fallunterscheidung Für den Beweis einer Aussage $p \rightarrow q$ oder $1 \rightarrow q$ ist es häufig hilfreich mehrere Fälle zu betrachten. Dies ist erlaubt, da $(p \wedge r) \rightarrow q \wedge (p \wedge \neg r) \rightarrow q$ logisch äquivalent zu $p \rightarrow q$ ist.

Die Fälle können feiner aufgeteilt werden, zum Beispiel $r \wedge t$, $r \wedge \neg t$ und $\neg r$.

Beispiel. Wir zeigen die folgende Aussage:


Es gilt $q^2 \geq q$ für alle $q \in \mathbb{Z}$

Fall 1: $q = 0$ Für den Fall $q = 0$ können wir das Ergebnis einfach ausrechnen. Es gilt $0^2 = 0$ und die Aussage gilt in diesem Fall.

Fall 2: $q < 0$ In diesem Fall ist $q^2 > 0$, da das Quadrat einer negativen Zahl immer positiv ist. Damit gilt $q < 0 < q^2$ und die Aussage wurde gezeigt.

Fall 3: $q > 0$ Da wir ganze Zahlen betrachten, ist dies äquivalent zu $q \geq 1$. Wir haben also

$$\begin{aligned}q &\geq 1 && | \cdot q \\ q^2 &\geq q\end{aligned}$$

und die Aussage ist gezeigt. 

⁴Dies ist eine gute Übungsaufgabe, um das Verständnis der Beweistechniken zu überprüfen

Selbsttest: Zeigen Sie, dass für alle $x \in \mathbb{R}$ und alle $y \in \mathbb{R}$ gilt, dass $\max(x, y) + \min(x, y) = x + y$ ist.

Welche Beweistechnik(en) haben Sie verwendet?



2.3 Beweis durch vollständige Induktion

Klost: In diesem Abschnitt werden eventuell nach und nach noch weitere Beispiele ergänzt.

Dieser Abschnitt beschäftigt sich mit dem Beweis durch vollständige Induktion. Diese Beweistechnik ist ein sehr mächtiges Werkzeug, um Allaussagen für die natürlichen Zahlen zu zeigen. Sie kann aber auch allgemeiner angewendet werden, um Aussagen über Strukturen zu zeigen, die mit natürlichen Zahlen beschrieben werden können.

2.3.1 Charakterisierung der natürlichen Zahlen

Die Grundlage für die Korrektheit von Beweisen mit vollständiger Induktion sind die *Peano-Axiome*. Dies sind 5 Grundannahmen, aus den alle aus der Schulmathematik bekannten Aussagen über den natürlichen Zahlen abgeleitet werden können.

Axiom 1 0 ist eine natürliche Zahl.

Axiom 2 Jede natürliche Zahl n hat einen eindeutigen Nachfolger $S(n)$, der auch eine natürliche Zahl ist.

Axiom 3 Aus $S(n) = S(m)$ folgt $n = m$.

Axiom 4 0 ist kein Nachfolger einer natürlichen Zahl.

Axiom 5 Jede Menge X , die 0 enthält und für die aus $n \in X$ aus $S(n) \in X$ folgt, enthält alle natürlichen Zahlen.

Eine Visualisierung der ersten vier Axiome findet sich in Abbildung 2.1.

Anmerkung: Oft schreibt man auch $n + 1$ statt $S(n)$. Für die Formulierung der Peano-Axiome wurde hier jedoch bewusst $S(n)$ gewählt, da $n + 1$ an dieser Stelle eben *nicht* die Addition mit 1, sondern ein syntaktisches Konstrukt für den Nachfolger einer Zahl darstellt.

2.3.2 Beweis durch vollständige Induktion

Aus den Peano-Axiomen folgt eine Möglichkeit, um eine Aussage für alle natürlichen Zahlen zu zeigen.

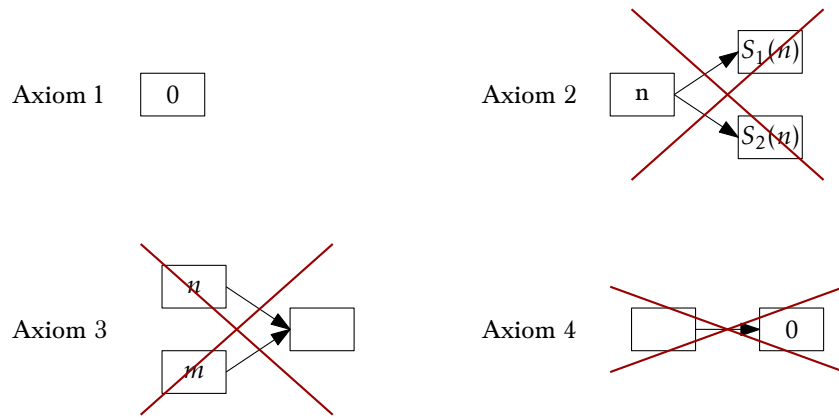


Abbildung 2.1: Visualisierung der ersten vier Peano-Axiome

Satz 2.1. Eine Aussage der Form $\forall n \in \mathbb{N} : P(n)$ ist wahr, wenn das folgende gilt:

1. $P(0)$ ist wahr.
2. Für beliebige $n \in \mathbb{N}$ gilt: Ist $P(n)$ wahr, dann ist auch $P(n+1)$ wahr.

Beweis. Sei $W \subseteq \mathbb{N}$ die Menge aller natürlichen Zahlen, für die $P(n)$ wahr ist. Wir wollen nun die Informationen aus dem Satz verwenden, um zu zeigen, dass das fünfte Peano-Axiom für W gilt, und daher $W = \mathbb{N}$ ist.

Wegen 1. ist $0 \in W$ und der erste Teil der Voraussetzung für das 5. Axiom gilt.

Nun besagt 2. ja genau, dass aus $n \in W$ auch $(n+1) \in W$ gilt. Dies ist der zweite Teil der Voraussetzung des 5. Peano-Axioms und damit gilt $W = \mathbb{N}$ und die Aussage gilt für alle $n \in \mathbb{N}$. \square

Eine Analogie für das Konzept der vollständigen Induktion ist das Hochklettern einer unendlich langen Leiter. Die Aufgabe ist es zu zeigen, dass es möglich ist alle Sprossen dieser unendlich langen Leiter zu erreichen⁵. Da die Leiter unendlich lang ist, ist es nicht möglich, dies für jede Sprosse einzeln zu zeigen. Aber angenommen wir haben gezeigt:

1. Die unterste Sprosse kann erreicht werden und
2. wenn eine Sprosse erreicht wurde, kann die nächste Sprosse erreicht werden.

Dann können alle Sprossen erreicht werden, denn es kann von der untersten Sprosse zunächst zur zweituntersten Sprosse, dann zur drittuntersten Sprosse und so weiter geklettert werden.

Genau das gleiche Prinzip verwendet die vollständige Induktion. Zunächst wird die Aussage für $n = 0$ gezeigt. Dann wird gezeigt, dass unter der Annahme, dass die Aussage für ein beliebiges n schon gilt (eine beliebige Sprosse wurde erreicht), die Aussage auch für $n+1$ gilt (die nächste Sprosse wird erreicht).

⁵Unter der Annahme, dass beim Klettern die Kraft nie ausgeht

Vor dem ersten Beispiel schauen wir uns noch einige Begriffe an. Die zu zeigende Aussage der Form $\forall n \in \mathbb{N} : P(n)$ ist die Aussage, die wir zeigen wollen (zu zeigen). $P(0)$, also der Fakt, dass die Aussage für 0 gilt, heißt **Induktionsanfang**, **Induktionsanker** oder auch **Induktionsbasis**⁶. Die Annahme, dass die Aussage für ein beliebiges n gilt, ist die **Induktionsvoraussetzung** oder **Induktionsannahme**. Die Folgerung, dass die Aussage dann auch für $n + 1$ gilt ist dann der **Induktionsschritt**. Innerhalb des Induktionsschritts wird die Aussage $P(n + 1)$ häufig auch als **Induktionsbehauptung** bezeichnet.

Wir schauen nun eine Folge von Beispielen für dieses Prinzip an.

Beispiel. In diesem Beispiel zeigen wir die geschlossene Form der Gaußschen Summenformel mithilfe von vollständiger Induktion.

Zu zeigende Aussage

$$\forall n \in \mathbb{N} : \sum_{i=0}^n i = \frac{n(n+1)}{2}$$

Induktionsanker Im Induktionsanker wird die Aussage für $n = 0$ gezeigt. Dies geschieht hier durch einfaches Ausrechnen: Es gilt $\sum_{i=0}^0 i = 0$ und $\frac{0 \cdot (0+1)}{2} = 0$, also ist die Aussage für $n = 0$ gezeigt.

Induktionsvoraussetzung Für ein beliebiges aber festes n' gilt $\sum_{i=0}^{n'} i = \frac{n'(n'+1)}{2}$

Induktionsschritt Nun zeigen wir, dass aus der Induktionsvoraussetzung folgt, dass die Aussage auch für $n' + 1$ gilt.

$\sum_{i=0}^{n'+1} i = \left(\sum_{i=0}^{n'} i \right) + (n' + 1)$	Summe auseinandernehmen
$= \frac{n'(n' + 1)}{2} + (n' + 1)$	I.V.
$= \frac{n'(n' + 1)}{2} + \frac{2(n' + 1)}{2}$	erweitern
$= \frac{n'(n' + 1) + 2(n' + 1)}{2}$	auf einen Nenner bringen
$= \frac{(n' + 2)(n' + 1)}{2}$	ausklammern

Da alle Schritte aus Satz 2.1 gezeigt wurden, ist die Aussage gezeigt. ◀

Hier haben wir die im Schritt verwendete Variable von der Variable aus der Behauptung abzugrenzen einmal n und einmal n' verwendet. In der Praxis wird allerdings an beiden Stellen meistens einfach n verwendet, in der Annahme das dem Lesenden aus dem Kontext klar ist, was genau gemeint ist.

⁶Dadurch, dass das Prinzip der vollständigen Induktion sehr weit verbreitet ist, haben sich verschiedene Begriffe eingebürgert.

Da die Beweise mittels vollständiger Induktion sehr stark standardisiert sind, wird vieles vom Fließtext oft weggelassen und die Namen der einzelnen Schritte werden abgekürzt.

Beispiel. Z.z.

$$\forall n \in \mathbb{N} : n < 2^n \quad (2.1)$$

I.A. ($n = 0$) $0 < 2^0 = 1$ ist wahr.

I.V. $n < 2^n$ gilt für ein beliebiges festes n .

I.S. ($n \rightarrow n + 1$)

$$\begin{aligned} n + 1 &\leq 2^n + 1 && | \text{I.V} \\ &\leq 2^n + 2^n && | 1 \leq 2^n \\ &= 2^{n+1} && \triangleleft \end{aligned}$$

Varianten der Induktion Es gibt einige Varianten der Induktion, die häufig auftreten.

1. Der Induktionsanker kann nicht nur bei $n = 0$, sondern bei $n = k$ für ein $k > 0$ sein. Wird der Rest des Beweises mittels vollständiger Induktion dann genauso geführt wie bisher, gilt die Aussage für alle $n \geq k$.
2. In einigen Fällen ist es hilfreich, im Induktionsschritt nicht nur zu verwenden, dass die Aussage für ein beliebiges n gilt, sondern dass sie für alle Werte zwischen dem Anker und n gilt. Also statt $P(n) \rightarrow P(n + 1)$ wird gezeigt $P(k) \wedge P(k + 1) \wedge \dots \wedge P(n) \rightarrow P(n + 1)$, wobei k der Induktionsanker ist.

Selbsttest: Zeigen Sie $\forall n \in \mathbb{N} : \sum_{i=0}^n (2i + 1) = (n + 1)^2$ mit Hilfe vollständiger Induktion.



Die zweite Variante wird häufig auch als **verallgemeinerte vollständige Induktion** bezeichnet.

Beispiel. Wir zeigen die folgende Aussage mithilfe von verallgemeinerter vollständiger Induktion.

Jede natürliche Zahl $n \geq 2$ kann als Produkt von Primzahlen geschrieben werden.⁷

I.A. Die Aussage soll für alle $n \geq 2$ gelten, der Anker ist also bei $n = 2$. Da 2 eine Primzahl ist, gilt die Aussage.

I.V. Sei n eine beliebige natürliche Zahl. Die Zahlen $2, \dots, n$ können als Produkt von Primzahlen geschrieben werden.

I.S. ($2, \dots, n \rightarrow n + 1$) Wir verwenden einen Beweis durch Fallunterscheidung.

⁷Wir erlauben Produkte, die nur aus einem Faktor bestehen.

Fall 1: $n + 1$ ist eine Primzahl In diesem Fall ist nichts zu tun, da die Aussage direkt gilt.

Fall 2: $n + 1$ ist keine Primzahl Wenn $n + 1$ keine Primzahl ist, dann kann $n + 1$ als $n + 1 = p \cdot q$ geschrieben werden, wobei p und q beliebige Zahlen zwischen 2 und n sind. Nach I.V. können p und q jeweils als Produkt $p = p_1 \cdot p_2 \cdots p_k$ und $q = q_1 \cdot q_2 \cdots q_l$ von Primzahlen geschrieben werden.

Damit ist $n + 1 = p_1 \cdot p_2 \cdots p_k \cdot q_1 \cdot q_2 \cdots q_l$ ein Produkt von Primzahlen. ◀

Beispiel. Wir zeigen die folgende Aussage mit verallgemeinerter vollständiger Induktion

Jede natürliche Zahl ≥ 12 lässt sich als Summe schreiben, in der alle Summanden 4 oder 5 sind.

Z.z. (Formale Version der Aussage)

$$\forall n \in \mathbb{N}, n \geq 12 \exists k, l \in \mathbb{N} : n = k \cdot 4 + l \cdot 5$$

I.A. Dass die Aussagen $P(12), P(13), P(14)$ und $P(15)$ gelten kann leicht nachgerechnet werden.

I.V. Die Aussagen $P(12), P(13), \dots, P(n)$ gelten für ein beliebiges festes $n \geq 15$.

I.S. $(12, \dots, n \rightarrow n + 1)$ Betrachte $P(n + 1 - 4) = P(n - 3)$. Da wir $n \geq 15$ annehmen, ist $n - 3 \geq 12$ und damit ist nach I.V. die Aussage $P(n - 3)$ wahr. Es gilt also:

$$n - 3 = k \cdot 4 + l \cdot 5$$

$$n + 1 = k \cdot 4 + l \cdot 5 + 4$$

$$n + 1 = (k + 1) \cdot 4 + l \cdot 5$$

und die Aussage wurde mit dem Prinzip der vollständigen Induktion gezeigt. ◀

Im letzten Beispiel sind wir nicht nur einen, sondern genau 4 Schritte „rückwärts“ von $n - 1$ gegangen. Damit dies für alle Zahlen, die nicht im Anker sind, möglich ist, reicht es nicht den Anker für $n = 12$ zu zeigen, sondern es müssen alle Werte bis einschließlich 15 im Anker betrachtet werden.

Zum Abschluss noch ein berühmter **falscher** Beweis mittels Induktion.

Beispiel. Die Aussage die betrachtet wird ist:

In einer Gruppe von n Pferden, haben alle Pferde die gleiche Farbe. (Induktionsbehauptung)

I.A. ($n = 1$) In einer Gruppe von einem Pferd gibt es nur eine Farbe.

I.V. Für ein beliebiges n gilt, dass alle Pferde in einer Gruppe von n Pferden die gleiche Farbe haben.

I.S. ($n \rightarrow n+1$) Betrachte eine Gruppe G aus $n+1$ Pferden p_1, \dots, p_n, p_{n+1} . Erzeuge daraus zwei Teilgruppen G_1 und G_2 . Dabei sind in G_1 die Pferde p_1, \dots, p_n und in G_2 die Pferde p_2, \dots, p_{n+1} .

Nach Induktionsvoraussetzung haben alle Pferde in G_1 die gleiche Farbe und alle Pferde in G_2 die gleiche Farbe. Da jedoch Pferde in beiden Mengen enthalten sind, müssen alle Pferde die gleiche Farbe haben.

Was ist falsch? Das Argument im Induktionsschritt funktioniert nicht für $n = 1$. In diesem Fall ist in G_1 nur p_1 enthalten und in G_2 nur p_2 . Für jede Menge an sich gilt die Aussage zwar, es gibt aber keine gemeinsamen Pferde in den Gruppen und daher gilt die letzte Schlussfolgerung nicht.

Wird nun versucht den Beweis anzupassen, dass $n = 2$ im Anker statt im Schritt betrachtet wird, ist dies nicht möglich, da sich Mengen von zwei Pferden finden, in denen nicht beide die gleiche Farbe haben. ◀

Vollständige Induktion kann nicht nur für Aussagen über den natürlichen Zahlen verwendet werden, sondern auch für Strukturen, die sich mithilfe von natürlichen Zahlen beschreiben lassen. Dieses Verfahren wird dann auch oft **strukturelle Induktion** genannt.

Beispiel. Wir zeigen die folgende Aussage mittels struktureller Induktion. Die passende vollständige Induktion läuft über die Anzahl der Ecken.

Z.z. Für alle $n \geq 3, n \in \mathbb{N}$ gilt: Ein konvexes n -eck⁸ hat $\frac{n \cdot (n-3)}{2}$ Diagonalen.

I.A. Für $n = 3$ betrachten wir ein Dreieck. Ein Dreieck hat keine Diagonalen. Da $\frac{3 \cdot 0}{2} = 0$ gilt die Aussage für $n = 3$.

I.V. Für ein beliebiges $n \geq 3$ gilt, dass ein konvexes n -eck $\frac{n \cdot (n-3)}{2}$ Diagonalen hat.

I.S. ($n \rightarrow n+1$) Siehe Abbildung 2.2 für eine Visualisierung. Seien p_1, \dots, p_n, p_{n+1} die Ecken eines konvexen $n+1$ -ecks. Wir nennen es P_{n+1} . Dann liegen an p_{n+1} genau $n-2$ Diagonalen an (zu p_2, \dots, p_{n-1}). Die Ecken p_1, \dots, p_n bilden wieder ein konvexes n -eck P_n . Nach I.V. hat P_n genau $\frac{n \cdot (n-3)}{2}$ Diagonalen. Zusätzlich wird die Verbindung zwischen p_1 und p_n jetzt neu zur Diagonale. Damit hat P_{n+1} genau $(n-2) + 1 + \frac{n \cdot (n-3)}{2}$ Diagonalen. Es gilt

$$\begin{aligned} & (n-2) + 1 + \frac{n \cdot (n-3)}{2} \\ &= (n-1) + \frac{n \cdot (n-3)}{2} \end{aligned}$$

⁸konvex bedeutet hier, dass alle Ecken mit allen anderen durch eine Strecke verbunden werden können, ohne das innere des n -ecks zu verlassen

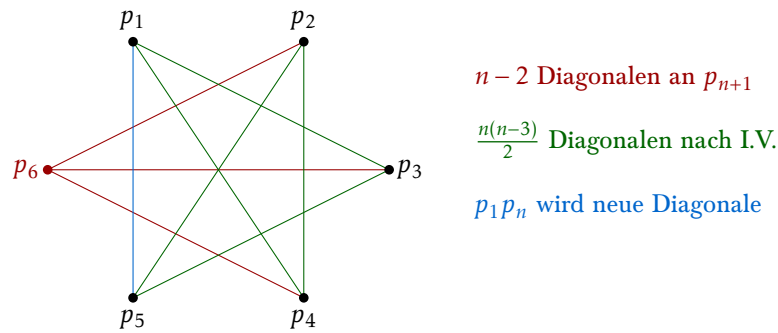


Abbildung 2.2: Induktionsschritt von $n = 5$ nach $n + 1 = 6$

$$\begin{aligned}
 &= \frac{2n - 2 + n \cdot (n - 3)}{2} \\
 &= \frac{2n - 2 + n^2 - 3n}{2} \\
 &= \frac{n^2 - n - 2}{2} \\
 &= \frac{(n + 1) \cdot (n - 2)}{2}
 \end{aligned}$$

2.4 Das Schubfachprinzip

Voraussetzung: Mengen (Abschnitt 3.1) und Funktionen (Abschnitt 3.3)

Das Schubfachprinzip, auch Taubenschlagprinzip genannt, ist eine Verallgemeinerung einer eigentlich einfachen Beobachtung. Dieses Prinzip kann jedoch an vielen Stellen in Beweisen angewendet werden und hat damit auch einen eigenen Namen bekommen. Um die Aussage zu verstehen, bietet es sich an, zunächst das erste Beispiel und dann erst den Beweis anzuschauen.

Satz 2.2 (Schubfachprinzip). Seien A und B endliche Mengen und $f : A \rightarrow B$ eine Funktion. Dann existiert ein $b_0 \in B$, sodass

$$|f^{-1}(b_0)| \geq \left\lceil \frac{|A|}{|B|} \right\rceil$$

Beweis. Sei $B' = \{b \in B \mid f^{-1}(b) \neq \emptyset\}$. Dann ist $(f^{-1}(b))_{b \in B'}$ eine Partition von A , also $|A| = \sum_{b \in B'} |f^{-1}(b)|$. Sei nun $b_0 = \arg \max_{b \in B'} |f^{-1}(b)|$ das Element aus B' mit dem größten

Urbild⁹, dann gilt:

$$\begin{aligned} |A| &= \sum_{b \in B} |f^{-1}(b)| \\ &\leq \sum_{b \in B} |f^{-1}(b_0)| && \text{da } |f^{-1}(b_0)| \geq |f^{-1}(b)| \quad \forall b \in B \\ &= |B| \cdot |f^{-1}(b_0)| \end{aligned}$$

Und $|f^{-1}(b_0)| \geq \frac{|A|}{|B|}$ folgt. Da $|f^{-1}(b_0)|$ ganzzahlig ist, $\frac{|A|}{|B|}$ aber im Allgemeinen nicht, dürfen wir aufrunden. \square

Im Folgenden schauen wir uns einige Beispiel an, die dieses Prinzip verwenden.

Beispiel. Das erste Beispiel macht klar, woher das Schubfachprinzip seinen Namen hat:

Sei A eine Menge von 10 Briefen und B eine Menge von 3 Schubfächern. Dann gibt es mindestens ein Schubfach, in dem mindestens 4 Briefe liegen. Dies ergibt sich aus dem Schubfachprinzip mit $f(a) = b$, wenn Brief a in Schubfach b liegt. Dann gibt es ein Schubfach $b_0 \in B$ mit mindestens $\left\lceil \frac{10}{3} \right\rceil$ Briefen. \blacktriangleleft

Nun wenden wir uns zwei Beispielen zu, die ebenfalls das Schubfachprinzip verwenden, jedoch etwas komplexer sind.

Beispiel. Wir wollen zeigen, dass es in einer Gruppe A von 6 Personen immer mindestens drei Personen gibt, die sich paarweise kennen oder drei Personen, die sich paarweise nicht kennen.

Wir stellen die Situation grafisch dar, siehe Abbildung 2.3. Jeder Punkt stellt eine Person dar. Eine grüne Strecke zwischen zwei Punkten bedeutet, dass sich die Personen kennen, eine rote, dass sie sich nicht kennen. Die Aussage, dass sich drei Personen paarweise kennen oder paarweise nicht kennen kann dann durch ein einfarbiges Dreieck dargestellt werden.

Seien a eine der Personen in der Gruppe. Dann hat a ausgehende Verbindungen zu allen anderen 5 Personen. Sei $f : A \setminus \{a\} \rightarrow \{\text{kennen, nicht kennen}\}$ die Funktion, mit $f(x) = \text{nicht kennen}$, wenn a Person x nicht kennt und $f(x) = \text{kennen}$, wenn a Person x kennt. Nach dem Schubfachprinzip gibt es jetzt $\left\lceil \frac{5}{2} \right\rceil$ Personen x, y, z , sodass a entweder x, y, z alle kennt oder nicht kennt.

Die folgende Argumentation ist in Abbildung 2.4 visualisiert.

Fall 1: a kennt x, y und z In diesem Fall gibt es zwei Unterfälle:

Fall 1a: x, y, z kennen sich paarweise nicht. In dem Fall sind wir fertig, da x, y, z das gesuchte Dreieck bilden.

⁹arg max bildet nicht das normale Maximum über den Ausdruck, sondern gibt das Element zurück, welches den Wert maximiert hat

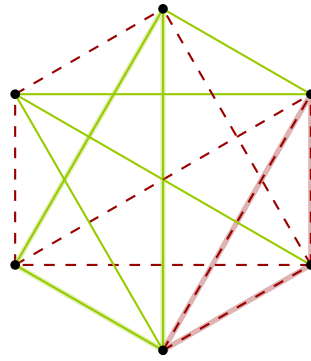


Abbildung 2.3: Zwei Personen sind mit einer grünen Linie verbunden, wenn sie sich kennen und mit einer roten gestrichelten Linie, wenn sie sich nicht kennen.

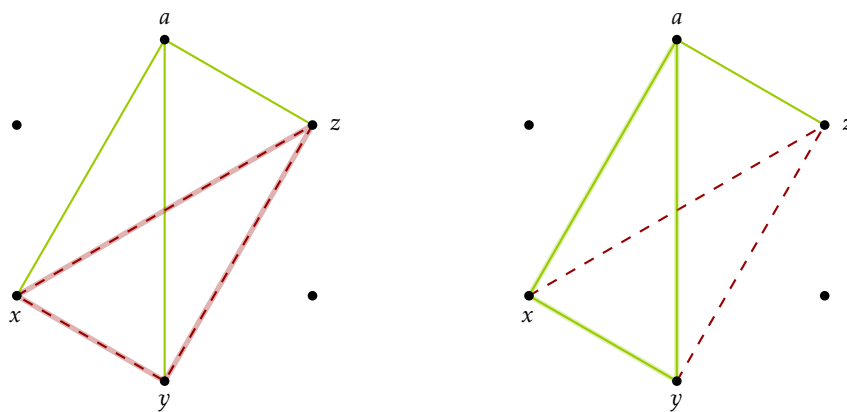


Abbildung 2.4: Links: Fall 1a, Rechts: Fall 1b

Fall 1b: Zwei Personen aus $\{x, y, z\}$ kennen sich. Wir gehen ohne Beschränkung der Allgemeinheit davon aus, dass dies x, y sind. Dann bilden a, x, y das gesuchte Dreieck.

Fall 2: a kennt x, y und z nicht Analog. ◀

Das dritte Beispiel ist geometrischer Natur:

Beispiel. Eine Visualisierung findet sich in Abbildung 2.5. Sei Q ein Quadrat, dessen Diagonale Länge 2 hat. Wenn wir jetzt die Mittelpunkte von 10 Kreisen, jeder mit Radius $\frac{1}{3}$ in Q verteilen, gibt es mindestens ein Paar von Kreisen, welches sich schneidet.

Um diese Aussage zu zeigen, unterteilen wir Q nochmal in 9 disjunkte Quadrate mit jeweils eine Diagonale mit Länge $\frac{2}{3}$. Nach dem Schubfachprinzip gibt es mindestens eines der kleineren Quadrate, in dem $\left\lceil \frac{10}{9} \right\rceil = 2$ Kreise sind.

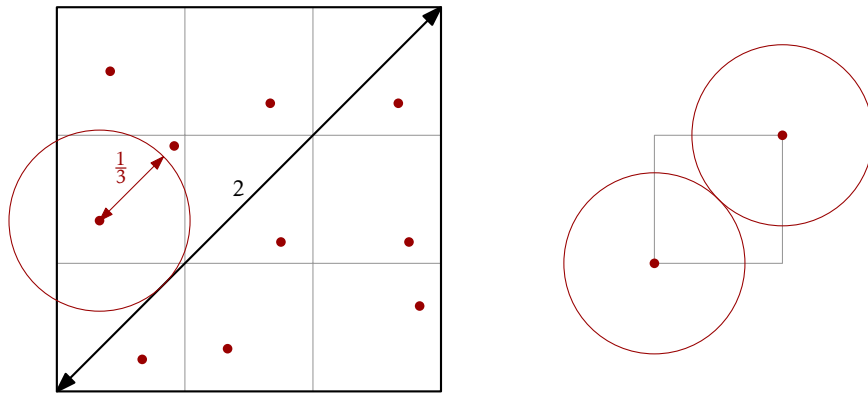


Abbildung 2.5: Links: Die Ausgangssituation mit 10 Mittelpunkten und der Unterteilung in kleine Quadrate, Rechts: Zwei Kreise mit Mittelpunkten im gleichen kleinen Quadrat schneiden sich.

Da der Radius der Kreise jeweils $\frac{1}{3}$ ist, der maximal mögliche Abstand von zwei Mittelpunkten in einem der kleinen Quadrate aber $\frac{2}{3}$, schneiden sich die beiden Kreise. ◀

KAPITEL 3

Mengen

Im Folgenden betrachten wir *Mengen* als mathematische Objekte. Die moderne Mengentheorie basiert auf einer Menge von Axiomen¹. Diese ist jedoch sehr technisch und die eigentlich sehr grundlegenden Konzepte werden durch den Formalismus verdeckt. Daher betrachten wir in diesem Abschnitt eine „intuitive“ oder „naive“ Mengenlehre betrachten. Diese ist vollkommen ausreichend für den weiteren Verlauf des Skripts und der Vorlesung.

3.1 Mengen

Voraussetzung:

- Aussagenlogik (Abschnitt 1.1), Prädikatenlogik (Abschnitt 1.2)
- Beweistechniken (Kapitel 2)
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$



Lernziele

Die Studierenden ...

- ... verwenden Mengenbegriffe
- ... führen Beweise zu Eigenschaften von Mengen



3.1.1 Einführung

¹als wahr angenommenen Grundsätzen

Definition 3.1 (G.Cantor (1895)). Eine **Menge** ist eine Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen. Die einzelnen Objekte in einer Menge werden **Elemente** genannt.

Wenn ein Element x in einer Menge A enthalten ist, schreiben wir $x \in A$, sonst schreiben wir $x \notin A$. Für endliche Mengen bezeichnet $|A|$ die **Größe** der Menge. Zwei Mengen sind **gleich** ($A = B$) genau dann, wenn sie die gleichen Elemente haben. Also, wenn gilt $x \in A \leftrightarrow x \in B$.

Darstellung von Mengen Mengen können auf verschiedene Art und Weisen dargestellt werden:

1. Die Elemente der Menge können explizit in geschweiften Klammern aufgezählt werden. Dabei ist die Reihenfolge, in der die Elemente aufgezählt werden, irrelevant. Werden Elemente mehrfach aufgezählt, hat dies keinen Einfluss auf die Menge.

Wenn unendliche Mengen nach einem Muster erzeugt werden, können diese auch explizit dargestellt werden.

Beispiel.

$$A = \{3, 2, 5, 4, 3\} = \{2, 3, 4, 5\}$$

$$B = \{0, 2, 4, \dots\}$$

$$\mathbb{N} = \{0, 1, \dots\}$$

alle gerade Zahlen aus \mathbb{N}

2. Meistens werden Mengen durch Angabe einer Grundmenge und eines Prädikats angegeben. Dies spiegelt das *Abstraktionsprinzip* wider, dass wir Mengen durch bestimmte Eigenschaften definieren können. Die Schreibweise dafür ist

$$M = \{x \in U \mid P(x)\}.$$

Die Menge M enthält dann alle Elemente aus der Grundmenge U , welche das Prädikat $P(x)$ erfüllen.

Beispiel. Die Menge der geraden natürlichen Zahlen kann dann als

$$\{x \in \mathbb{N} \mid x \text{ gerade} \}$$

dargestellt werden.

Hinweis: In dieser Darstellung kann man schnell zu paradoxen Definitionen kommen, wie $M = \{x \mid x \notin M\}$. Dann ist $x \in M$ genau dann, wenn $x \notin M$, also ein Widerspruch.

3. Zur Veranschaulichung von Argumenten, können Mengen durch Flächen in der Ebene dargestellt werden. Die Grundmenge wird dann oft als Kasten um die Flächen herum gezeichnet. Solch eine Darstellung nennt man auch *Venn-Diagramm*.

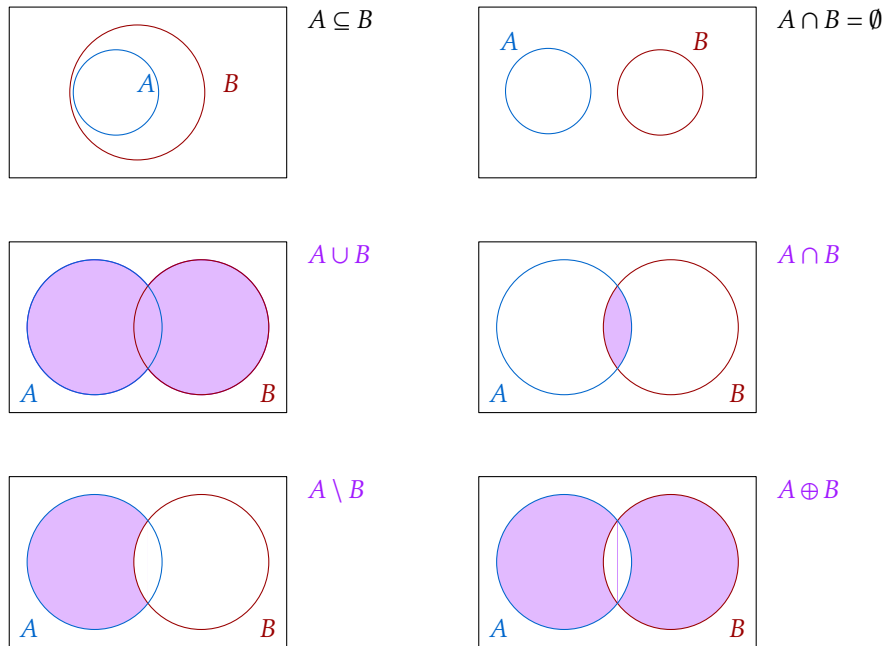
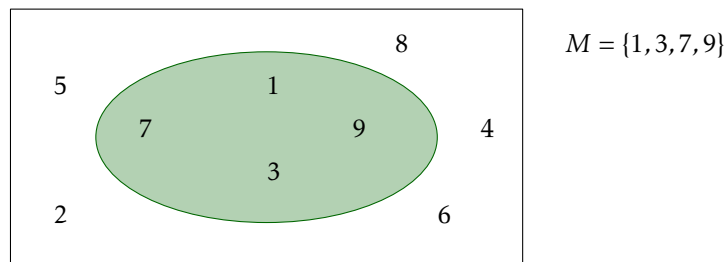


Abbildung 3.1: Visualisierung der Begriffe für Menge

Beispiel.



Selbsttest: Je zwei der folgenden sechs Mengen sind gleich. Welche Paare sind das?

$$\{0, 3, 6, 9, \dots\}$$

$$\{5, 10, 15, 30\}$$

$$\{5, 10, 15, \dots\}$$

$$\{x \in \mathbb{N} \mid x > 0 \wedge 5 \text{ teilt } x\}$$

$$\{x \in \mathbb{N} \mid 3 \text{ teilt } x\}$$

$$\{5, 15, 30, 10, 5\}$$

3.1.2 Eigenschaften von Mengen

Wir definieren nun einige Begriffe, die im Zusammenhang mit Mengen wichtig sind, siehe Abbildung 3.1 für eine Visualisierung.

Definition 3.2.

Teilmenge Eine Menge A ist **Teilmenge** ($A \subseteq B$) von B , wenn alle Elemente aus A auch in B enthalten sind. B heißt dann auch **Obermenge** von A .

Leere Menge Die Menge, die kein Element enthält, wird leere Menge genannt und häufig mit \emptyset bezeichnet.

Vereinigung Die **Vereinigung** ($A \cup B$) von zwei Mengen A und B besteht aus allen Elementen, die in A oder in B enthalten sind.

$$A \cup B = \{x \mid x \in A \vee x \in B\}$$

Schnittmenge Die **Schnittmenge** ($A \cap B$) von zwei Mengen A und B besteht aus allen Elementen, die in A und in B enthalten sind.

$$A \cap B = \{x \mid x \in A \wedge x \in B\}$$

Disjunkt Zwei Menge A und B sind **disjunkt**, wenn sie keine gemeinsamen Elemente haben ($A \cap B = \emptyset$).

Differenz Die **Differenz** ($A \setminus B$) von zwei Mengen A und B besteht aus allen Elementen, die in A aber nicht in B sind.

$$A \setminus B = \{x \mid x \in A \wedge x \notin B\}$$

Symmetrische Differenz Die **symmetrische Differenz** ($A \oplus B$) von zwei Mengen A und B ist definiert als

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

Komplement Das **Komplement** (\overline{A}) einer Menge A bezüglich eines Universums U ist definiert als

$$\overline{A} = U \setminus A$$

Anmerkungen zu den Definitionen:

- Die leere Menge \emptyset ist etwas anderes, als die Menge, die die leere Menge enthält ($\{\emptyset\}$).
- $\{x \mid P(x)\} = \{x \mid Q(x)\}$ genau dann, wenn $P(x) \equiv Q(x)$.
- Wenn $A = \{x \in U \mid P(x)\}$, dann ist $\overline{A} = \{x \in U \mid \neg P(x)\}$.
- Es gilt $\emptyset = \{x \in U \mid \text{False}\}$ und $U = \{x \in U \mid \text{True}\}$.

Selbsttest: Betrachten Sie die Mengen $A = \{1, 3, 5, 7, 9\}$ und $\{1, 2, 5, 6, 9, 10\}$. Bestimmen Sie

1. $A \cup B$

2. $A \cap B$

3. $A \setminus B$

Sind die Mengen disjunkt?

Mit diesen Definitionen können wir nun einige komplexere Mengenidentitäten zeigen.

Satz 3.1. Es gelten die folgenden Identitäten für Mengen:

$$A \cup B = B \cup A \quad \text{Kommutativität}$$

$$A \cap B = B \cap A$$

$$A \cup (B \cup C) = (A \cup B) \cup C \quad \text{Assoziativität}$$

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

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{Distributivität}$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cup A = A \quad \text{Idempotenz}$$

$$A \cap A = A$$

$$A \cup \emptyset = A \quad \text{Identität}$$

$$A \cap U = A$$

$$A \cup U = U \quad \text{Dominanz}$$

$$A \cap \emptyset = \emptyset$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{DeMorgansche Regel}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$A \cup \overline{A} = U \quad \text{Komplementierung}$$

$$A \cap \overline{A} = \emptyset$$

$$\overline{(\overline{A})} = A \quad \text{Doppelte Negation}$$

$$A \setminus B = A \cap \overline{B} \quad \text{Differenz}$$

Beweis. Die Aussagen können durch Rückführung auf die Booleschen Gesetze gezeigt werden. Wir zeigen die Aussage beispielhaft für die Kommutativität von \cup und die erste Version der Distributivität. Die anderen Beweise folgen dann analog.

Kommutativität von \cup

$$\begin{aligned} A \cup B &= \{x \mid x \in A \vee x \in B\} && | \text{ Definition } \cup \\ &= \{x \mid x \in B \vee x \in A\} && | \text{ Kommutativität für } \vee \\ &= B \cup A && | \text{ Definition } \cup \end{aligned}$$

Distributivität

$$\begin{aligned} A \cup (B \cap C) &= \{x \mid x \in A \vee x \in (B \cap C)\} && | \text{ Definition } \cup \\ &= \{x \mid x \in A \vee (x \in B \wedge x \in C)\} && \text{ Definition } \cap \\ &= \{x \mid (x \in A \vee x \in B) \wedge (x \in A \vee x \in C)\} && \text{ Distributivität von } \vee \text{ und } \wedge \\ &= \{x \mid (x \in A \vee x \in B) \wedge (x \in A \vee x \in C)\} && \text{ Definition } \cap \\ &= \{x \mid x \in A \vee x \in B\} \cap \{x \mid x \in A \vee x \in C\} && \text{ Definition } \cup \\ &= (A \cup B) \cap (A \cup C) \end{aligned}$$

□

Bemerkung: Wegen der Assoziativität können beim Schnitt/der Vereinigung von mehreren Mengen die Klammern weggelassen werden. Analog zur Schreibweise von Summen mit \sum gibt es auch hier eine kompakte Schreibweise.

$$\begin{aligned} A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n &= \bigcup_{i=1}^n A_i \\ A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n &= \bigcap_{i=1}^n A_i \end{aligned}$$

Definition 3.3 (Mengenfamilie). Sei I eine beliebige Menge. Wenn für jedes $i \in I$ eine Menge A_i definiert ist, dann ist diese Menge von Mengen eine **Mengenfamilie**. Diese wird mit $(A_i)_{i \in I}$ bezeichnet. Die Menge I wird auch oft **Indexmenge** genannt.

Definition 3.4. Schnitt und Vereinigung einer Mengenfamilie Es gilt:

$$\begin{aligned} \bigcup_{i \in I} A_i &= \{x \mid \text{es gibt ein } i \in I, \text{ sodass } x \in A_i\} \\ \bigcap_{i \in I} A_i &= \{x \mid \text{für alle } i \in I \text{ gilt } x \in A_i\} \end{aligned}$$

Beispiel. Mit Indexmenge $I = \{1, 2, 3\}$ ist $I_1 = \{1, 3, 6, 10\}, I_2 = \{3, 9, 15, 21\}, I_3 = \{2, 4, 6, 8, 10, 3, 5, 7, 9\}$ eine Mengenfamilie.

Es gilt

$$\bigcup_{i \in I} A_i = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 21\}$$
$$\bigcap_{i \in I} A_i = \{3\}$$



Die Definition der Mengenfamilie brauchen wir nun für die nächste Definition. Diese beschreibt eine *Partition* einer Grundmenge A . Eine Partition kann als Aufteilung der Menge A in eine Mengenfamilie betrachtet werden. Dabei sind die einzelnen Elemente der Mengenfamilie nicht leer und jedes Element aus A ist in genau einer Menge der Mengenfamilie enthalten. Formal kann dieses Konzept wie folgt definiert werden.

Definition 3.5 (Partition). Eine Familie $\{A_i \mid i \in I\}$ von nichtleeren Mengen wird **Partition** einer Menge A genannt, wenn

- (i) $A = \bigcup_{i \in I} A_i$ (Jedes Element aus A ist in mindestens einer Menge der Mengenfamilie enthalten)
- (ii) Für beliebige voneinander verschiedene $i, j \in I$ gilt: $A_i \cap A_j = \emptyset$. (Jedes Element aus A ist in maximal einer Menge der Mengenfamilie enthalten.)

Beispiel. Sei $A = \{0, 1, 2, 3, 4, 5\}$ die Grundmenge.

1. $A_0 = \{0, 3\}, A_1 = \{1, 4\}, A_2 = \{2, 5\}$ ist eine Partition. Um dies zu sehen, schauen wir beide Eigenschaften an und zeigen, dass diese gelten.
 - (i) $A_0 \cup A_1 \cup A_2 = \{0, 1, 2, 3, 4, 5\}$.
 - (ii) Werden alle Paare von Mengen betrachtet, ist zu sehen, dass jedes Paar disjunkt ist und damit ist die Eigenschaft erfüllt.
2. $A_0 = \{0, 1, 2, 3, 4, 5\}$ ist eine Partition mit nur einer Menge in der Mengenfamilie.
 - (i) Die erste Eigenschaft gilt direkt nach Definition.
 - (ii) Da es nur eine Menge gibt, können keine Paare gebildet werden und die Aussage gilt.
3. $A_0 = \{0, 1, 2\}, A_1 = \{2, 3, 4\}, A_2 = \{3, 4, 5\}$ ist *keine* Partition von A
 - (i) Die erste Eigenschaft gilt, da die Vereinigung aller Mengen genau $\{0, 1, 2, 3, 4, 5\}$ ist.
 - (ii) Die zweite Eigenschaft gilt nicht, da $A_1 \cap A_2 = \{3, 4\} \neq \emptyset$ ist.
4. $A_0 = \{0, 1\}, A_1 = \{3, 4\}, A_2 = \{5\}$ ist *keine* Partition von A

- (i) Die erste Eigenschaft gilt nicht, da die Vereinigung aller Mengen nur $\bigcup_{i \in I} A_i = \{0, 1, 3, 4, 5\} \neq A$ ist.
- (ii) Die zweite Eigenschaft gilt, alle Paare von Mengen sind disjunkt. ◀

Selbsttest:

1. Geben Sie eine Partition der Menge \mathbb{N} an.
2. Finden Sie Mengenfamilien auf \mathbb{N} , die keine Partitionen sind. Welche Eigenschaften werden jeweils verletzt?

Nun betrachten wir noch eine letzte Definition im Bezug auf Mengen.

Definition 3.6. Ist A eine Menge, dann wird die Menge aller Teilmengen von A die **Potenzmenge** von A genannt. Die Potenzmenge wird mit $\mathcal{P}(A)$ bezeichnet.

Beispiel. Die Potenzmenge der Menge $A = \{1, 2, 3\}$ ist

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}. \quad \blacktriangleleft$$

Selbsttest: Was ist $\mathcal{P}(\emptyset)$?

Lemma 3.2. Wenn $|A| = n$ für eine endliche Menge gilt, dann ist $|\mathcal{P}(A)| = 2^n$.

Beweis. Wir zeigen die Aussage mit vollständiger Induktion über $|A|$.

I.A. ($n = 0$) Wenn $n = 0$ ist, dann ist $A = \emptyset$ und $\mathcal{P}(A) = \{\emptyset\}$. Es gilt also $|\mathcal{P}(A)| = 1 = 2^0$ und die Aussage gilt für $n = 0$.

I.V. Für ein beliebiges $n \in \mathbb{N}$ hat die Potenzmenge einer n -elementigen Menge 2^n Elemente.

I.S. ($n \rightarrow n + 1$) Sei $A = \{a_1, \dots, a_{n+1}\}$ eine Menge mit $n + 1$ Elementen. Dann ist

$$\mathcal{P}(A) = \mathcal{P}(\{a_1, \dots, a_n\}) \cup \{X \cup \{a_{n+1}\} \mid X \in \mathcal{P}(\{a_1, \dots, a_n\})\}$$

Die Potenzmenge von A besteht also zunächst aus Potenzmenge von $\{a_1, \dots, a_n\}$. Nun fehlen jedoch noch alle Teilmengen, die a_{n+1} enthalten. Diese ergeben sich, indem in jede Teilmenge aus $\mathcal{P}(\{a_1, \dots, a_n\})$ das Element a_{n+1} einmal hinzugefügt wird.

Nach I.V. gilt $|\mathcal{P}(\{a_1, \dots, a_n\})| = 2^n$. Also ist $|\mathcal{P}(A)| = 2^n + 2^n = 2^{n+1}$ und die Aussage ist gezeigt. \square

3.2 Relationen

Voraussetzung:

- Abschnitt 3.1
- Kennen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ und elementare Rechenoperationen auf diesen Mengen.

Lernziele

Die Studierenden ...

- ... verwenden Operationen auf Relationen
- ... untersuchen Relationen auf ihre Eigenschaften
- ... beweisen Aussagen über bestimmte Relationen

In diesem Abschnitt beschäftigen wir uns mit einer besonderen Art von Mengen, den sogenannten *Relationen*. Bevor wir Relationen formal definieren können, brauchen wir zunächst noch einige weitere Definitionen.

Definition 3.7 (Geordnetes Paar). Ein **geordnetes Paar** ist ein (den Objekten a und b zugeordnetes) Konstrukt mit der folgenden Eigenschaft: $(a, b) = (a', b')$, genau dann, wenn $a = a'$ und $b = b'$.

Ein geordnetes Paar wird auch oft als **Tupel** oder **2-Tupel** bezeichnet. Die Definition kann auf mehr als zwei Objekte erweitert werden und wird dann **k-Tupel** genannt.

Sind nun zwei oder mehr Mengen gegeben, hat die Menge aller möglichen Tupel bei denen das erste Element aus der ersten Menge, das zweite aus der zweiten Menge usw. kommt einen speziellen Namen:

Definition 3.8 (Kartesisches Produkt). Das **kartesische Produkt** $A \times B$ von zwei Mengen A und B ist definiert als die Menge aller geordneten Paare (a, b) mit $a \in A$ und $b \in B$.

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

Verallgemeinert ist das kartesische Produkt $A_1 \times A_2 \times \dots \times A_k$ von k Mengen A_1, \dots, A_k definiert als:

$$A_1 \times A_2 \times \dots \times A_k = \{(a_1, \dots, a_k) \mid \forall i : a_i \in A_i\}$$

Sind alle Mengen A_i gleich, schreibt man auch kurz A^k statt $A \times \dots \times A$.

Beispiel. Für die Menge $A = \{1, 2, 3\}$ und $B = \{a, b\}$ ist

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$

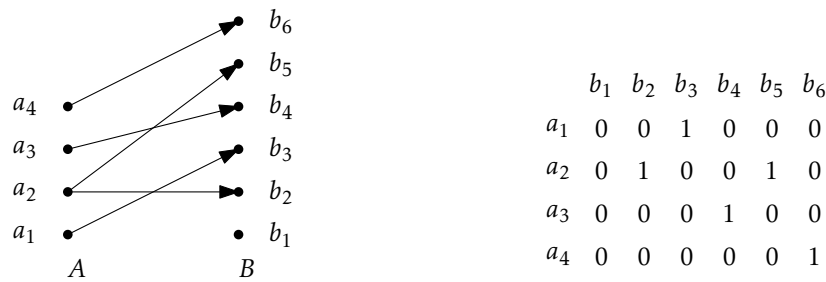


Abbildung 3.2: Grafische Darstellung einer Relation

Selbsttest: Was ist $\{1, 2\} \times \{x, y, z\}$?

Das kartesische Produkt von Mengen ist auch wieder eine Menge. Also können wir die bekannten Mengenoperationen und Begriffe darauf verwenden. Eine Teilmenge vom kartesischen Produkt von zwei Mengen hat einen besonderen Namen.

Definition 3.9 (Binäre Relation). Eine Teilmenge $R \subseteq A \times B$ wird **binäre Relation** zwischen A und B genannt. Statt $(a, b) \in R$ wird oft auch aRb geschrieben. Eine Teilmenge $R \subseteq A \times A$ des kartesischen Produkts von A mit sich selbst, wird (binäre) Relation in A genannt.

Die Relation $\emptyset \subseteq A \times B$ ist die **leere Relation**, die Relation $A \times B \subseteq A \times B$ ist die **Allrelation**. Die Relation $Id_A = \{(a, a) \mid a \in A\}$ wird als **identische Relation** bezeichnet.

Relationen können auf verschiedene Arten dargestellt werden. Es kann zum Beispiel eine Tabelle angegeben werden, in der alle Tupel aufgezählt werden. Eine weitere geläufige Darstellung ist eine grafische Darstellung, bei der die Elemente als Punkte aufgezeichnet werden und ein Pfeil von a nach b gezeichnet wird, wenn $(a, b) \in R$, siehe Abbildung 3.2.

3.2.1 Operationen auf Relationen

Da Relationen auch nur wieder Mengen sind, sind alle Mengenoperationen auch definiert.

Definition 3.10 (Operationen auf Relationen).

Mengenoperationen Seien R, S Relationen zwischen A und B . Dann sind $R \cup S$, $R \cap S$ und $\overline{R} = (A \times B) \setminus R$ Relationen zwischen A und B .

Inverse Relation Die Relation $R^{-1} = \{(b, a) \in B \times A \mid (a, b) \in R\}$ ist die **inverse Relation** zu $R \subseteq A \times B$.

Verkettung von Relationen Die **Verkettung** $R \circ S$ von zwei Relation $R \subseteq A \times B$ und $S \subseteq B \times C^a$ ist:

$$R \circ S = \{(a, c) \mid \exists b \in B : (a, b) \in R \wedge (b, c) \in S\}$$

^aAchtung, S ist hier anders als im ersten Punkt definiert

Beispiel. 1. Die folgenden Relationen sind Relationen auf \mathbb{N} . Es stehen jeweils zwei Zahlen in Relation zueinander, wenn die erste Zahl kleiner gleich (gleich, echt kleiner, größer gleich) als die zweite Zahl ist.

$$R_{\leq} = \{(a, b) \in \mathbb{N} \times \mathbb{N} \mid a \leq b\}$$

$$R_{=} = \{(a, b) \in \mathbb{N} \times \mathbb{N} \mid a = b\}$$

$$R_{<} = \{(a, b) \in \mathbb{N} \times \mathbb{N} \mid a < b\}$$

$$R_{\geq} = \{(a, b) \in \mathbb{N} \times \mathbb{N} \mid a \geq b\}$$

Dann gelten die folgenden Aussagen:

$$R_{<} \cup R_{=} = R_{\leq}$$

$$\overline{R_{<}} = R_{\geq}$$

$$R_{\leq}^{-1} = R_{\geq}$$

$$R_{=} = Id_{\mathbb{N}}$$

2. Ein Beispiel für die Verkettung ist das Folgende. Sei A die Menge aller Studierenden, B die Menge aller Tutor:innen und C die Menge aller Dozent:innen. Dann kann eine Relation $R \subseteq A \times B$ definiert werden als aRb , wenn Studi a bei Tutor:in b im Tutorium ist. Eine Relation $S \subseteq B \times C$ zwischen Tutor:innen und Dozent:innen kann definiert werden, als bSc , wenn b Tutor:in in einer Veranstaltung von c ist. Wir nehmen der Einfachheit halber an, dass Tutor:innen jeweils nur für eine Veranstaltung eingesetzt werden. Dann beinhaltet die Relation $R \circ S$ dann alle Studierenden a , welche ein Tutorium in einer Veranstaltung von Dozent:in c besuchen.

Etwas genauer: Nach der Definition der Verkettung steht $a \in A$ in Relation zu $c \in C$ bezüglich $R \circ S$, wenn es ein $b \in B$ gibt, mit $(a, b) \in R$ und $(b, c) \in S$. In unserem Beispiel steht dann also ein Studi a in Relation zu einer Dozierenden Person c , wenn es einen Tutor/eine Tutorin b gibt, sodass a ein Tutorium von b besucht, und b das Tutorium für eine Veranstaltung bei c leitet.

3. Ein weiteres Beispiel ist die Elternteilrelation über der Menge aller Menschen. Sei M die Menge aller Menschen und sei aRb , wenn a Elternteil von b ist.² Es gilt also:

- $aR^{-1}b$, wenn a Kind von b ist.

²Wir verwenden jetzt die aRb Schreibweise, statt $(a, b) \in R$.

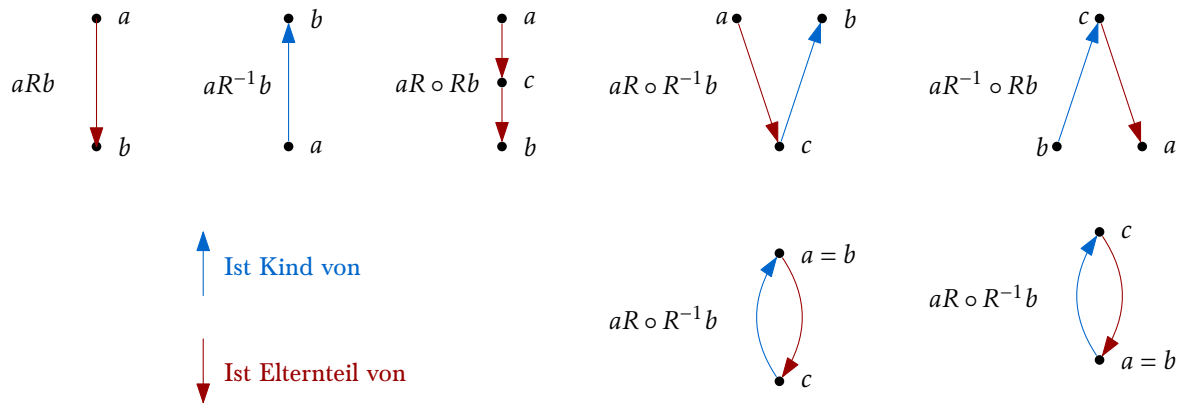


Abbildung 3.3: Visualisierung des Beispiels zur Elternrelation

- $aR \circ Rb$, wenn a Großelternteil von b ist.
- $aR \circ R^{-1}b$, wenn a und b ein gemeinsames Kind haben, oder wenn $a = b$ und a hat ein Kind.
- $aR^{-1} \circ Rb$, falls $a = b$ oder a und b sind (Halb)-geschwister.

Dieses Beispiel ist in Abbildung 3.3 visualisiert.

Selbsttest: Wählen Sie eine feste Relation $R \subseteq A \times B$ und $S \subseteq B \times C$ mit $A = \{1, 2, 3\}$, $B = \{7, 8, 9\}$ und $C = \{a, b\}$. Bestimmen Sie $R \circ S$, R^{-1} , S^{-1} , \overline{R} und \overline{S} .

Für binäre Relationen in einer Menge gibt es einige Eigenschaften, welche die Relation haben kann:

Definition 3.11 (Eigenschaften von Relationen in einer Menge). Sei $R \subseteq A \times A$ eine Relation.

Reflexivität R ist **reflexiv**, falls für jedes $a \in A$ gilt, dass aRa .

Für jedes Element a der Grundmenge ist das Tupel (a, a) in der Relation enthalten.

Symmetrie R ist **symmetrisch**, falls aus aRb folgt, dass bRa .

Wenn ein Tupel (a, b) in der Relation enthalten ist, dann ist auch (b, a) in der Relation enthalten.

Transitiv R ist **transitiv**, falls aus aRb und bRc folgt, dass aRc .

Wenn Tupel (a, b) und (b, c) in der Relation enthalten sind, dann ist auch (a, c) enthalten.

Antisymmetrie R ist **antisymmetrisch**, falls aus aRb und bRa folgt, dass $a = b$.
Wenn ein Tupel (a, b) mit $a \neq b$ in der Relation ist, dann ist das Tupel (b, a) nicht in der Relation.

Asymmetrie R ist **asymmetrisch**, falls aus aRb folgt, dass $(b, a) \notin R$.
Für kein Tupel (a, b) ist (b, a) auch in der Relation enthalten. Das schließt das Tupel (a, a) ein.

Beispiel. 1. Wir betrachten die Grundmenge $A = \{1, 2, 3, 4\}$.

Reflexivität Die Relation $R_1 = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (4, 1)\}$ ist reflexiv, da jedes Element der Grundmenge mit sich selbst in Relation steht.

Die Relation $R_2 = \{(1, 1), (2, 2)\}$ ist *nicht* reflexiv, da die Tupel $(3, 3)$ und $(4, 4)$ fehlen.

Symmetrie Die Relation $R_3 = \{(1, 2), (2, 1), (2, 2)\}$ ist symmetrisch, da für jedes Tupel in der Relation das umgedrehte Tupel auch enthalten ist.

Die Relation $R_4 = \{(1, 2), (2, 2), (2, 3), (3, 2)\}$ ist *nicht* symmetrisch, da das Tupel $(2, 1)$ in der Relation fehlt.

Transitivität Die Relation $R_5 = \{(1, 2), (2, 3), (1, 3), (2, 1), (1, 1), (2, 2)\}$ ist transitiv. Das Tupel $(1, 3)$ muss in der Relation enthalten sein, da $(1, 2)$ und $(2, 1)$ enthalten sind. Genauso muss $(2, 2)$ enthalten sein, da $(1, 2)$ und $(2, 1)$ enthalten sind.

Die Relation $R_6 = \{(1, 2), (2, 3)\}$ ist nicht transitiv, da das Tupel $(1, 3)$ fehlt.

Antisymmetrie Die Relation $R_7 = \{(1, 2), (2, 3), (2, 2), (4, 1)\}$ ist antisymmetrisch, da außer für $(2, 2)$ keins der „umgedrehten“ Tupel enthalten ist.

Die Relation $R_8 = \{(1, 2), (2, 1), (2, 3), (2, 2)\}$ ist nicht antisymmetrisch³, da die Tupel $(1, 2)$ und $(2, 1)$ enthalten sind, aber $1 \neq 2$ gilt.

Asymmetrie Die Relation $R_9 = \{(1, 2), (4, 1), (2, 3)\}$ ist asymmetrisch, da es kein Tupel gibt, für das das „umgedrehte“ Tupel auch in der Relation enthalten ist.

Die Relation $R_{10} = \{(1, 2), (4, 1), (2, 3), (2, 2)\}$ ist nicht asymmetrisch⁴, da $(2, 2)$ enthalten ist. Damit ist $(2, 2)$ und das (identische) umgedrehte Tupel $(2, 2)$ in der Relation enthalten.

Die Relation $R_{11} = \{(1, 2), (2, 1), (4, 1), (2, 3)\}$ ist nicht asymmetrisch⁵, da $(2, 1)$ und $(1, 2)$ beide enthalten sind.

2. Die Relation $R_<$ aus dem vorherigen Beispiel ist transitiv, antisymmetrisch und asymmetrisch, aber *nicht* reflexiv und symmetrisch. Wir schauen uns die Argumente für jede der Eigenschaften einmal im Detail an:

(i) $R_<$ ist nicht reflexiv: Da $3 < 3$ eine falsche Aussage ist, ist $(3, 3) \notin R$ und damit ist $R_<$ nicht reflexiv.

³aber auch nicht symmetrisch

⁴aber auch nicht symmetrisch

⁵und auch nicht symmetrisch oder antisymmetrisch

- (ii) $R_<$ ist nicht symmetrisch: Da zwar $3 < 4$ aber nicht $4 < 3$ gilt, ist $3R_<4$ aber $(4, 3) \notin R_<$. Damit ist $R_<$ nicht symmetrisch. Diese Argumentation kann auf alle Paare von Zahlen erweitert werden.
- (iii) $R_<$ ist transitiv: Es gilt $(a < b) \wedge (b < c) \rightarrow a < c$. Also, wenn $a < b$ und $b < c$ dann ist auch $a < c$. Dies ist die Definition der Transitivität und damit ist $R_<$ transitiv.
- (iv) $R_<$ ist antisymmetrisch. Da es kein Paar gibt, für das aRb und bRa (also $a < b$ und $b < a$) gilt, ist die Aussage direkt wahr.
- (v) $R_<$ ist asymmetrisch: Siehe Argument für Antisymmetrie. ◀

Selbsttest: Untersuchen Sie $R_<$ auf die fünf oben eingeführten Eigenschaften einer Relation. ?

Anmerkungen:

- Jede asymmetrische Relation ist auch antisymmetrisch, aber nicht umgekehrt.
- Die einzigen Relationen die antisymmetrisch und symmetrisch gleichzeitig sein können sind Teilmengen der Identitätsrelation.
- Die leere Relation ist die einzige Relation, die asymmetrisch und symmetrisch gleichzeitig ist.

3.2.2 Äquivalenzrelationen

Binäre Relationen auf einer Grundmenge A können auch so interpretiert werden, dass zwei Elemente aus A miteinander in eine Beziehung gesetzt werden. Eine besondere Art dieser Relationen werden wir in diesem Abschnitt betrachten, die sogenannten Äquivalenzrelationen. Diese haben die Besonderheit, dass sie die Grundmenge A in Klassen unterteilen. In jeder Klasse stehen alle Elemente in Relation zueinander, zwischen den Klassen gibt es keine Relationen.

Definition 3.12 (Äquivalenzrelation). Eine Relation in einer Menge A wird **Äquivalenzrelation** genannt, wenn sie reflexiv, symmetrisch und transitiv ist.

Beispiel. 1. Die Identitätsrelation $Id_A = \{(a, a) \mid a \in A\}$ ist eine Äquivalenzrelation.

Reflexivität Id_A ist reflexiv: $(a, a) \in Id_A$ nach Definition.

Symmetrie Id_A ist symmetrisch: Alle Elemente von Id_A haben die Form (a, a) . Also folgt aus aRb direkt $a = b$ und damit auch bRa .

Transitivität Id_A ist transitiv: Aus aRb folgt $a = b$. Aus bRc folgt $b = c$. Also gilt $a = b = c$. Es ist also zu zeigen, dass aRc gilt, da aber $a = c$ ist dies das Gleiche, wie aRa , was per Definition gilt.

2. Seien $a, b \in \mathbb{Z}$. Die Relation, die a und b in Relation setzt, wenn sie beim Teilen durch eine Zahl $m \in \mathbb{Z}$ den gleich Rest haben ist eine Äquivalenzrelation. Dies wird oft auch als $a \equiv b \pmod{m}$ geschrieben⁶. Eine äquivalente Bedingung ist auch $m \mid (a - b)$, also m teilt $(a - b)$. Wir definieren $R_m = \{(a, b) \mid m \mid (a - b)\}$. Für jedes feste m ist R_m eine Äquivalenzrelation:

Reflexivität R_m ist reflexiv. Wir müssen zeigen $m \mid (a - a)$. Das ist äquivalent zu $m \mid 0$. Da die 0 von jeder Zahl geteilt wird, gilt die Aussage.

Symmetrie R_m ist symmetrisch. Wir müssen also zeigen $m \mid (a - b) \rightarrow m \mid (b - a)$. Es gilt, m teilt eine Zahl x genau dann, wenn m auch $-x$ teilt. Zudem gilt $(a - b) = -(b - a)$. Also

$$\begin{aligned} m \mid (a - b) \\ \Leftrightarrow m \mid -(b - a) \\ \Leftrightarrow m \mid (b - a) \end{aligned}$$

und wir sind fertig.

Transitivität R_m ist transitiv. Wir müssen zeigen $m \mid (a - b) \wedge m \mid (b - c) \rightarrow m \mid (a - c)$. Als Hilfseigenschaft können wir benutzen, dass $m \mid x \wedge m \mid y \implies m \mid (x + y)$ gilt.⁷

Mit der Hilfseigenschaft und dem Fakt, dass $a - c = (a - b) + (b - c)$ gilt, sind wir fertig.



Definition 3.13. Ist $R \subseteq A \times A$ eine Äquivalenzrelation und ist $a \in A$, dann nennt man $[a]_R = \{x \in A \mid xRa\}$ die **Äquivalenzklasse** von a . Das Element a wird dann der **Repräsentant** der Klasse genannt.

Beispiel. Für die Identitätsrelation ist jedes Element seine eigene Äquivalenzklasse. Für die Teilbarkeitsrelation mit $m = 5$ gilt zum Beispiel $[3]_{R_m} = \{3, 8, 13, \dots\}$.



Lemma 3.3. Sei R eine Äquivalenzrelation. Dann sind je zwei Äquivalenzklassen entweder gleich oder disjunkt.

Beweis. Seien $a, b \in A$. Wir führen einen Beweis mit Fallunterscheidung und unterscheiden die Fälle, ob die Äquivalenzklassen von a und b disjunkt sind, oder nicht.

Fall 1: $[a]_R \cap [b]_R = \emptyset$ in diesem Fall ist nichts zu tun, da die zweite Bedingung direkt erfüllt ist.

⁶Gelesen als „ a ist kongruent zu b modulo m “

⁷Dies kann einfach mit einem direkten Beweis und der Definition der Teilbarkeit gezeigt werden.

Fall 2: $[a]_R \cap [b]_R \neq \emptyset$. Wir betrachten nun ein Element $c \in [a]_R \cap [b]_R$. Dann ist cRa und cRb nach Definition der Äquivalenzklassen. Wegen der Symmetrie gilt dann auch bRc . Es gilt also bRc und cRa , wegen der Transitivität gilt dann bRa und nach Symmetrie auch aRb .

Nun zeigen wir $[a]_R = [b]_R$ in zwei Schritten. Zuerst wird $[a]_R \subseteq [b]_R$ gezeigt, danach $[b]_R \subseteq [a]_R$.

Schritt 1: ($[a]_R \subseteq [b]_R$) Sei $d \in [a]_R$ ein Element aus der Äquivalenzklasse von a . Wir wollen nun zeigen, dass auch $d \in [b]_R$ gilt. Da $d \in [a]_R$ gilt dRa und mit Symmetrie auch aRd . Die Überlegung oben hatte bRa ergeben. Wegen der Transitivität folgt aus bRa und aRd auch bRd . Mit Symmetrie ergibt sich dann dRb und daher gilt $d \in [b]_R$ nach der Definition der Äquivalenzklassen.

Schritt 2: ($[a]_R \subseteq [b]_R$) Analog mit der Rolle von a und b vertauscht. \square

Nun können wir die Hauptaussage zum Thema Äquivalenzrelationen zeigen:

Satz 3.4. Ist $R \subseteq A \times A$ eine Äquivalenzrelation, dann bildet die Menge der Äquivalenzklassen eine Partition von A .

Umgekehrt kann aus jeder Partition $\{A_i \mid i \in I\}$ eine Äquivalenzrelation mit aRb genau dann, wenn $a \in A_i \wedge b \in A_i$ gebildet werden.

Beweis. Wir zeigen nur den ersten Teil, der zweite ist eine schöne Übung.

Wir müssen nun zeigen, dass die Menge $\{[a]_R \mid a \in A\}$ eine Partition ist. Zunächst ist leicht zu sehen, dass wegen der Reflexivität $a \in [a]_R$ gilt, und damit keine der Äquivalenzklassen leer ist.

Nun zeigen wir, dass die Vereinigung aller Äquivalenzklassen genau die Menge A ergeben. Wegen $\{a\} \subseteq [a]_R$ gilt

$$A = \bigcup_{a \in A} \{a\} \subseteq \bigcup_{a \in A} [a]_R,$$

also sind schon einmal alle Elemente aus A in der Vereinigung enthalten. Auf der anderen Seite sind in $[a]_R$ per Definition nur Elemente aus A enthalten, also ist $\bigcup_{a \in A} [a]_R = A$.

Das jeweils zwei Mengen der Familie disjunkt sind, folgt aus Lemma 3.3. \square

Definition 3.14. Sei $R \subseteq A \times A$ eine Äquivalenzrelation. Eine Teilmenge $T \subseteq A$ wird **Repräsentantensystem** genannt, wenn sie aus jeder Äquivalenzklasse genau ein Element enthält.

Beispiel.

1. Für die Identitätsrelation ist jedes Element seine eigene Äquivalenzklasse. Daher ist die Menge A das Repräsentantensystem.

2. Für die Teilbarkeitsrelation mit $m = 5$ gibt es 5 Äquivalenzklassen, die jeweils die Elemente beinhalten, welche den gleichen Rest beim Teilen durch 5 haben. Ein Repräsentantensystem ist $\{0, 1, 2, 3, 4\}$. Ein anderes mögliches Repräsentantensystem ist $\{0, 6, -3, 4\}$.
3. Ein weiteres Beispiel ist die folgende Relation über der Menge $A = \{1, 2, 3, 4, 5\}$

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

Es kann explizit geprüft werden, dass diese Relation eine Äquivalenzrelation ist. Die Äquivalenzklassen sehen wie folgt aus:

$$[1]_R = \{1, 3, 5\}$$

$$[2]_R = \{2, 4\}$$

$$[3]_R = \{1, 3, 5\}$$

$$[4]_R = \{2, 4\}$$

$$[5]_R = \{1, 3, 5\}$$

Damit gibt es zwei verschiedene Äquivalenzklassen. Ein mögliches Repräsentantensystem ist $\{1, 2\}$, aber auch $\{1, 4\}$ ist ein Repräsentantensystem.

Selbsttest: Die Relation, welche zwei Wörter der deutschen Sprache in Relation miteinander setzt, genau dann, wenn sie mit dem gleichen Buchstaben beginnen ist eine Äquivalenzrelation.

1. Warum ist diese Relation eine Äquivalenzrelation?
2. Wie sehen die Äquivalenzklassen aus?
3. Wie würde ein Repräsentantensystem aussehen?

3.2.3 Halbordnungsrelationen und totale Ordnung

Ähnlich zu Äquivalenzrelationen, gibt es für Relationen die reflexiv, transitiv und antisymmetrisch sind auch einen Namen.

Definition 3.15. Halbordnungsrelation Eine **Halbordnungsrelation** ist eine Relation R in A , die reflexiv, transitiv und antisymmetrisch ist.

Das Paar (A, R) nennt man auch eine **halb/partiell geordnete Menge** oder kurz ein **Poset** (partially ordered set).

Statt (A, R) wird oft auch (A, \leq) geschrieben. Für aRb wird auch $a \leq b$ und für aRb mit $a \neq b$ wird $a < b$ geschrieben.

Beispiel.

1. Für jede Menge M ist $(\mathcal{P}(M), \subseteq)$ eine halbgeordnete Menge.


Reflexivität Für jede Teilmenge M' von M gilt $M' \subseteq M$.

Transitivität Seien M_1, M_2, M_3 Teilmengen von M . Dann ist zu zeigen $M_1 \subseteq M_2 \wedge M_2 \subseteq M_3 \rightarrow M_1 \subseteq M_3$.

$$\begin{aligned} M_1 \subseteq M_2 \wedge M_2 \subseteq M_3 &\rightarrow M_1 \subseteq M_3 \\ &\equiv (m \in M_1 \rightarrow m \in M_2) \wedge (m \in M_2 \rightarrow m \in M_3) \rightarrow (m \in M_1 \rightarrow m \in M_3) \end{aligned}$$

Die letzte Zeile ist eine wahre Aussage wegen der Transitivität der Implikation.

Antisymmetrie Seien M_1 und M_2 Teilmengen von M . Wieder mit Rückführung auf die Definition von \subseteq gilt direkt $M_1 \subseteq M_2 \wedge M_2 \subseteq M_1 \rightarrow M_1 = M_2$.

2. (\mathbb{R}, \leq) ist halbgeordnete Menge. (Ohne Beweis)
3. $R_1 \subseteq \mathbb{N} \times \mathbb{N}$ wobei a in Relation zu b steht, wenn $a \mid b$ gilt ist eine Halbordnungsrelation. Also ist $(\mathbb{N}, |)$ eine halbgeordnete Menge. (Ohne Beweis) 

Selbsttest: Zeigen Sie durch explizites Prüfen der Eigenschaften einer Halbordnungsrelation, dass $(\mathbb{N}, |)$ eine partiell geordnete Menge ist.

Verwenden Sie dafür die aus Kapitel 2 bekannte Definition der Teilbarkeit.

Nun betrachten wir einige weitere Definitionen, die im Kontext von Halbordnungsrelationen relevant sind:

Definition 3.16 (Vergleichbar). Zwei Elemente a, b aus einer halbgeordneten Menge (A, \leq) sind **vergleichbar**, wenn $a \leq b$ oder $b \leq a$.^z

^zHierbei ist das \leq als „Stehen in Relation zueinander“ zu lesen. Die Relation kann, muss aber nicht die kleiner gleich Relation auf Zahlen sein.

Wenn $a \leq b$ und $a \neq b$ gilt, schreibt man auch $a < b$.

Definition 3.17 (Totale Ordnung). Eine Halbordnungsrelation \leq in einer Menge A wird **totale (lineare) Ordnungsrelation** genannt, wenn jedes Paar von Elementen vergleichbar ist.

Beispiel.

1. $(\mathcal{P}(M), \subseteq)$ ist keine totale Ordnung für alle Mengen mit $|M| > 1$. Um dies zu zeigen, reicht es ein Paar zu finden, das nicht vergleichbar ist. Für $M = \{a, b\}$ sind die Teilmengen $\{a\}$ und $\{b\}$ nicht vergleichbar, da $\{a\} \not\subseteq \{b\}$ und $\{b\} \not\subseteq \{a\}$ gilt.

$$M_1 = \{a, b, c\} \quad (\mathcal{P}(M_1), \subseteq)$$

$$M_2 = \{2, 5, 7, 10, 15, 20, 25, 30, 35, 100\} \quad (M_2, |)$$

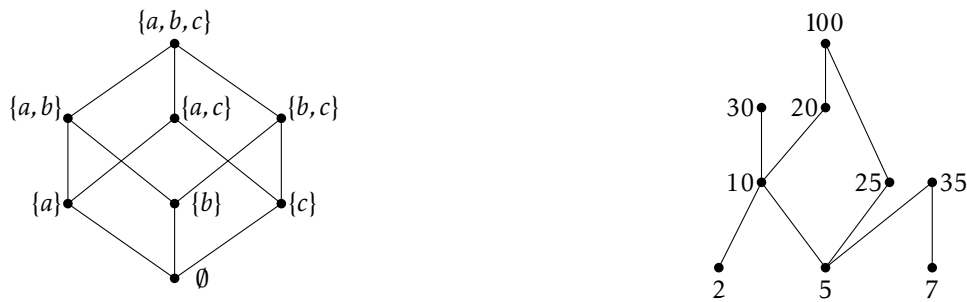


Abbildung 3.4: Beispiele für Hasse-Diagramme

2. (\mathbb{R}, \leq) ist totale Ordnung, da jedes Paar von reellen Zahlen miteinander verglichen werden kann.
3. Die Relation $(\mathbb{N}, |)$, ist keine totale Ordnung, da $3 \nmid 4$ und $4 \nmid 3$ gilt. ◀

Im Folgenden wollen wir uns dem etwas detaillierter dem Ordnungsaspekt der Ordnungsrelationen zuwenden. Dazu definieren wir zunächst den Begriff eines direkten Nachfolgers eines Elements. Intuitiv gesehen ist ein direkter Nachfolger eines Elements a ein Element b , sodass $a < b$, und es keine Elemente zwischen a und b gibt.

Definition 3.18 (Direkter Nachfolger). Gegeben eine Ordnungsrelation (M, \leq) . Ein Element $b \in M$ ist direkter Nachfolger von $a \in M$, wenn

- (i) $a \neq b$
- (ii) $a \leq b$
- (iii) $\forall c \in M : (a \leq c) \wedge (c \leq b) \rightarrow (a = c) \vee (c = b)$

Wir nennen dann auch a den direkten Vorgänger von b .

Selbsttest: Was sind die direkten Nachfolger von 2 in der partiell geordneten Menge $(\{2, 4, 6, 12\}, |)$? ?

Halbordnungsrelationen können mit einem sogenannten **Hasse-Diagramm** dargestellt werden. Hierbei werden die Elemente in Ebenen angeordnet. Ein direkter Nachfolger eines Elements wird dabei in der Ebene über dem Element dargestellt. Elemente die in Relation zueinander stehen und direkte Vorgänger/Nachfolger voneinander sind, werden durch Strecken verbunden, siehe Abbildung 3.4 für Beispiele.

Selbsttest: Zeichnen Sie das Hasse-Diagramm für $(\{2, 4, 6, 12\}, |)$. ?

Eine häufig auftretende Ordnung ist die sogenannte lexikografische Ordnung.

Definition 3.19. Seien $(A_1, \leq_1), \dots, (A_n, \leq_n)$ Halbordnungen bzw. totale Ordnungen. Dann sei $(A_1 \times A_2 \times \dots \times A_n, \leq)$ das kartesische Produkt mit einer Relation \leq , wobei $(a_1, \dots, a_n) \leq (b_1, \dots, b_n)$ gilt, falls $\forall 1 \leq i \leq n : a_i = b_i$ oder $\exists 1 \leq i < n : \forall 1 \leq j \leq i : a_j = b_j \wedge (a_{i+1} <_{i+1} b_{i+1})$

Mithilfe dieser Definition kann die folgende Aussage gezeigt werden. Dies werden wir hier im Skript jedoch nicht tun. Der Beweis ist aber eine einfache Übung.

Lemma 3.5. Für halb (bzw. total) geordnete Mengen $(A_1, \leq_1), \dots, (A_n, \leq_n)$, ist die lexikografische Ordnung wieder eine halb (bzw. total) geordnete Menge.

Wenn eine Halbordnung (M, \leq) zusammen mit einer Teilmenge $A \subseteq M$ gegeben ist, gibt es einige Elemente, die besondere Namen haben. Bei der folgenden Definition ist es wichtig im Blick zu behalten, wann ein Element aus der Grundmenge M und wann ein Element aus der Teilmenge A betrachtet wird.

Definition 3.20 (Schränken in Halbordnungen). Sei (M, \leq) eine Halbordnung und A eine nichtleere Teilmenge von M .

Maximales Element $a \in A$ ist **maximales Element** von A , falls es kein $a' \in A$ mit $a < a'$ gibt.

Minimales Element $a \in A$ ist **minimales Element** von A , falls es kein $a' \in A$ mit $a' < a$ gibt.

Obere Schranke $m \in M$ ist **obere Schranke** von A , falls $a \leq m$ für alle $a \in A$ gilt.

Untere Schranke $m \in M$ ist **untere Schranke** von A , falls $m \leq a$ für alle $a \in A$ gilt.

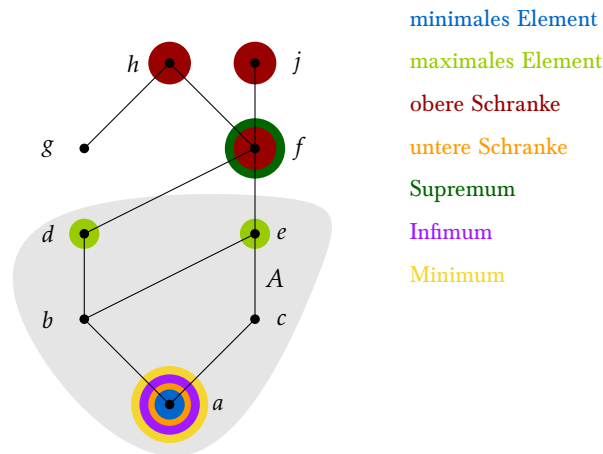
Supremum $m \in M$ ist **Supremum** von A , falls m obere Schranke ist, und $m \leq m'$ für alle oberen Schranken m' von A gilt.

Infimum $m \in M$ ist **Infimum** von A , falls m untere Schranke ist, und $m' \leq m$ für alle unteren Schranken m' von A gilt.

Maximum $a \in A$ ist **Maximum** von A , falls für alle $a' \in A$ gilt: $a' \leq a$. Falls das Maximum existiert, ist es auch das Supremum.

Minimum $a \in A$ ist **Minimum** von A , falls für alle $a' \in A$ gilt: $a \leq a'$. Falls das Minimum existiert, ist es auch das Infimum.

Beispiel. Wir betrachten eine Halbordnung, die durch ein Hasse-Diagramm dargestellt wird.



Die Elemente d, e sind die maximalen Elemente der Menge A , a ist minimales Element von A .

Die Elemente h, j, f sind obere Schranken für A , da für alle Element $a \in A$ gilt, dass $a \leq h, a \leq j$ und $a \leq f$. Das Element g ist keine obere Schranke für A , da g und d nicht vergleichbar sind und damit nicht $m \leq g$ für alle $m \in A$ gilt. Da für f gilt, dass $f \leq h$ und $f \leq j$, ist f das Supremum von A . Es gibt kein Maximum, da keine der oberen Schranken auch in A ist.

Das Element a ist die einzige untere Schranke für A . Damit ist a auch Infimum und da $a \in A$ auch Minimum von a .

Selbsttest: Wählen Sie verschiedene Teilmengen von M_2 aus der partiell geordneten Menge, die rechts in Abbildung 3.4 abgebildet sind. Bestimmen Sie jeweils alle Schranken, soweit vorhanden.

Im Folgenden wollen wir nun zeigen, dass jede endliche Halbordnung mindestens ein minimales und mindestens ein maximales Element hat. Für diesen Beweis brauchen wir noch die folgende Definition:

Definition 3.21 (Kette, Antikette). Eine **Kette** in einer Halbordnung (A, \leq) ist eine Teilmenge von Elementen aus A , die bezüglich \leq eine totale Ordnung bilden. Eine **Antikette** ist eine Menge von paarweisen nicht vergleichbaren Elementen.

Beispiel. Im Beispiel von oben ist zum Beispiel $\{a, c, e, f\}$ eine Kette. Die Menge $\{g, e, d\}$ ist eine Antikette.

Satz 3.6. Jede endliche Halbordnung hat mindestens ein minimales und mindestens ein maximales Element.

Beweis. Sei $a_1 < a_2 < \dots < a_n$ eine Kette maximaler Länge⁸. Wir zeigen mit einem Widerspruchsbeweis, dass dann a_1 minimales und a_n maximales Element ist. Angenommen, a_1 wäre nicht minimales Element. Dann gibt es ein a' mit $a' \neq a_1$ und $a' < a_1$.

⁸Also eine Teilmenge von A , die keine Kette mehr ist, wenn ein weiteres Element hinzugefügt wird

Die Kette $a' < a_1 < a_2 < \dots a_n$ wäre dann länger als die ursprüngliche Kette. Nach Annahme hatte die ursprüngliche Kette aber maximale Länge, ein Widerspruch.


Ein analoges Argument zeigt, dass a_n ein maximales Element ist. \square


Zum Abschluss noch eine letzte Definition, die der linearen Erweiterung. Grob gesagt ist eine lineare Erweiterung einer Halbordnung ist eine totale Ordnung (Reihenfolge der Elemente), welche die Ordnung innerhalb der Halbordnung berücksichtigt. Wenn also a mit b bezüglich der Halbordnung in Relation steht, muss a vor b in der totalen Ordnung stehen.

Definition 3.22 (Lineare Erweiterung). Eine totale Ordnung (M, \preceq) heißt **lineare Erweiterung** einer Halbordnung (M, \leq) , falls aus $a \leq b$ folgt, dass $a \preceq b$ in der totalen Ordnung gilt.

Beispiel. Eine lineare Erweiterung für $(\mathcal{P}(\{1, 2, 3\}), \subseteq)$ ist die totale Ordnung, welche die Elemente in der folgenden Reihenfolge aufzählt:

$$\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$$

Für jedes Paar M_1, M_2 mit $M_1 \subseteq M_2$ gilt, dass M_1 vor M_2 in der totalen Ordnung ist. 

Selbsttest: Finden Sie eine lineare Erweiterung für die Halbordnung rechts in Abbildung 3.4. 

Im Folgenden zeigen wir nun, dass jede endliche Halbordnung eine lineare Erweiterung hat. Dieser Beweis ist insbesondere interessant da er *konstruktiv* ist. Das bedeutet, dass der Beweis nicht nur zeigt, dass es eine lineare Erweiterung gibt, er beschreibt auch ein Verfahren, um die lineare Ordnung zu konstruieren.

Satz 3.7. Jede endliche Halbordnung (M, \leq) hat eine lineare Erweiterung.

Beweis. Finde ein minimales Element $m \in M$. Ein solches Element existiert immer, nach Satz 3.6. Setze m als Minimum der linearen Erweiterung. $M \setminus \{m\}$ ist weiter eine endliche Halbordnung. Iteriere den Prozess, bis M leer ist.

Dass dieses Verfahren in der Tat korrekt funktioniert, kann mit Hilfe eines induktiven Beweises gezeigt werden. Für diesen unterteilen wir M während des iterativen Verfahrens in zwei Mengen F und U . In F werden alle Elemente gespeichert, die als minimales Element aus M entfernt wurden, in U alle Elemente, die noch nicht aus M entfernt wurden. Es gilt also $F \cup U = M$. Wir zeigen nun, dass nach jeder Iteration die folgende Invariante⁹ gilt: Die Elemente in F bilden in der Reihenfolge, in der sie entfernt wurden, eine lineare Erweiterung von F und es gibt kein Paar $m' \in U, m \in F$ mit $m' < m$.

⁹Eigenschaft

Nach der ersten Iteration ist ein Element m entfernt worden. Dieses Element alleine ist eine lineare Erweiterung der Menge $\{m\}$. Da m ein minimales Element ist, gibt es kein Element m' in $M \setminus \{m\}$ mit $m' < m$ und beide Teile der Invariante sind erfüllt.

Vor einer beliebigen Iteration i können wir davon ausgehen, dass die Invariante gilt. Also die Reihenfolge in der bisher aus M entfernten Elemente betrachtet wurden bilden eine lineare Erweiterung und es gibt kein $m' \in U$, sodass es ein $m \in F$ gibt mit $m' < m$. Nun wird das minimale Element m_i aus U in F verschoben. Da $m < m_i$ für alle $m \in F$ gilt, reiht sich dieses Element in die lineare Erweiterung ein. Da m_i minimales Element ist, gibt es kein Element $m'_i \in F \setminus \{m_i\}$ mit $m'_i < m_i$ und die Invariante gilt weiter.

Nachdem alle Elemente aus U in F verschoben wurden, gilt $F = M$ und da die Invariante weiter gilt, bilden die Elemente in der Reihenfolge, in der sie entfernt wurden, eine lineare Erweiterung von M . \square

Die oben gezeigte lineare Erweiterung von $(\mathcal{P}(\{1, 2, 3\}), \subseteq)$ kann mit dem beschriebenen Verfahren gefunden werden.

Wir werden das Konzept der linearen Erweiterung in Kapitel 6 wieder aufgreifen, wenn wir über das Konzept der *topologischen Sortierung* sprechen. Im Allgemeinen haben lineare Erweiterungen bei Scheduling-Problemen Anwendungen. Angenommen, wir haben eine Menge von Aufgaben, die bearbeitet werden müssen. Dabei stehen a und b in Relation zueinander, wenn Aufgabe a vor Aufgabe b erledigt werden muss. Diese Relation ist eine Halbordnungsrelation. Eine lineare Erweiterung der Relation ist dann eine Reihenfolge, in der die Aufgabe erledigt werden können, sodass die Einschränkungen für die Reihenfolge eingehalten werden.

3.3 Funktionen

Voraussetzung:

- Abschnitt 3.1, Abschnitt 3.2
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ und elementare Rechenoperationen auf diesen Mengen.

Lernziele

Die Studierenden ...

- ... untersuchen Funktionen auf Eigenschaften
- ... beweisen Aussagen zu Funktionen mit gegebenen Eigenschaften

Wir werden nun sehen, wie das Konzept der *Funktion*, welches schon zum Beispiel aus der Schule bekannt ist in den Kontext der Relationen eingebettet werden kann.

An einigen Stellen werden wir aber auch sehen, dass sich die Bezeichner und Begriffe zwischen Funktionen und Relationen doch unterscheiden.

Definition 3.23 (Funktion). Eine **Funktion** f ist eine Relation zwischen zwei Mengen A und B , sodass für jedes $a \in A$ genau ein $b \in B$ mit $a f b$ existiert.

Wir schreiben auch:

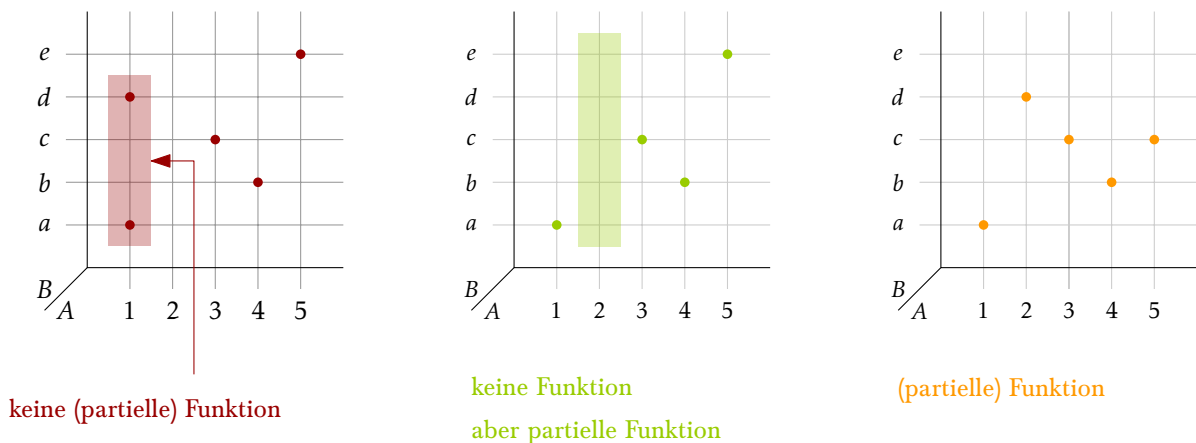
$$f : A \rightarrow B$$

$$f(a) = b$$

Die Menge A wird **Definitionsbereich** und die Menge B wird **Wertebereich** genannt.

Eine Relation ist eine **partielle Funktion**, wenn jedem a höchstens ein b zugeordnet wird.

Beispiel. In der folgenden Abbildung sind drei Relationen abgebildet, nur eine davon ist auch eine Funktion. Im Vergleich zur Darstellung der Relationen, die wir bis jetzt gesehen haben, ist die Darstellung gespiegelt. Dies ist näher an den klassischen Graphen, die zur Darstellung von Funktionen bekannt sind.



Wir definieren nun zwei wichtige Begriffe für Funktionen:

Definition 3.24 (Bild, Urbild). Sei $f : A \rightarrow B$ eine Funktion und $M \subseteq A, N \subseteq B$. Dann ist das **Bild** von M unter f definiert, als

$$f(M) = \{y \in B \mid \text{es gibt ein } x \in M \text{ mit } f(x) = y\}$$

$$= \bigcup_{x \in M} \{f(x)\}$$

Das Urbild von N unter f ist definiert als:

$$f^{-1}(N) = \{x \in A \mid f(x) \in N\}$$

Anmerkung: Für Mengen N mit nur einem Element x wird auch manchmal $f^{-1}(x)$ statt $f^{-1}(\{x\})$ geschrieben.



Intuitiv gibt das Bild einer Teilmenge $M \subseteq A$ alle Elemente aus B an, die von mindestens einem Element aus M getroffen werden. Das Bild einer Teilmenge von A ist also eine Teilmenge von B . Das Urbild einer Teilmenge $N \subseteq B$ sind alle Elemente, aus A , die eines der Elemente aus N treffen. Das Urbild ist also eine Teilmenge von A .

Beispiel. Wenn f die einzige Funktion aus dem Beispiel oben ist, dann ist das Bild $f(\{1, 3, 4\})$ von $\{1, 3, 4\}$ die Menge $\{c, b, a\}$ und das Urbild von $\{b, c\}$ ist $f^{-1}(\{b, c\}) = \{3, 4, 5\}$. ◀

Nun können wir einige Eigenschaften von Funktionen definieren:

Definition 3.25 (Surjektivität, Injektivität, Bijektivität). Sei $f : A \rightarrow B$ eine Funktion.

Surjektivität f heißt **surjektiv**, falls jedes Element aus B im Bild auftritt.

$$f(A) = B$$

Intuitiv: Jedes Element aus B wird von mindestens einem Element aus A getroffen.

Injektivität f heißt **injektiv**, wenn aus $f(a) = f(a')$ folgt, dass $a = a'$.

Intuitiv: Jedes Element aus B wird von höchstens einem Element aus A getroffen.

Bijektivität f heißt **bijektiv**, wenn f surjektiv und injektiv ist.

Intuitiv: Jedes Element aus B wird von genau einem Element aus A getroffen.

Beispiel.

1. In Abbildung 3.5 sind vier Funktionen zusammen mit ihren Eigenschaften abgebildet.
2. Ein weiteres Beispiel, an dem die Eigenschaften betrachtet werden können, ist die Funktion $\sin : \mathbb{R} \rightarrow \mathbb{R}$. Für Definitions- und Wertebereich \mathbb{R} ist die Funktion weder injektiv noch surjektiv. Sie ist nicht injektiv, da $\sin(\pi) = \sin(0)$ gilt, aber $\pi \neq 0$. Sie ist nicht surjektiv, da es kein $x \in \mathbb{R}$ gibt, sodass $\sin(x) = 2$ gilt.

Wird nun der Wertebereich auf $[-1, 1]$ eingeschränkt, ist die Funktion surjektiv, da jeder Wert zwischen -1 und 1 mindestens einmal getroffen ist. Die Funktion ist aber immer noch nicht injektiv, da weiterhin $\sin(\pi) = \sin(0)$ aber $\pi \neq 0$ gilt.

Wird der Definitionsbereich auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$ eingeschränkt, der Wertebereich aber wieder auf \mathbb{R} gesetzt, ist die Funktion injektiv, aber nicht surjektiv.

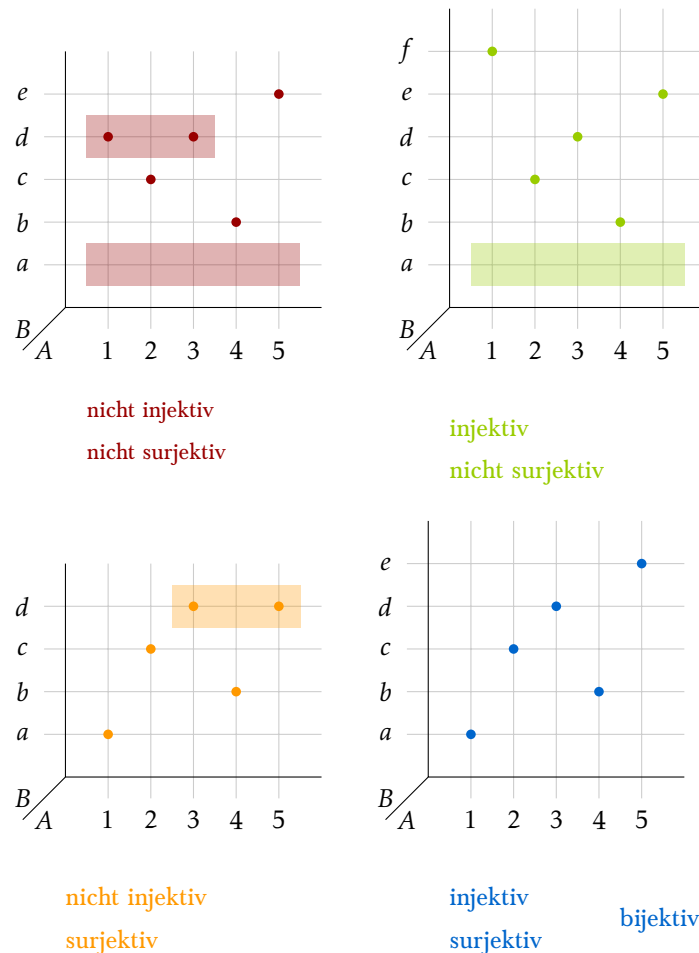


Abbildung 3.5: Beispiel für die Eigenschaften von Funktionen

Erst für Definitionsbereich $[-\frac{\pi}{2}, \frac{\pi}{2}]$ und Wertebereich $[-1, 1]$ ist die Funktion surjektiv und injektiv und damit bijektiv. Siehe Abbildung 3.6 für eine grafische Darstellung der Argumente. ◀

Wird $f : A \rightarrow B$ als Relation betrachtet, ist die inverse Relation f^{-1} eine Funktion mit Definitionsbereich B und Wertebereich A , genau dann, wenn f bijektiv ist. Diese Aussage werden wir hier im Skript nicht beweisen. f^{-1} heißt dann die zu f *inverse Funktion*.

Anmerkung: Es wurde jetzt zweimal die Notation f^{-1} für verschiedene, aber verwandte, Konzepte verwendet. Wird f als Relation betrachtet, also als Menge von Tupeln, kann die inverse Relation f^{-1} auch wieder als Menge von Tupeln gesehen werden. In Definition 3.24 wird mit $f^{-1}(N)$ jedoch das Urbild einer Teilmenge von $N \subseteq B$ bezeichnet. Man kann mit dieser Definition eine Funktion $f^{-1} : \mathcal{P}(B) \rightarrow \mathcal{P}(A)$



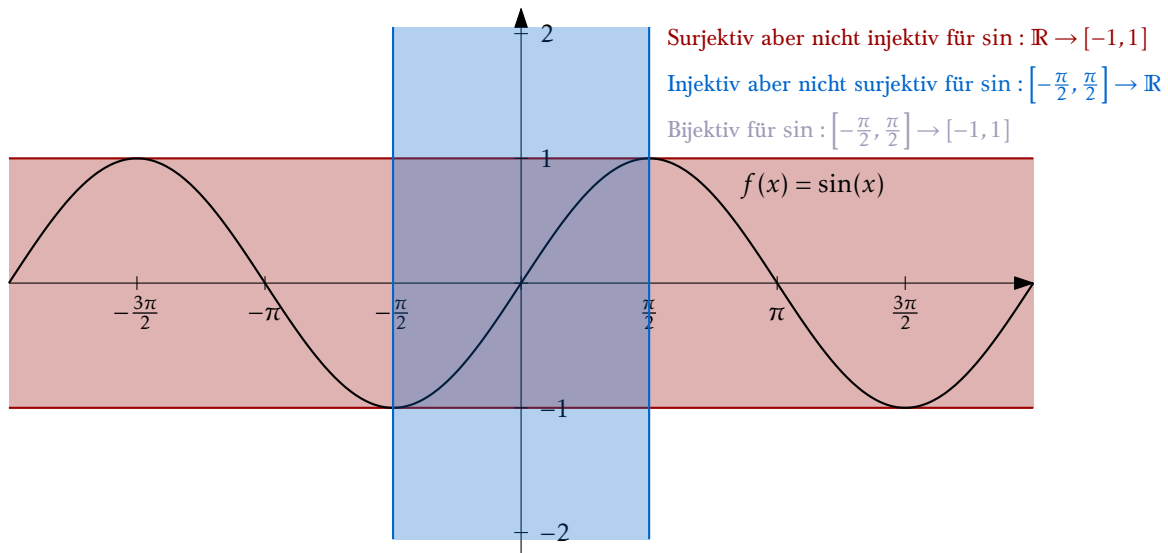
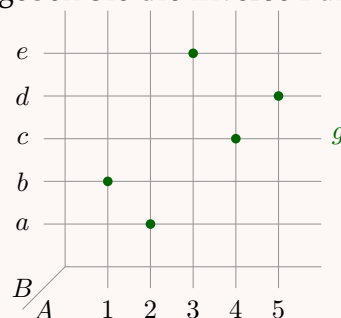
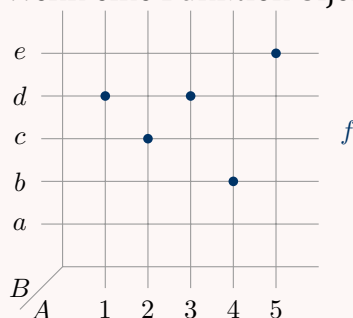


Abbildung 3.6: Einschränkung von Definitions- und Wertebereich und der Einfluss auf Injektivität, Surjektivität und Bijektivität

definieren, die einer Teilmenge von B ihr Urbild, eine Teilmenge von A zuweist. Bei allgemeinen Funktionen, kann das Urbild einer einelementigen Menge $\{x\}$ weniger oder mehr als ein Element beinhalten. Ist die Funktion bijektiv, besteht das Urbild aus nur einem Element. In dem Fall schließt sich wieder der Kreis zur Interpretation als Relation, weil dann $f^{-1}(\{x\})$ das eindeutige Element y mit $(x, y) \in f^{-1}$ beinhaltet.

Selbsttest: Betrachten Sie die Funktionen, welche in der folgenden Grafik dargestellt sind. Untersuchen Sie die Funktionen auf Injektivität, Surjektivität und Bijektivität. Wählen Sie verschiedene Teilmengen von A und B und bestimmen Sie jeweils das Bild und das Urbild dieser Teilmengen.

Wenn eine Funktion bijektiv ist, geben Sie die inverse Funktion an.



Wir betrachten nun noch eine weitere Definition. An dieser ist gut zu sehen, dass einige Begriffe im Kontext der Funktionen und im Kontext von Relationen anders definiert werden.

Definition 3.26 (Funktionsverknüpfung). Sind $f : A \rightarrow B$ und $g : B \rightarrow C$ Funktionen, dann ist die Relationsverkettung $f \circ g$ eine Funktion von A in C . Sie wird **Verknüpfung** von f mit g genannt und durch $g \circ f : A \rightarrow C$ bezeichnet, wobei $g \circ f = g(f(a))$ für alle $a \in A$.

Achtung: Die Relationsverkettung (\circ) wird von links nach rechts gelesen. Die Funktionsverknüpfung (\circ) hingegen von rechts nach links. Wir werden häufig einfach gf statt $g \circ f$ schreiben, wenn aus dem Kontext klar ist, dass die Verknüpfung gemeint ist.

Selbsttest: Betrachten Sie die Funktionen $f(x) = \sin(x)$ und $g(x) = x^2$. Was sind die Funktionsverknüpfungen $f \circ g$ und $g \circ f$?



Lemma 3.8. Die Funktionsverknüpfung ist assoziativ. Es gilt also $(f \circ g) \circ h = f \circ (g \circ h)$

Beweis. Ohne Beweis. □

Bevor wir im Folgenden einige Anwendungen von Funktionen betrachten, hier noch einen Satz, der einige Fakten zu surjektiven, injektiven und bijektiven Funktionen zusammenfasst.

Satz 3.9. Seien $f : A \rightarrow B$ und $g : B \rightarrow C$ zwei Funktionen. Dann gilt:

1. Ist f bijektiv, dann gilt: $f^{-1}f = \text{Id}_A$ und $ff^{-1} = \text{Id}_B$
2. f ist injektiv, genau dann, wenn eine Funktion $h : B \rightarrow A$ existiert mit $hf = \text{Id}_A$
3. f ist surjektiv, genau dann, wenn eine Funktion $h : B \rightarrow A$ existiert mit $fh = \text{Id}_B$
4. Sind f und g injektiv, dann ist auch gf injektiv
5. Sind f und g surjektiv, dann ist auch gf surjektiv
6. Sind f und g bijektiv, dann ist auch gf bijektiv und es gilt $(gf)^{-1} = f^{-1}g^{-1}$

Wir werden nicht alle Aussagen zeigen, sondern uns auf Punkt 3 fokussieren. Die anderen Aussagen lassen sich ähnlich zeigen. Statt direkt nur Punkt 3 zu zeigen, zeigen wir die folgende stärkere Aussage, aus der Punkt 3 direkt folgt.

Lemma 3.10. Die folgenden Bedingungen sind äquivalent:

- (1) $f : A \rightarrow B$ ist surjektiv.
- (2) $\forall b \in B : f^{-1}(b) \neq \emptyset$, oder verbal, Es gibt kein $b \in B$ mit leerem Urbild.
- (3) $\exists g : B \rightarrow A : f \circ g = \text{Id}_B$, oder verbal: Es gibt eine Funktion, g , sodass $f(g(b)) = b$ für alle $b \in B$.

(4) $\forall C : \forall r, s : B \rightarrow C : (r \circ f = s \circ f) \rightarrow r = s$, oder verbal: Wenn die Verknüpfung von zwei Funktionen mit f gleich ist, sind auch die Funktionen gleich.

Beweis. Wir zeigen das Lemma mit einem sogenannten Ringbeweis. Statt paarweise die Äquivalenzen zu zeigen, zeigen wir, (1) \rightarrow (2), (2) \rightarrow (3), (3) \rightarrow (4) und (4) \rightarrow (1). Durch die Definition der Implikation ergibt sich dann, dass alle vier Aussagen äquivalent sind.

- (1) \rightarrow (2) Wir führen einen direkten Beweis. Nach Annahme ist f surjektiv, also $f(A) = B$. Dies kann geschrieben werden als $\forall b \in B : \exists a \in A : f(a) = b$. Es gibt also für jedes $b \in B$ mindestens ein a mit $f(a) = b$. Aus der Definition des Urbilds folgt dann direkt, dass $a \in f^{-1}(b)$ ist. Dies impliziert direkt (2), da $f^{-1}(b)$ dann nicht leer sein kann.
- (2) \rightarrow (3) Wir führen wieder einen direkten Beweis. Für jedes $b \in B$ wählen wir ein festes $a_b \in f^{-1}(b)$. Da kein Urbild leer ist, gibt es für jedes b ein solches a_b . Nach der Definition des Urbilds gilt also $f(a_b) = b$. Wir definieren nun eine Funktion g mit $g(b) = a_b$. Dann ist $f(g(b)) = f(a_b) = b$ für alle $b \in B$.
- (3) \rightarrow (4) Sei C eine beliebige Menge und seien r, s Funktionen von B in C . Um (3) \rightarrow (4) zu zeigen, können wir zunächst die Aussage aus (3) als Voraussetzung annehmen. Da (4) eine Aussage der Form *Aus x folgt y* ist, können wir auch die Voraussetzung der Implikation aus (4) also $r \circ f = s \circ f$ als wahr annehmen. Dann kann mit Hilfe von Lemma 3.8 und diesen Annahmen, kann nun $r = s$ wie folgt geschlussfolgert werden. Es gilt:

$$\begin{aligned}
 r &= r \circ \text{Id}_B \\
 &= r \circ (f \circ g) && (3) \\
 &= (r \circ f) \circ g && \text{Assoziativität von } \circ \\
 &= (s \circ f) \circ g && \text{Annahme aus Punkt (4)} \\
 &= s \circ (f \circ g) && \text{Assoziativität von } \circ \\
 &= s \circ \text{Id}_B && (3) \\
 &= s
 \end{aligned}$$

- (4) \rightarrow (1) Wir zeigen diese Richtung durch Kontraposition, also wir zeigen $\neg(1) \rightarrow \neg(4)$. Die Negation von (1) ist $\exists b_0 : \forall a \in A : f(a) \neq b_0$ oder verbal es gibt ein $b_0 \in B$, sodass es kein $a \in A$ mit $f(a) = b_0$ gibt. Die Negation von (4) ist $\exists C : \exists r, s : B \rightarrow C : (r \circ f) = (s \circ f) \wedge r \neq s$. Verbal kann die Negation gelesen werden als *Es gibt eine Menge C und Funktionen r, s von B nach C , sodass $r \circ f = s \circ f$ und r und s sind verschiedene Funktionen.*

Um zu zeigen, dass eine Auswahl von C, r, s existiert, konstruieren wir diese direkt. Sei b_0 ein Element mit $f(a) \neq b_0$ für alle $a \in A$. Dieses Element existiert nach der Negation von (1). Setze $C = \{0, 1\}$ und definiere r, s wie folgt:

$$r(b) = 0 \quad \forall b \in B$$

$$\begin{aligned}s(b_0) &= 1 \\ s(b) &= 0 \quad \forall b \in B \setminus \{b_0\}\end{aligned}$$

Die Funktion r bildet also alle Elemente aus B auf 0 ab, während s das Element b_0 auf 1 abbildet, die restlichen Elemente aber auf 0.

Dann ist $r \circ f = s \circ f$, da durch Anwenden von f , nie das Element b_0 getroffen werden kann und die Funktionen r und s für alle anderen Werte beide immer 0 zurückgeben. Aber $r \neq s$, da für b_0 das Ergebnis verschieden ist. Damit ist gezeigt, dass aus $\neg(1)$ die Aussage $\neg(4)$ folgt. \square

Ähnlich könnte auch das folgende Lemma gezeigt werden, aus dem dann direkt die zweite Aussage aus Satz 3.9 folgt:

Lemma 3.11. *Die folgenden Bedingungen sind äquivalent.*

- (1) $f : A \rightarrow B$ ist injektiv.
- (2) $\forall b \in B : |f^{-1}(b)| \leq 1$
- (3) $\exists g : B \rightarrow A : g \circ f = \text{Id}_A$
- (4) $\forall C : \forall r, s : C \rightarrow A : (f \circ r = f \circ s) \rightarrow r = s$

Anwendungen von Funktionen In diesem Skript gibt es verschiedene Anwendungen der Begriffe und Konzepte, die in diesem Abschnitt eingeführt wurden. Die Interpretation von Booleschen Junktoren als Funktionen mit allen Auswirkungen dazu werden in Abschnitt 1.3 betrachtet. Das Schubfachprinzip ist eine allgemeine Beweismethode, welche für die Definition der Konzepte der Funktionen benötigt, siehe Abschnitt 2.4. Auswirkungen auf die Abzählbarkeit von Mengen werden im folgenden Abschnitt 3.4 behandelt.

3.4 Abzählbarkeit von Mengen

Voraussetzung:

- Mengen (Abschnitt 3.1), Funktionen (Abschnitt 3.3)
- Kennen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ und elementare Rechenoperationen auf diesen Mengen.





Lernziele

Die Studierenden ...

- ... untersuchen gegebene Mengen auf Abzählbarkeit.
- ... zeigen Aussagen im Bereich der Abzählbarkeit.

Eine Anwendung von bijektiven Funktionen ist die Betrachtung von Kardinalitäten von Mengen. In Abschnitt 3.1 haben wir schon definiert, dass für endliche Mengen M die Größe oder auch Mächtigkeit/Kardinalität $|M|$ als die Anzahl der Elemente in der Menge definiert ist.

Wir werden nun eine allgemeine Definition dafür angeben, dass zwei Mengen die gleiche Mächtigkeit haben, welche auch für unendliche Mengen benutzt werden kann.

Definition 3.27. Zwei Mengen A und B heißen **gleichmächtig**, wenn es eine bijektive Abbildung von A nach B gibt.

Sind A und B endlich kann die Bijektion einfach erstellt werden, indem eine Reihenfolge der Elemente in A und B festgelegt wird und das i te Element aus A auf das i te Element in B abgebildet wird. Für unendliche Mengen werden wir später Beispiele der Bijektionen sehen.

Definition 3.28. Eine unendliche Menge A heißt **abzählbar unendlich**, falls sie gleichmächtig mit \mathbb{N} ist.

Eine Menge heißt **abzählbar**, wenn sie endlich oder abzählbar unendlich ist.

Mit diesen beiden Definitionen, können wir nun die folgende Beobachtung machen und beweisen:

Beobachtung 3.12. Ist $f : \mathbb{N} \rightarrow A$ eine Surjektion, so ist A abzählbar.

Beweis. Ist A endlich, dann muss nicht viel bewiesen werden, da A dann per Definition direkt abzählbar ist. Wenn f schon eine Bijektion ist, sind wir fertig, weil dann nach Definition 3.27 beide Mengen gleichmächtig sind.

Wir betrachten nun also den Fall, dass f keine Bijektion ist und A unendlich ist. Wir wollen nun eine Funktion g definieren, bei der jedes Element aus A von genau einem Element aus \mathbb{N} getroffen wird. Wir definieren die neue Funktion $g : \mathbb{N} \rightarrow A$ wie folgt:

$$\begin{aligned} g(0) &= f(0) \\ g(n) &= f(m) \quad \text{mit } m = \min\{i \mid f(i) \notin \{g(0), \dots, g(n-1)\}\} \end{aligned}$$

Die grundlegende Idee ist die Folgende: Die Zahlen aus \mathbb{N} werden nach und nach betrachtet. Sei n die aktuelle Zahl. Wenn $f(n)$ (ein Element aus A) noch nicht als

Funktionswert $g(n')$ einer Zahl $n' < n$ gesetzt wurde, ist $g(n) = f(n)$. Ansonsten wird das kleinste $m > n$ gesucht, sodass $f(m)$ noch nicht Funktionswert einer Zahl n' mit $n' < n$ ist und es wird $g(n) = f(m)$ gesetzt.

Auf diese Art und Weise wird jedem Element aus \mathbb{N} ein eindeutiges Element aus A zugewiesen. Da f surjektiv ist, wird jedes Element aus A irgendwann als Funktionswert gewählt. Dann ist g eine Bijektion $\mathbb{N} \rightarrow A$ und A ist damit abzählbar unendlich. \square

Selbsttest: Definieren Sie eine surjektive aber nicht bijektive Funktion von \mathbb{N} in eine beliebige unendliche Menge. Stellen Sie die Wertetabelle für g auf, die sich aus dem Verfahren im Beweis ergibt.

Aus Beobachtung 3.12 folgt direkt die nächste Beobachtung, die hier ohne Beweis genannt wird.

Beobachtung 3.13. Ist A abzählbar und $B \subseteq A$, so ist B abzählbar.

Selbsttest: Beweisen Sie die Beobachtung 3.13.

Im Folgenden betrachten wir einige Mengen für die gezeigt werden kann, dass diese abzählbar oder auch nicht abzählbar sind.

Lemma 3.14. \mathbb{Z} ist abzählbar.

Beweis. Wir zeigen die Aussage, indem wir eine Bijektion zwischen \mathbb{N} und \mathbb{Z} angeben. Da jede bijektive Funktion eine Umkehrfunktion hat, geben wir die Funktion als $f : \mathbb{Z} \rightarrow \mathbb{N}$ an.

$$f : \mathbb{Z} \rightarrow \mathbb{N}$$

$$f(n) = \begin{cases} 2n & , n \geq 0 \\ -2n - 1 & , n < 0 \end{cases}$$

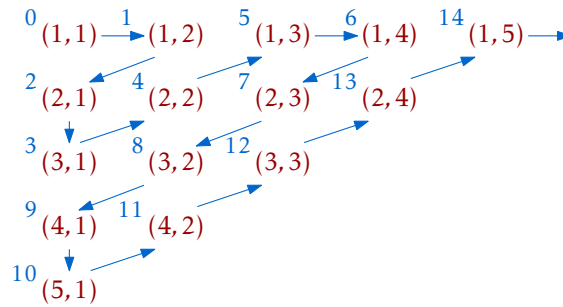
Zu zeigen, dass f eine Bijektion ist, ist eine Übungsaufgabe. Grafisch kann man sich die Zuordnung wie folgt vorstellen:

...	-5	-4	-3	-2	-1	0	1	2	3	4	5	...	\mathbb{Z}
...	9	7	5	3	1	0	2	4	6	8	10	...	\mathbb{N}

\square

Lemma 3.15. $\mathbb{N}_+ \times \mathbb{N}_+$ ist abzählbar.

Beweis. Wir geben eine Bijektion $f : \mathbb{N}_+ \times \mathbb{N}_+ \rightarrow \mathbb{N}$ an, zunächst durch das folgende Bild:



Die Nummerierung folgt dabei den Pfeilen in einem Schlangenmuster. Es ist zu sehen, dass dabei jede Zahl aus \mathbb{N} einmal getroffen wird. Formal kann die Abbildung als

$$f(a, b) = \begin{cases} \frac{(a+b-1)(a+b-2)}{2} + (b-1) & , a+b \text{ gerade} \\ \frac{(a+b-1)(a+b-2)}{2} + (a-1) & , a+b \text{ ungerade} \end{cases}$$

angegeben werden. Dieses Argument geht auf Georg Cantor zurück. \square

Lemma 3.16. \mathbb{Q}^+ ist abzählbar.

Beweis. Jede Zahl aus \mathbb{Q}^+ kann als $\frac{p}{q}$ mit p, q teilerfremd dargestellt werden. Nach Beobachtung 3.13 und Lemma 3.15 ist \mathbb{Q}^+ abzählbar. \square

Selbsttest: Ist \mathbb{N}^3 abzählbar?



Lemma 3.17. Die Menge aller endlichen Teilmengen von \mathbb{N} ist abzählbar.

Beweis. Wir weisen jeder der Teilmengen die Summe der in ihr enthaltenen Elemente zu. Das ist eine natürliche Zahl. Zu jeder möglichen Summe gibt es nur endliche viele Teilmengen mit genau dieser Summe, diese können dann alle nacheinander aufgezählt werden. Zunächst werden also alle endlichen Teilmengen mit Summe 0 aufgeschrieben, dann die beiden endlichen Teilmengen mit Summe 1 und so weiter¹⁰. \square

Lemma 3.18. Die Vereinigung von abzählbar vielen abzählbaren Mengen ist abzählbar.

Beweis. Seien M_1, M_2, \dots die Mengen. Wir verwenden eine ähnliche Idee, wie in Lemma 3.15. Statt dass wir Tupel in die Reihen und Spalten schreiben, werden in die erste Zeile die Elemente von M_1 , in die zweite Zeile die Elemente von M_2 und so weiter.

Nun können die Elemente wieder wie im Beweis von Lemma 3.15 in Schlangenlinien aufgezählt werden, wobei eventuell doppelte Elemente übersprungen werden. \square

Bis jetzt haben wir viele unendliche Mengen gesehen, die abzählbar sind. Nun werden wir sehen, dass nicht alle Mengen abzählbar sind.

¹⁰Erinnerung: 0 ist eine natürliche Zahl

Satz 3.19. Die Menge aller^a Untermengen von \mathbb{N} ist nicht abzählbar.

^aauch der unendlichen

Beweis. Die Menge aller Untermengen von \mathbb{N} ist genau $\mathcal{P}(\mathbb{N})$, also die Potenzmenge von \mathbb{N} . Der Beweis ist eine andere Art eines Diagonalisierungsarguments. Wir führen einen Widerspruchsbeweis. Wir nehmen also an, dass $\mathcal{P}(\mathbb{N})$ abzählbar ist und leiten daraus einen Widerspruch her.

Angenommen $\mathcal{P}(\mathbb{N})$ ist abzählbar, dann gibt es eine Bijektion $f : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N})$. Für diese Funktion, definieren wir nun eine spezifische Teilmenge von \mathbb{N} :

$$S_f = \{n \in \mathbb{N} \mid n \notin f(n)\}$$

Um diese Definition zu verstehen, ist es wichtig zu sehen, dass $f(n) \subseteq \mathbb{N}$ für jedes n gilt, daher ist die Funktion wohldefiniert. Es gibt also keinen Selbstbezug, da der Index der Menge mit den Elementen der Menge verglichen wird.

Beispiel. Wir betrachten die ersten Funktionswerte einer hypothetischen¹¹ Funktion f und erstellen die Menge S_f für diese ersten Werte:

$$\begin{aligned}f(0) &= \{1, 2, 3, \dots\} \\f(1) &= \{2, 4, 6, \dots\} \\f(2) &= \{2, 7, 12, \dots\} \\f(3) &= \{1, 3, 5, \dots\} \\f(4) &= \{3, 6, 9, \dots\} \dots\end{aligned}$$

Dann ist $0 \in S_f$, da $0 \notin f(0)$, aber $2 \notin S_f$, da $2 \in f(2)$. ◀

Da f eine Bijektion ist, werden alle möglichen Teilmengen von \mathbb{N} getroffen insbesondere also auch S_f , also gibt es ein n_0 mit $f(n_0) = S_f$. Nun können wir Fragen, ob $n_0 \in S_f$ ist. Nach der Definition von S_f gilt, dass $n_0 \in S_f$ ist, genau dann, wenn $n_0 \notin S_f$, was genau der gesuchte Widerspruch ist. □

Da wir im Beweis nie wirklich benutzt haben, dass wir über Potenzmengen von \mathbb{N} reden, kann mit dem gleichen Beweis gezeigt werden, dass es für jede beliebige Menge X keine Bijektion zwischen X und der Potenzmenge von X gibt.

Definition 3.29. Eine Menge, die nicht abzählbar ist, heißt **überabzählbar**.

Die Potenzmenge von \mathbb{N} ist also überabzählbar.

Wir schauen nun noch eine weitere Menge an und zeigen, dass diese überabzählbar ist.

¹¹Wir wissen nicht genau, wie das f aus dem Beweis aussehen würde. Um die Definition von S_f nachvollziehen zu können, denken wir uns also eine konkrete Funktion aus.

Satz 3.20. \mathbb{R} ist überabzählbar.

Beweis. Wir zeigen nicht, dass ganz \mathbb{R} überabzählbar ist, sondern beschränken uns auf das offene Intervall $(0, 1)$ ¹². Als Beweistechnik verwenden wir wieder einen Widerspruchsbeweis mit der Cantorschen Diagonalisierungsmethode.

Angenommen $f : \mathbb{N} \rightarrow (0, 1)$ ist eine Bijektion und wir stellen f als unendliche Wertetabelle dar, in der jeder Wert als unendlicher Dezimalbruch dargestellt wird.

0:	0,	1	4	5	8	8...	$r_f = 0,08726$
1:	0,	0	7	5	7	9...	
2:	0,	1	8	7	8	3...	
3:	0,	5	1	7	0	2...	
4:	0,	6	1	6	2	3...	
5:	0,	8	5	1	7	6...	
⋮							

Wir ordnen dieser festen Wertetabelle eine Zahl $r_f \in (0, 1)$ zu, die als die Zahl auf der Diagonale definiert ist. Im Beispiel ist das also $r_f = 0.08726$.

Wir erzeugen nun aus r_f eine neue Zahl t_f und argumentieren, dass diese nicht in der Wertetabelle enthalten sein kann. Die neue Zahl ist definiert, indem in r_f jede Nachkommastelle x durch $(x + 1) \bmod 10$ ersetzt wird, es sei denn $x = 8$, dann ist die Nachkommastelle 0. Wenn die erste Nachkommastelle dieser Zahl der ersten Nachkommastelle von $f(0)$ entspricht, wird die erste Nachkommastelle noch einmal erhöht. Im Beispiel wird also aus 0.08726 zunächst 0.10837 und dann $t_f = 0.20837$.

Da wir angenommen haben, dass f eine Bijektion ist, die alle Elemente aus \mathbb{R} trifft, muss es ein n_0 geben, mit $f(n_0) = t_f$. Wir wissen $n_0 \neq 0$, da wir darauf geachtet haben, dass sich $f(0)$ und t_f an der ersten Nachkommastelle unterscheiden. Für alle anderen $k \in \mathbb{N}$ gilt ebenfalls $n_0 \neq k$. Die k -te Nachkommastelle von $f(k)$ ist nach Definition von r_f genau die k -te Nachkommastelle von r_f . Da sich t_f und r_f an allen Nachkommastellen unterscheiden, kann in der Zeile nicht t_f stehen, ein Widerspruch. \square

Hinweis: Die Regeln um t_f aus r_f zu bauen sind so nicht fest. Wichtig ist, dass sich am Ende $f(0)$ und t_f in der ersten Stelle unterscheiden, r_f und t_f an allen Stellen verschieden sind und die Zahl nicht auf einer Periode von 9 endet.

Es gibt noch einige weitere Anmerkungen zum Thema Abzählbarkeit:

1. Wie wissen jetzt, dass die Menge $\mathbb{R} \setminus \mathbb{Q}$ der irrationalen Zahlen überabzählbar ist.

Selbsttest: Warum gilt das?

2. Eine Zahl heißt **algebraisch**, wenn Sie Lösung einer Gleichung der Form $a_0 + a_1x + a_2x^2 + \dots + a_nx^n = 0$ mit allen $a_i \in \mathbb{Z}$ ist. Zum Beispiel ist jede rationale Zahl $\frac{p}{q}$ algebraisch, da sie die Lösung von $-p + qx = 0$ ist. Auch die irrationale Zahl $\sqrt{7}$

¹²Also die Menge $\{x \in \mathbb{R} \mid 0 < x < 1\}$

ist algebraisch ($-7 + x^2 = 0$). Analog zu Lemma 3.17 kann gezeigt werden, dass es nur abzählbar viele solche Gleichungen gibt und daher auch nur abzählbar viele algebraische irrationale Zahlen.

Daraus folgt, dass es überabzählbar viele nichtalgebraische irrationale Zahlen gibt. Diese Zahlen heißen *transzendent*. Es ist nicht einfach zu zeigen, dass eine Zahl transzendent ist. Beispiele für transzendente Zahlen sind π und e .

3. Die Anzahl $|\mathbb{N}|$ wird oft mit \aleph_0 (gesprochen Aleph-Null) bezeichnet. \aleph ist der erste Buchstabe des hebräischen Alphabets. Es gibt die sogenannte *Kontinuumshypothese*, welche besagt, dass es keine Menge gibt mit $\aleph_0 < |X| < |\mathbb{R}|$. Es ist bekannt, dass diese Hypothese aus den Standardaxiomen der Mengenlehre weder bewiesen noch widerlegt werden kann.

Entwurf

II

Kombinatorik und Wahrscheinlichkeitsrechnung

KAPITEL 4

Kombinatorik

Voraussetzung:

- Mengen Abschnitt 3.1
- Funktionen Abschnitt 3.3



Lernziele

Die Studierenden ...

- ...bestimmen die Größe von strukturiert beschriebenen Mengen.
- ...zeigen Aussagen aus der Kombinatorik



4.1 Grundregeln und Inklusion-Exklusion

In diesem Kapitel beschäftigen wir uns damit, wie die Größe von strukturiert beschriebenen endlichen Mengen bestimmt werden kann. Als Beispiel kann man sich die Frage stellen, wie viele Binärstrings der Länge 8 existieren, die auf 01 oder 11 enden. Für dieses Beispiel ist die Antwort relativ einfach zu sehen, alle Binärstrings, die auf 01 oder 11 enden sind einfach alle Binärstrings, die auf 1 enden. Damit können nur die vordersten 7 Stellen frei gewählt werden und es gibt 2^7 solcher Strings. Im Folgenden werden wir einige Grundregeln kennenlernen mit deren Hilfe solche Größen bestimmt werden können und diese auf verschiedene Beispiel anwenden.

Sei S die Menge, deren Größe wir bestimmen wollen, dann gibt es die folgenden drei Regeln:

1. Gleichheitsregel: $|S| = |T|$ für eine andere Menge T , wenn es eine bijektive Abbildung zwischen S und T gibt.

Diese Regel ist einfach zu sehen, wenn man sich erinnert, das bei einer bijektiven Abbildung jedem Element aus S genau ein Element aus T zugeordnet wird und umgekehrt, jedes Element aus T von einem Element aus S getroffen wird.

2. **Summenregel:** Wenn S die Vereinigung von disjunkten Mengen S_1, \dots, S_k ist, dann ist $|S| = \sum_{i=1}^k |S_i|$.

Auch diese Regel lässt sich schnell nachvollziehen. Da jedes Element in genau einer der Teilmengen ist, wird jedes Element in der Summe über die Teilmengen genau einmal betrachtet.

3. **Produktregel:** Ist S das kartesische Produkt der endlichen Mengen S_1, \dots, S_t , dann ist $|S| = \prod_{i=1}^t |S_i|$.

Für $t = 2$ lässt sich die Regel auf einfach sehen, mit jedem Element aus S_1 müssen alle Elemente aus S_2 kombiniert werden. Für $t > 2$ kann die Regel mit vollständiger Induktion gezeigt werden.

Beispiel. Wir betrachten nun einige Beispiele für die Anwendung der Regeln.

1. *Wie viele Stühle passen in einen Hörsaal mit 10 Reihen à 15 Stühlen?*

Wir können jedem Stuhl (bijektiv) eine eindeutige Koordinate aus $\{1, \dots, 10\} \times \{1, \dots, 15\}$ zuweisen. Also gibt es mit der Produktregel genau

$$|\{1, \dots, 10\} \times \{1, \dots, 15\}| = 150$$

Stühle.

2. *Wie viele Möglichkeiten eine Unteraufgabe auszuwählen gibt es, wenn es 3 Aufgaben mit jeweils 5, 12 und 17 Unteraufgaben zur Auswahl gibt?*

Die Menge aller Unteraufgaben ist disjunkt in die drei Aufgaben unterteilt, also gibt es mit der Summenregel $5 + 12 + 17 = 34$ Möglichkeiten.

3. *Wie viele Binärstrings der Länge 8, die auf 01 oder 10 enden gibt es?*

Es gibt 2^6 Wörter, die auf 01 Enden (Produktregel mit \mathbb{B}^6) und 2^6 Wörter, die auf 10 Enden. Die Menge der Wörter die auf 01 Enden und die Menge der Wörter, die auf 10 enden sind disjunkt. Mit der Summenregel gibt es also $2 \cdot 2^6 = 2^7$ Wörter, die auf 01 oder 10 enden. ◀

Selbsttest: Argumentieren Sie mit der Produktregel, dass es 2^n Belegungen für die Variablen einer n -stelligen Booleschen Funktion gibt. ?

Die Summenregel ist nur für disjunkte Mengen hilfreich, häufig sind die Mengen, die wir betrachten, jedoch nicht disjunkt. Für diese Fälle gilt der folgende Satz:

Satz 4.1 (Inklusion-Exklusionsprinzip). Für beliebige Mengen A und B gilt:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Beweis. Wir suchen eine Darstellung von $A \cup B$ als Vereinigung von disjunkten Mengen, auf die wir dann die Summenregel anwenden können. Es gilt:

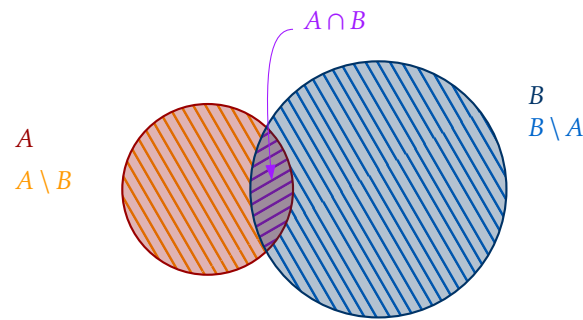


Abbildung 4.1: Darstellung von $A \cup B$ durch disjunkte Teilmengen

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B) \quad (4.1)$$

$$|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B| \quad (4.2)$$

Dabei gilt (4.2) da die drei Mengen aus (4.1) alle disjunkt sind, siehe Abbildung 4.1 für eine Illustration. Nach der Summenregel gilt dann:

$$\begin{aligned} |A| &= |A \setminus B| + |A \cap B| & \Leftrightarrow & & |A \setminus B| &= |A| - |A \cap B| \\ |B| &= |B \setminus A| + |A \cap B| & \Leftrightarrow & & |B \setminus A| &= |B| - |A \cap B| \end{aligned}$$

Fügt man all diese Überlegungen zu zusammen, ergibt sich:

$$\begin{aligned} |A \cup B| &= |A \setminus B| + |B \setminus A| + |A \cap B| \\ &= |A| - |A \cap B| + |B| - |A \cap B| + |A \cap B| \\ &= |A| + |B| - |A \cap B| \end{aligned} \quad \square$$

Beispiel. Was ist die Anzahl der Binärstrings, die mit 001 beginnen oder mit 10 Enden und Länge 12 haben?

Sei A die Menge der Binärstrings, die mit 001 beginnen und sei B die Menge der Binärstrings, die mit 10 Enden. Dann gilt:

$$\begin{aligned} |A| &= 2^9 \\ |B| &= 2^{10} \\ |A \cap B| &= 2^7 \\ |A \cup B| &= 2^9 + 2^{10} - 2^7 = 1408 \end{aligned} \quad \blacktriangleleft$$

Selbsttest: Wie viele Ternärstrings, also Folgen von Zeichen aus $\{0, 1, 2\}$, der Länge 10 gibt es, die mit 20 oder 11 beginnen und mit 0 enden gibt es. ?

Das Prinzip der Inklusion und Exklusion kann auch auf drei (und mehr) Mengen verallgemeinert werden.

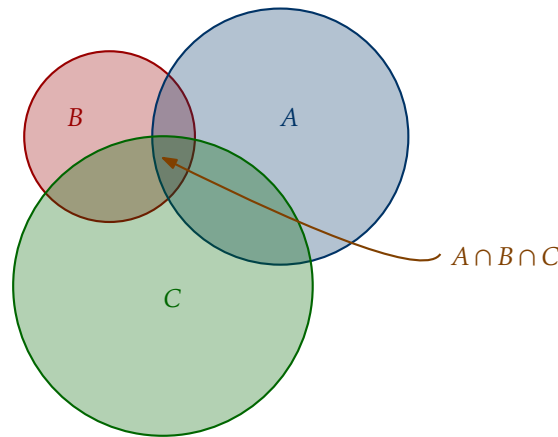


Abbildung 4.2: Darstellung von $A \cup B \cup C$ durch disjunkte Teilmengen

Satz 4.2. Für endliche Mengen A, B und C gilt:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Beweis. Wir betrachten zuerst die Intuition für die Formel. Betrachte den Teil $|A \cap B \cap C|$. Dieser wird bei $|A| + |B| + |C|$ dreimal addiert und bei $-|A \cap B| - |A \cap C| - |B \cap C|$ dreimal wieder abgezogen. Daher muss er noch einmal wieder hinzugefügt werden, siehe Abbildung 4.2.

Formal lässt sich der Beweis auf Satz 4.1 zurückführen:

$$\begin{aligned} |A \cup B \cup C| &= |A \cup B| + |C| - |(A \cup B) \cap C| \\ &= |A \cup B| + |C| - |(A \cap C) \cup (B \cap C)| \\ &= |A \cup B| + |C| - (|A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|) \\ &= |A \cup B| + |C| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \\ &= |A| + |B| - |A \cap B| + |C| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \\ &= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \quad \square \end{aligned}$$

Dieses Prinzip lässt sich beliebig weiter erweitern:

Satz 4.3. Für endliche Mengen A_1, \dots, A_k $k \geq 1$ gilt:

$$\begin{aligned} &\left| \bigcup_{i=1}^k A_i \right| \\ &= \sum_{1 \leq i \leq k} |A_i| - \sum_{1 \leq i_1 < i_2 \leq k} |A_{i_1} \cap A_{i_2}| + \sum_{1 \leq i_1 < i_2 < i_3 \leq k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \dots + (-1)^{k+1} \left| \bigcap_{i=1}^k A_i \right| \end{aligned}$$

Beispiel. Sei $S = \{n \in \mathbb{N} \mid ((2 \mid n) \vee (3 \mid n) \vee (5 \mid n)) \wedge 1 \leq n \leq 3000\}$. Die Frage ist nun: Wie groß ist $|S|$?

Sei $D_i = \{n \in \mathbb{N} \mid (i \mid n), 1 \leq n \leq 3000\}$ die Menge aller durch i teilbaren ganzen Zahlen, die zwischen 1 und 3000 liegen.

Dann ist $S = D_2 \cup D_3 \cup D_5$. Wir berechnen alle nötigen Teilgrößen, um Satz 4.2 anzuwenden.

$$\begin{aligned} |D_2| &= 1500 & |D_2 \cap D_3| &= |D_6| = 500 \\ |D_3| &= 1000 & |D_2 \cap D_5| &= |D_{10}| = 300 \\ |D_5| &= 600 & |D_3 \cap D_5| &= |D_{15}| = 200 \\ |D_2 \cap D_3 \cap D_5| &= |D_{30}| = 100 \end{aligned}$$

Also ist $|D_2 \cup D_3 \cup D_5| = 1500 + 1000 + 600 - (500 + 300 + 200) + 100 = 2200$ ◀

4.2 Permutationen

Im Folgenden werden wir eine Menge A mit n Elementen auch kurz als **n -Menge A** bezeichnen.

Dieser Abschnitt beschäftigt sich mit Permutationen. Intuitiv gesehen, ist eine Permutation einer Menge, eine feste Reihenfolge, in der die Elemente der Menge aufgeschrieben werden können.

Definition 4.1. Eine **Permutation** einer n -Menge A ist eine Bijektion

$$\pi : \{1, 2, \dots, n\} \rightarrow \{1, \dots, n\}$$

Eine Permutation kann als geordnete Aufzählung der Elemente $a_i \in A$ in der Reihenfolge $a_{\pi(1)}, \dots, a_{\pi(n)}$ betrachtet werden. Häufig werden wir statt der Funktion π auch einfach die Aufzählung der Elemente angeben.

Beispiel. Für $A = \{a, b, c\}$ kann die Reihenfolge cba als Permutation π_1 mit

n	$\pi_1(n)$
1	3
2	2
3	1

dargestellt werden.

Alle möglichen Reihenfolgen für die Elemente sind: $abc, acb, bac, bca, cab, cba$. ◀

Selbsttest: Was sind alle möglichen Permutationen der Menge $\{a, b, c, d\}$?



Satz 4.4. Es gibt $n!$ viele Permutationen einer n -Menge A .

Beweis. Wir zeigen die Aussage mittels vollständiger Induktion nach n .

I.A. Für $n = 1$ ist die Aussage klar.

I.V. Für ein festes n gibt es $n!$ Permutationen jeder n -Menge

I.S. $n \rightarrow n+1$ Sei $A = \{a_1, \dots, a_{n+1}\}$ eine $n+1$ -elementige Menge und sei $A' = A \setminus \{a_{n+1}\}$ eine n -elementige Menge. Nach Induktionsvoraussetzung gibt es $n!$ Permutationen von A' .

Wir bilden nun $n+1$ paarweise disjunkte Mengen $\Pi(A_i)$ mit $1 \leq i \leq n+1$ die jeweils Permutationen von A enthalten. Die Menge $\Pi(A_i)$ entsteht, indem a_{n+1} für jede Permutation aus A' an der i -ten Stelle eingeschoben wird. Diese Beschreibung von $\Pi(A_i)$ gibt direkt eine Bijektion zwischen A' und jedem der $\Pi(A_i)$, also gilt $|\Pi(A_i)| = n!$ für alle $1 \leq i \leq n+1$.

Da die Menge aller Permutationen von A die disjunkte Vereinigung aller $\Pi(A_i)$ ist, gibt es nach der Summenregel $(n+1) \cdot n! = (n+1)!$ Permutationen von A . \square

Selbsttest: Wie viele Möglichkeiten gibt es 5 verschiedene Kuchenstücke auf 5 Personen zu verteilen?



4.3 Binomialkoeffizienten

Im Folgenden betrachten wir den sogenannten Binomialkoeffizienten und dessen Eigenschaften.

Definition 4.2. Der **Binomialkoeffizient** $\binom{n}{k}$ ist definiert als die Anzahl der k -elementigen Teilmengen einer n -Menge M .

Diese Definition ist zwar eine genaue Beschreibung des Binomialkoeffizienten, hilft uns aber nicht dabei, konkrete Werte auszurechnen. Bevor wir uns verschiedene konkretere Formeln für die Berechnung anschauen, stellen wir zunächst das Folgende fest:

Beobachtung 4.5. Es gilt:

- $\sum_{i=0}^n \binom{n}{i} = 2^n$
- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = 0$ für $k > n$

Nun geben wir eine direkte Formel für $\binom{n}{k}$ an und Beweisen diese kombinatorisch.

Satz 4.6. Für $n \geq k$ gilt:

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

Beweis. Wir führen einen kombinatorischen Beweis für die Anzahl der k -Teilmengen einer n -Menge. Zunächst einmal betrachten wir eine Folge (also feste Reihenfolge) der k ausgewählten Elemente für die Teilmenge. Mit der Produktregel sind dies $n \cdot (n-1) \cdot \dots \cdot (n-k+1)$ Möglichkeiten. Diese Formel können wir auch als $\frac{n!}{(n-k)!}$ schreiben.

Es gibt also $\frac{n!}{(n-k)!}$ Möglichkeiten eine feste Reihenfolge von k Elementen zu wählen. Jede mögliche Teilmenge ist nun aber mehrfach in diesen festen Reihenfolgen enthalten. Genauer ist jede der $k!$ möglichen Permutationen jeder Auswahl genau einmal enthalten. Um also die Anzahl der k -Teilmengen zu erhalten, muss dieser Wert noch durch $k!$ geteilt werden und es ergibt sich:

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \quad \square$$

Definition 4.3. Eine Permutation der Elemente einer k -elementigen Teilmenge einer n -Menge von M heißt k -Permutation von M . Die Anzahl dieser Permutationen wird **fallende Faktorielle** genannt und mit $n^{\underline{k}}$ bezeichnet.

Korollar 4.7. Es gilt $n^{\underline{k}} = \frac{n!}{(n-k)!}$.

Beweis. Ergibt sich aus dem Beweis von Satz 4.6. □

Beispiel. Betrachte die 4-Menge $M = \{a, b, c, d\}$. Dann ist die Menge der k -Permutationen

$$\{ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc\}$$

mit $\frac{4!}{2!} = 12$ Elementen. Die Menge der 2-elementigen Teilmengen ist

$$\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}$$

mit $\frac{4!}{2! \cdot 2!} = 6$ Elementen. ◀

Wir betrachten nun weitere Beispiele, die Binomialkoeffizienten und die anderen Regeln mischen.

Beispiel. 1. Lotto 6 aus 49. Wie viele Möglichkeiten für eine Wette beim Lotto gibt es? Jede Ziehung ergibt eine k -Permutation mit $49^{\underline{6}}$ Möglichkeiten. Für das Bestimmen der Gewinnzahlen ist die Reihenfolge egal, also ist die Anzahl der Möglichkeiten

$$\frac{49^{\underline{6}}}{6!} = \binom{49}{6} = 13983816$$

2. Es gibt $\binom{n}{k}$ Bitstrings der Länge n mit genau k Einsen. Wir betrachten die Menge aller Stellen der Zahl als Grundmenge und wählen daraus die k Stellen aus, an denen eine Eins steht. Es gibt $\binom{n}{k}$ Möglichkeiten die Stellen für die Einsen auszuwählen und die Aussage folgt.

- Es gibt $\binom{n}{k} \cdot 2^{n-k}$ Strings der Länge n über dem Alphabet¹ $\{0, 1, 2\}$ mit genau k Einsen. Zunächst werden die Positionen für die Einsen ausgewählt, das sind $\binom{n}{k}$ Stück. Die restlichen Stellen können frei aus 0 und 2 gewählt werden. Das kann bijektiv auf Binärstrings der Länge $n-k$ abgebildet werden. Mit der Produktregel folgt die Formel.
- Es gibt $\binom{n}{k} \cdot \binom{n-k}{l}$ Strings der Länge n über $\{0, 1, 2\}$ mit genau k Einsen und l Zweien. Zunächst werden die Positionen der Einsen gewählt, das sind $\binom{n}{k}$ Stück. Danach werden aus den restlichen Positionen die Zweien gewählt, dafür gibt es noch $\binom{n-k}{l}$ Möglichkeiten. Die Produktregel gibt wieder das Ergebnis. ◀

Selbsttest: Wie viele Möglichkeiten gibt es, drei verschiedene Sorten Kuchen und vier verschiedene Sorten Kekse auszuwählen?



Der Binomialkoeffizient kann auch rekursiv definiert werden.

Satz 4.8. Es gilt $\binom{n}{0} = 1$, $\binom{n}{k} = 0$ für $k > n$ und

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

Beweis. Sei $a \in M$ ein Element. Wir unterteilen die k -elementigen Teilmengen von M in zwei disjunkte Mengen, um dann die Summenformel anwenden zu können.

Die erste Menge enthält alle Untermengen, in denen a nicht enthalten sind. Um diese Mengen zu bilden, werden k Elemente aus den $n-1$ Elementen von $M \setminus \{a\}$ ausgewählt. Dafür gibt es $\binom{n-1}{k}$ Möglichkeiten.

Die zweite Menge enthält alle Untermengen, in denen a enthalten ist. Damit ist ein Element der Menge schon fest. Die übrigen $k-1$ Elemente können aus $M \setminus \{a\}$ gewählt werden. Dafür gibt es dann $\binom{n-1}{k-1}$ Möglichkeiten. Die Formel ergibt sich dann aus der Summenregel. ◻

Wir betrachten nun stichpunktartig weitere Identitäten. In den meisten Fällen geben wir kombinatorische Beweise an. In vielen Fällen könnte der Beweis auch über die Formel aus Satz 4.6 oder mittels vollständiger Induktion geführt werden.

- $\binom{n}{k} = \binom{n}{n-k}$ für $0 \leq k \leq n$. Diese Symmetrie folgt entweder direkt aus der Formel oder auch kombinatorisch. Alle k -elementigen Teilmengen können bijektiv auf die $n-k$ -elementigen Teilmengen abgebildet werden. Statt die k Elemente auszuwählen, die in die Teilmenge kommen, werden die $n-k$ Elemente ausgewählt, die nicht in der Menge enthalten sind.
- Eine Möglichkeit die Binomialkoeffizienten darzustellen ist eine Tabelle mit n in den Reihen und k in den Spalten. Diese Darstellung wird auch das **Pascalsche Dreieck** genannt.

¹Also die Buchstaben, aus denen das Wort zusammengesetzt werden kann.

$n \backslash k$	0	1	2	3	4	5	6
0	1	0					
1	1	1	0				
2	1	2	1	0			
3	1	3	3	1	0		
4	1	4	6	4	1	0	
5	1	5	10	10	5	1	0
6	1	6	15	20	15	6	1

3. Die *Vandermond'sche Identität* besagt:

$$\binom{n+m}{k} = \sum_{i=0}^k \binom{n}{i} \cdot \binom{m}{k-i}$$

Beweis. Wir zeigen die Gleichung kombinatorisch. Sei A eine $(n+m)$ -Menge und seien B und C disjunkte Teilmengen von A mit $|B| = n$ und $|C| = m$.

Die linke Seite der Vandermond'schen Identität besagt, dass wir eine k -Teilmenge von A wählen wollen. Jede solche Teilmenge besteht aus i Elementen aus B und $k-i$ Elementen aus C für alle $0 \leq i \leq k$. Nach der Produktregel gibt es $\binom{n}{i} \cdot \binom{m}{k-i}$ Möglichkeiten diese Mengen auszuwählen und die k -Teilmengen von A sind disjunkt für verschiedene i . Nach der Summenregel können nun die Ausdrücke für jedes i aufsummiert werden und erhalten die rechte Seite der Gleichung. \square

4. Es gilt der *binomische Satz*

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} \cdot x^{n-k} \cdot y^k$$

Beweis. Bevor wir uns einige bekannte Spezialfälle anschauen, beweisen wir zunächst die Aussage. Wir zählen wie oft jeder Term $x^{n-k}y^k$ beim Ausmultiplizieren von $(x+y)^n$ entsteht. Dies ist genau die Anzahl der Möglichkeiten aus k der Klammern das y und aus den restlichen $n-k$ Klammern das x zu wählen. Also $\binom{n}{k}$ Möglichkeiten für $0 \leq k \leq n$. \square

Einige Spezialfälle kennen wir schon:

$$2^n = (1+1)^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} \cdot 1^k = \sum_{k=0}^n \binom{n}{k}$$

$$(a+b)^2 = \sum_{k=0}^2 a^{2-k} b^k = a^2 + 2ab + b^2$$

$$0 = (1-1)^n = \sum_{k=0}^n (-1)^k \binom{n}{k}$$

5. Verallgemeinerte Rekursion. Für die Spaltensummen des Pascalschen Dreiecks gilt:

$$\binom{n+1}{k+1} = \sum_{i=0}^n \binom{i}{k}$$

Beweis. Wir betrachten eine $(n+1)$ Menge $L = \{0, 1, \dots, n\}$ an Losen. Dann beschreibt $\binom{n+1}{k+1}$ die Anzahl der Möglichkeiten $k+1$ Lose zu ziehen. Sei Z die Menge aller Teilmengen von Losen. Wir unterteilen Z nun in $n+1$ disjunkte Teilmengen Z_i . Dabei beschreibt Z_i alle Möglichkeiten $k+1$ Lose zu ziehen, wobei das größte Los das Los mit Nummer i ist. Es gilt $|Z_i| = \binom{i}{k}$, da ein Los fix ist² und die restlichen Lose aus $\{0, \dots, i-1\}$ ergänzt werden.

Es gilt nun $Z = \bigcup_{i=0}^n Z_i$ und da alle Z_i disjunkt sind folgt die Aussage mit der Summenregel. \square

Anmerkung: Da $\binom{i}{k} = 0$ für $i < k$ kann die verallgemeinerte Rekursion auch als $\binom{n+1}{k+1} = \sum_{i=k}^n \binom{i}{k}$ geschrieben werden.



6. Newton-Identität. Für $0 \leq l \leq k \leq n$ gilt:

$$\binom{n}{k} \cdot \binom{k}{l} = \binom{n}{l} \cdot \binom{n-l}{k-l}$$

Beweis. Auch diese Aussage zeigen wir kombinatorisch. Ein Beweis über die Formel ist zwar theoretisch möglich aber schwer zu rechnen.

Wir finden zunächst eine kombinatorische Interpretation der linken Seite. Der erste Faktor gibt die Anzahl aus einer n -Menge eine k -Teilmenge auszuwählen. Aus dieser k -Teilmenge werden dann weitere l Elemente ausgewählt. Also gibt die Linke Seite die Möglichkeit an, l Elemente aus einer k Teilmenge einer n -Menge auszuwählen.

Auf der rechten Seite werden zunächst l Elemente aus der n -Menge gewählt. Anschließend werden noch $k-l$ Elemente aus $n-l$ Elementen gewählt. Es wird also zunächst festgelegt, welche l Elemente aus der k Menge gewählt werden und anschließend werden die restlichen Elemente gewählt, um die k -Menge aufzufüllen. \square

4.3.1 Monotone Gitterwege

Anmerkung: Dieser Abschnitt wird in der Vorlesung aktuell meistens nicht behandelt. Er gibt eine schöne Anwendung der Binomialkoeffizient und kann bei Interesse gelesen werden.



²und zwar das Los mit Nummer i

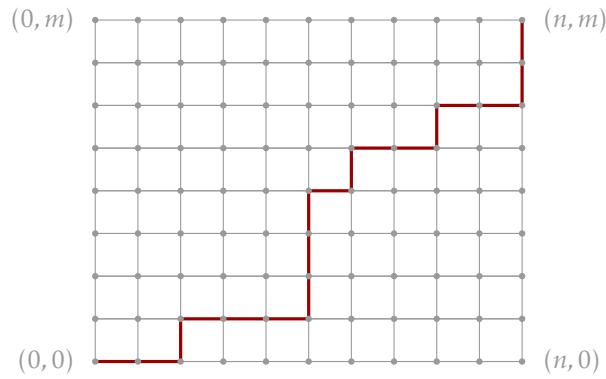


Abbildung 4.3: Beispiel für ein $n \times m$ Gitter mit einem möglichen monotone Gitterweg

In diesem Abschnitt betrachten wir eine geometrische Interpretation der Binomialkoeffizienten. Bei dieser werden wir auch einige der eben betrachteten Identitäten wieder sehen.

Definition 4.4. Ein monotoner Weg in einem ganzzahligen Gitter der Form $\{0, 1, \dots, n\} \times \{0, 1, \dots, m\}$ ist ein Weg von $(0, 0)$ zum Knoten (n, m) , der nur Schritte nach rechts oder oben macht, siehe Abbildung 4.3.

Beobachtung 4.9. Es gelten die folgenden Aussagen:

1. Jeder monotone Gitterweg von $(0, 0)$ zum Knoten (n, m) enthält $n + m$ Kanten, davon sind n waagerecht und m senkrecht.
2. Ein solcher Weg ist eindeutig dadurch bestimmt, bei welchen Schritten nach rechts gegangen wird.
3. Es gibt $\binom{n+m}{n}$ verschiedene Wege.

Statt die Schritte nach rechts festzulegen, können auch die Schritte nach oben festgelegt werden. Die Anzahl der Wege bleibt gleich, daraus folgt direkt die Symmetrie:

$$\binom{n+m}{n} = \binom{n+m}{m}$$

Die rekursive Definition des Binomialkoeffizienten kann auch durch die Betrachtung der monotonen Gitterwege gesehen werden. Hierzu werden die Gitterwege disjunkt unterteilt in alle Gitterwege, die mit einem Schritt nach rechts beginnen und alle Gitterwege, die mit einem Schritt nach oben beginnen. Es gibt $\binom{n+m-1}{n-1}$ Gitterwege, die mit einem Schritt nach rechts beginnen und $\binom{n+m-1}{n}$ Gitterwege, die mit einem Schritt

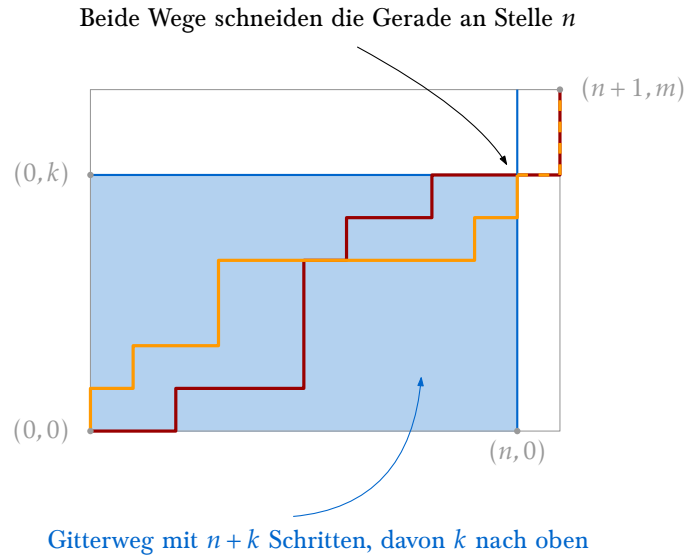


Abbildung 4.4: Verallgemeinerte Rekursion in Gitterwegen

nach oben beginnen. Damit ergibt sich nach der Summenregel:

$$\binom{n+m}{n} = \binom{n+m-1}{n} + \binom{n+m-1}{n-1}$$

Auch die verallgemeinerte Rekursionsgleichung kann im Kontext von monotonen Gitterwegen gezeigt werden. Betrachte die monotonen Gitterwege von $(0,0)$ nach $(n+1,m)$. Jeder Weg muss an einer Stelle die vertikale Gerade durch $(n,0)$ kreuzen. Das ist genau der Zeitpunkt, an dem der letzte Schritt nach rechts gegangen wird. Wenn dieser Schritt auf Höhe k geschieht, ist der Weg bis zu diesem Punkt ein monotoner Gitterweg von $(0,0)$ nach (n,k) . Von diesen gibt es $\binom{n+k}{k}$ viele. Danach gibt es nur noch einen Schritt nach rechts und $m-k$ Schritte nach oben, diese sind dann schon festgelegt, siehe Abbildung 4.4.

Also gilt die folgende Gleichung.

$$\sum_{k=0}^m \binom{n+k}{k} = \binom{m+n+1}{m}$$

Auch die Vandermond'sche Identität kann mithilfe von monotonen Gitterwegen dargestellt werden, diesmal betrachten wir die Wege in einem $(n+m) \times k$ Gitter. Dazu werden die Wege unterteilt in alle Wege, die in den ersten $n+k$ Schritten genau i Schritte nach oben machen, siehe Abbildung 4.5. Für ein solches i gibt es $\binom{n+k}{i} \cdot \binom{m}{k-i}$ solche Wege. Über alle Wert i summiert ergibt sich nach der Summenregel:

$$\binom{n+k+m}{k} = \sum_{i=0}^k \binom{n+k}{i} \cdot \binom{m}{k-i}$$

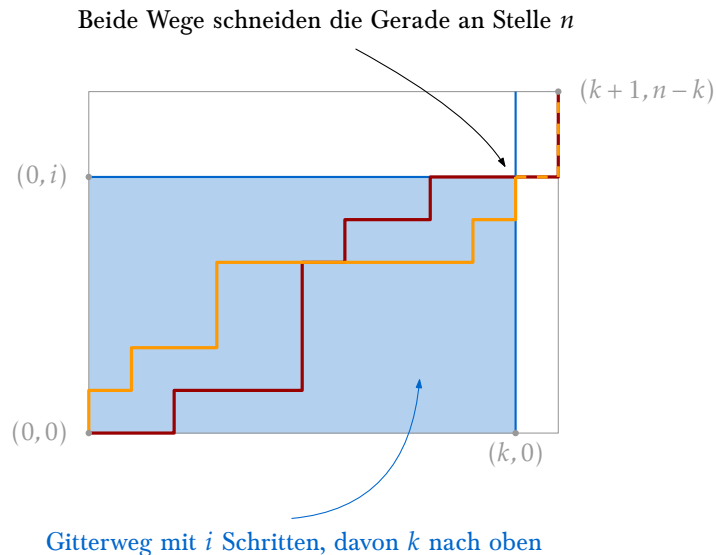


Abbildung 4.5: Vandermonde'sche Identität in Gitterwegen

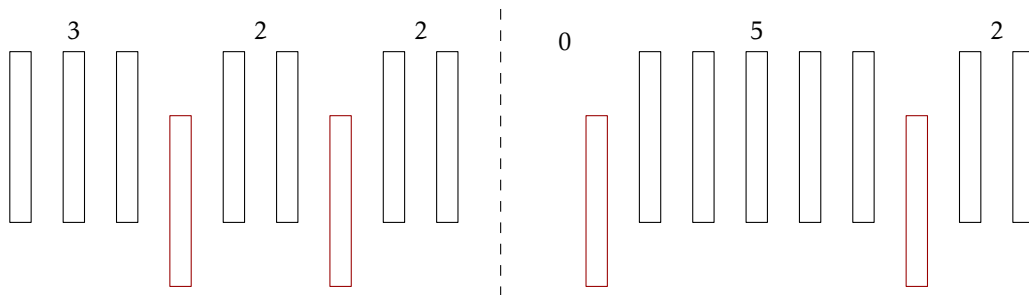


Abbildung 4.6: Zwei mögliche Verteilungen von Kreidestücke auf die Dozierenden

4.3.2 Zahlpartitionen und geordnete Summendarstellungen

Eine weitere Anwendung von Binomialkoeffizienten sind sogenannte *Zahlpartitionen*. Als Hinführung zu diesen, schauen wir uns zunächst *geordnete k -Summendarstellungen von n* an, bezeichnet mit $S_{n,k}$.

Beispiel. Ein motivierendes Beispiel für eine geordnete Summendarstellung ist das Folgende. Gegeben seien 7 Stücke Kreide, die auf 3 Dozierende aufgeteilt werden sollen, wobei einzelne Dozierende auch keine Kreide abbekommen können. Wie viele Möglichkeiten die Kreide aufzuteilen gibt es?

Anders gefragt, wie viele Zerlegungen n_1, n_2, n_3 mit $7 = n_1 + n_2 + n_3$ und $n_i \geq 0$ gibt es? Abbildung 4.6 zeigt zwei mögliche Verteilungen. Das Bild gibt auch direkt schon einen Hinweis auf die Lösung. Es werden zwei „Dummy“ Kreidestücke hinzugefügt und dann aus den 9 neuen Stücken 2 als Trenner ausgewählt. Damit gibt es $\binom{9}{2} = 36$ verschiedenen Verteilungen.

Max:
Verein-
heitli-
chung
der Na-
mens-
ge-
bung?

Satz 4.10. Es gibt $\binom{n+k-1}{k-1}$ viele geordnete k -Summendarstellungen der Zahl $n = n_1 + n_2 + \dots + n_k$ mit $n_i \geq 0$.

Beweis. Der Beweis folgt aus der Verallgemeinerung des Beispiels. \square

Beobachtung 4.11.

1. Wird das Problem auf $n \geq n_1 + n_2 + \dots + n_k$ mit $n_i \geq 0$ erweitert, gibt es $\binom{n+k}{k}$ geordnete Summendarstellungen.
2. Gibt es zusätzliche Nebenbedingungen, wie $n_1 \geq m \geq 1$, dann gibt es $\binom{n+k-1-m}{k-1}$ Möglichkeiten.

Selbsttest: Warum gilt Beobachtung 4.11?



Wenn alle Summanden in der Darstellung ≥ 1 sind, spricht man von einer **Zahlpartition**.

Definition 4.5. Eine **ungeordnete k -Partition** einer ganzen Zahl $n > 0$ ist eine (Multi)-Menge^a von positiven ganzen Zahlen $\{n_1, n_2, \dots, n_k\}$ mit $\sum_{i=1}^k n_i = n$. Diese Anzahl wird auch als $P_{n,k}$ bezeichnet.^b

Eine **geordnete k -Partition** einer ganzen Zahl $n > 0$ ist eine Folge von positiven ganzen Zahlen n_1, n_2, \dots, n_k mit $\sum_{i=1}^k n_i = n$. Diese Anzahl wird auch als $GP_{n,k}$ bezeichnet.

^aAlso eine Menge, die ein Element mehrfach enthalten kann

^bDiese werden wir in der Vorlesung nicht weiter behandeln

Mit ungeordneten Zahlpartitionen werden wir uns im Weiteren nicht beschäftigen. Die Definition ist hier nur als Abgrenzung zu den geordneten Zahlpartitionen gegeben.

Satz 4.12. Es gibt $\binom{n-1}{k-1}$ viele geordnete k -Partitionen einer ganzen Zahl $n > 0$.

Beweis. Anders als im Beweis zu Satz 4.10 ist die 0 nicht als Summand erlaubt, daher wählen wir einen leicht anderen Ansatz. Wir schreiben n als Folge von n Einsen mit $n - 1$ Pluszeichen dazwischen:

$$n = \underbrace{1 + 1 + 1 + 1 + \dots + 1}_{n \text{ mal}}$$

Um nun k positive Summanden zu erhalten, können $k - 1$ Pluszeichen ausgewählt werden. Dadurch wird eine geordnete k -Partition erzeugt. Umgekehrt kann jede geordnete k -Partition auf eine Auswahl von $k - 1$ Pluszeichen abgebildet werden. \square

Selbsttest: Satz 4.12 kann auch mit Hilfe von Beobachtung 4.11 gezeigt werden. Wie?



Beispiel.

1. Wie viele Möglichkeiten gibt es, 10 nicht unterscheidbare Bälle in 8 unterscheidbare Boxen zu verteilen?

Die Aufgabe ist genau eine k -Summendarstellung mit $n = 10$ und $k = 8$. Also gibt es $\binom{10+8-1}{8-1} = \binom{17}{7}$ Möglichkeiten.

2. Wie viele Möglichkeiten gibt es, 10 nicht unterscheidbare Bälle in 8 unterscheidbare Boxen zu verteilen, sodass in jeder Box mindestens ein Ball landet?

Dies ist genau das Problem einer geordneten k -Partition, es gibt also $\binom{10-1}{8-1} = \binom{9}{7}$ viele Möglichkeiten.



4.4 Doppeltes Abzählen

In diesem Abschnitt betrachten wir eine Möglichkeit, die Anzahl der Elemente in einer Relation auf zwei verschiedene Arten abzuzählen. Die grundlegende Idee ist, ähnlich wie beim Taubenschlagprinzip, sehr naheliegend, das Konzept kann aber an vielen Stellen verwendet werden.

Definition 4.6. Ein **Inzidenzsystem** (S, T, I) besteht aus zwei (endlichen) Mengen S und T und einer Inzidenzrelation $I \subseteq S \times T$.

Dabei ist die Inzidenzrelation eigentlich eine ganz normale Relation. Wir fassen hier nur die beiden Mengen, auf denen die Relation definiert ist, noch mit in das betrachtete mathematische Objekt dazu und geben dem ganzen einen neuen Namen.

Satz 4.13 (Regel vom zweifachen Abzählen). Sei (S, T, I) ein Inzidenzsystem und sei

$$r(a) = |\{b \in T \mid (a, b) \in I\}|$$

$$r(b) = |\{a \in S \mid (a, b) \in I\}|$$

für alle $a \in S$ und $b \in T$, dann gilt:

$$\sum_{a \in S} r(a) = \sum_{b \in T} r(b)$$

Beweis. Um zu sehen, warum die Aussage gilt, machen wir uns klar, was wir genau zeigen wollen. Die linke Seite der Gleichung, ist eine Summe über alle Elemente $a \in S$, in der jeweils die Elemente der Relation I betrachtet werden, in denen a an erste Stelle steht. Also ist das Ergebnis der Summe auf der linken Seite genau die Anzahl der

Elemente in I . Ähnlich werden auf der rechten Seite über alle Elemente $b \in T$, bei der in jedem Schritt die Elemente aus I betrachtet werden, bei denen b an zweiter Stelle steht. Nach der Summenregel werden hier auch wieder alle Elemente aus I gezählt und damit sind beide Summen gleich. \square

Eine andere Darstellung der Regel ergibt sich, wenn wir eine $m \times n$ Matrix M anschauen. Eine $m \times n$ Matrix kann man sich als eine Tabelle mit m Zeilen und n Spalten vorstellen. Das Element in Zeile i und Spalte j bezeichnen wir mit m_{ij} .

Ist nun $z_i = \sum_{j=1}^n m_{ij}$ für $1 \leq i \leq m$ die Summe der Elemente in Zeile i und $s_j = \sum_{i=1}^m m_{ij}$ für $1 \leq j \leq n$ die Summe der Elemente in Spalte j gilt:

$$\sum_{i=1}^m z_i = \sum_{j=1}^n s_j \quad (4.3)$$

Die Gleichung 4.3 ergibt sich daraus, dass auf beiden Seiten wieder alle Elemente der Matrix aufsummiert werden. Dabei ist es egal, ob erst die Zeilen oder erst die Spalten summiert werden.

Satz 4.13 ergibt sich dann als Sonderfall. Die Elemente aus S und T werden nummeriert als $A = \{a_1, \dots, a_m\}$ und $T = \{b_1, \dots, b_n\}$. Dann kann eine Matrix definiert werden als:

$$m_{ij} = \begin{cases} 1 & \text{falls } (a_i, b_j) \in I \\ 0 & \text{sonst} \end{cases}$$

und Satz 4.13 ergibt sich direkt.

Beispiel. Sei $S = T = \{1, 2, \dots, n\}$ und sei I die Teilbarkeitsrelation mit $(a, b) \in I$ genau dann wenn $a \mid b$.

Dann ergibt sich als Inzidenzmatrix für $n = 8$ die folgende Matrix:

	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	0	1	0	1	0	1	0	1
3	0	0	1	0	0	1	0	0
4	0	0	0	1	0	0	0	1
5	0	0	0	0	1	0	0	0
6	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	1

Sei $t(j)$ die Anzahl der Teiler von j . Das ist dann genau die Summe der Einträge in Spalte j .

Angenommen uns interessiert die durchschnittliche Anzahl $t^*(n)$ von Teilern einer Zahl $\leq n$? Um diese auszurechnen, verwenden wir doppeltes Abzählen, sowie die Beobachtung, dass in Zeile i jeder i -te Eintrag 1 ist.

$$t^*(n) = \frac{1}{n} \sum_{j=1}^n t(j)$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n \left\lfloor \frac{n}{i} \right\rfloor \\
 &\approx \sum_{i=1}^n \frac{1}{i} \\
 &= H_n \approx \ln n
 \end{aligned}$$

Dabei gibt der Schritt von der vorletzten zur letzten Zeile die Definition der Zahl H_n . $H_n = \sum_{i=1}^n \frac{1}{i}$ wird auch als ***n-te Harmonische Zahl*** bezeichnet. ◀

4.5 Zusammenfassung und eine Anwendung

Wie haben die folgenden Sorten des Abzählens kennengelernt:

Untermengen einer n -Menge:	2^n
k -Untermengen einer n -Menge:	$\binom{n}{k}$
Permutationen einer n -Menge:	$n!$
geordnete k -Teilmengen einer n -Menge:	$n^k = n \cdot (n-1) \cdot \dots \cdot (n-k+1)$
geordnete k -Zahlpartitionen der Zahl n :	$\binom{n-1}{k-1}$
Strings mit Länge n über einem Alphabet mit k Zeichen:	k^n
geordnete Summendarstellungen von n mit k Summanden ≥ 0 :	$\binom{n+k-1}{k-1}$

Einige der Formeln können auch mit dem folgenden Modell zusammengefasst werden. m Bälle B werden in ℓ Kisten K geworfen. Die Zuordnung von Bällen zu Kisten kann als eine Funktion $f : B \rightarrow K$ betrachtet werden. Je nachdem ob die Bälle unterscheidbar sind und welche Eigenschaften die Funktion hat, ergeben sich verschiedene Fälle:

Bälle	f beliebig	f injektiv	f surjektiv	f bijektiv
unterscheidbar	m^ℓ	$m^{\underline{\ell}}$	-	$n!$ oder 0
nicht unterscheidbar	$\binom{m+\ell-1}{\ell-1}$	$\binom{\ell}{m}$	$\binom{m-1}{\ell-1}$	1 oder 0

4.5.1 Ein Kartentrick

Anmerkung: Dieser Abschnitt ist Zusatzmaterial und wird in der Vorlesung im Allgemeinen nicht behandelt. 💡

Zum Abschluss dieses Kapitels schauen wir uns noch einen Kartentrick an, welcher kombinatorische Argumente mit dem Taubenschlagprinzip verbindet. Die Beteiligten beim Kartentrick, sind das Publikum (P), ein Assistent (A) und eine Magierin (M). A zieht fünf Karten aus einem Kartendeck mit 52 Karten und zeigt M vier dieser fünf Karten. Daraufhin kann M immer sagen, welche die fünfte Karte war.

Satz 4.14. Werden fünf Karten gezogen, ist es möglich durch das Zeigen von vier Karten in einer gegebenen Reihenfolge die fünfte Karte eindeutig zu codieren.

Beweis. Um die fünfte Karte eindeutig identifizieren zu können, muss sowohl die Farbe, also auch der Wert der Karte kommuniziert werden. Die Farbe ist einfach zu codieren. Nach dem Taubenschlagprinzip gibt es mindestens eine Farbe, die mindestens zwei mal gezogen wird. Eine der Karten mit doppelter Farbe wird als erste Karte gezeigt, welche das ist, ergibt sich daraus, welche Werte die Karten haben.

Um den Wert zu kommunizieren, werden die Werte auf einem Kreis angeordnet. Nach dem Taubenschlagprinzip gibt es eine der doppelten Karte mit doppelter Farbe, von der aus maximal 6 Schritte im Uhrzeigersinn gelaufen werden müssen, um zur zweiten Karte zu kommen. Diese Karte wird dann als erste Karte gezeigt. Die restlichen drei Karten codieren dann wie viele Schritte im Uhrzeigersinn gelaufen werden müssen.

Dazu wird eine totale Ordnung auf allen 52 Karten definiert. Sei k die kleinste, m die mittlere und g die größte der drei Karten. Dann gibt es $3! = 6$ mögliche Reihenfolgen, in denen die Karten gezeigt werden. Jeder der Reihenfolgen kann eine Zahl zwischen 1 und 6 bijektiv zugewiesen werden, welche dann die Schritte im Uhrzeigersinn darstellt. \square

Beobachtung 4.15. *Werden vier Karten gezogen ist es nicht möglich durch das zeigen von drei Karten die vierte eindeutig zu identifizieren.*

Beweis. Es gibt $\binom{52}{4}$ mögliche Mengen an Karten, die gezogen werden können. Es gibt aber nur $52 \cdot 51 \cdot 50$ mögliche Folgen mit 3 Karten. Da $\binom{52}{4} > 52 \cdot 51 \cdot 50$, können nicht alle möglichen Mengen eindeutig codiert werden. \square

Diskrete Wahrscheinlichkeitstheorie

Voraussetzung:

- Mengen, Relationen und Funktionen (Kapitel 3)
- Kombinatorik (Kapitel 4)



Lernziele

Die Studierenden ...

- ... bestimmen Wahrscheinlichkeiten von Ergebnissen
- ... berechnen Erwartungswerte
- ... zeigen Aussagen im Kontext der Wahrscheinlichkeitsrechnung



Dieses Kapitel beschäftigt sich mit den Grundlagen der diskreten Wahrscheinlichkeitstheorie. Diese wird in der Informatik in vielen verschiedenen Kontexten benötigt und ist daher in verschiedenen Bereichen wichtig. Ein großes Feld ist die Analyse des Verhaltens von Algorithmen bei einer zufälligen Eingabe. Häufig wird für die Analyse der Zeit, die ein Algorithmus für ein bestimmtes Problem braucht die schlecht mögliche Eingabe betrachtet. Dies ist jedoch häufig nicht repräsentativ für das Verhalten des Algorithmus bei Eingaben aus der echten Welt. Daher wird für einige Algorithmen auch oft die zu erwartende Laufzeit bei einer zufälligen Eingabe betrachtet.

Eine weitere Anwendung des Zufalls ist der Entwurf von Algorithmen. So gibt es Algorithmen in denen Entscheidungen mithilfe des Zufalls getroffen werden. Bei diesen Algorithmen hängt dann die Laufzeit oder die Korrektheit vom Zufall ab. Es wird also nur die erwartete Laufzeit bestimmt, oder es kann sein, dass der Algorithmus mit einer bestimmten Wahrscheinlichkeit ein falsches Ergebnis ausgibt.

In weiteren Kontexten wird Zufall für die Modellierung und Analyse von komplexen Zusammenhängen verwendet. Ein weiteres großes Feld in der Wahrscheinlichkeitsrechnung verwendet wird, ist die künstliche Intelligenz.

5.1 Grundlegende Begriffe und Beispiele

Wir betrachten nun zunächst einige grundlegende Begriffe und Beispiele zu diesen.

Definition 5.1. Ein **diskreter Wahrscheinlichkeitsraum** ist ein Paar (Ω, Pr) , wobei Ω eine endliche (oder abzählbar unendliche) Menge von **Elementarereignissen** und $\text{Pr} : \Omega \rightarrow [0, 1]$ eine Wahrscheinlichkeitsverteilung ist^a. Pr ordnet jedem $\omega \in \Omega$ seine **Wahrscheinlichkeit** zu.

Zusätzlich gilt als Normierung, dass $\sum_{\omega \in \Omega} \text{Pr}(\omega) = 1$.

^a $[0, 1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$

Definition 5.2. Ist $\text{Pr}(\omega) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$, wird Pr eine Gleichverteilung genannt.

Definition 5.3. Eine Teilmenge $A \subseteq \Omega$ von Ω wird Ereignis genannt und es gilt $\text{Pr}(A) = \sum_{\omega \in A} \text{Pr}(\omega)$.

Nun betrachten wir einige Beispiele zu diesen Definitionen.

Beispiel.

1. Würfeln eines fairen 6-seitigen Würfels. Es gilt $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $\text{Pr}(1) = \text{Pr}(2) = \text{Pr}(3) = \text{Pr}(4) = \text{Pr}(5) = \text{Pr}(6) = \frac{1}{6}$, also beschreibt Pr eine Gleichverteilung. Wir betrachten nun das Ereignis eine gerade Zahl zu würfeln. Es gilt also $A = \{2, 4, 6\}$. Die Wahrscheinlichkeit von A ist $\text{Pr}(A) = \text{Pr}(2) + \text{Pr}(4) + \text{Pr}(6) = \frac{1}{2}$.
2. Wurf einer fairen Münze. Es gilt $\Omega = \{k, z\}$, $\text{Pr}(k) = \text{Pr}(z) = \frac{1}{2}$, also haben wir wieder eine Gleichverteilung.
3. Zwei faire Münzwürfe hintereinander (oder zwei unterscheidbare Münzen). Es gilt $\Omega = \{(k, k), (k, z), (z, k), (z, z)\}$ mit Gleichverteilung, also $\text{Pr}(i, j) = \frac{1}{4}$.
4. Zwei gleiche Münzen gleichzeitig. Es gilt $\Omega = \{\{k, k\}, \{z, k\}, \{z, z\}\}$ mit $\text{Pr}(\{k, k\}) = \text{Pr}(\{z, z\}) = \frac{1}{4}$ und $\text{Pr}(\{z, k\}) = \frac{1}{2}$. ◀

Selbsttest: Betrachten Sie die folgenden Zufallsexperimente und geben Sie die Wahrscheinlichkeitsräume an. Denken Sie daran, dass ein Wahrscheinlichkeitsraum immer aus dem Tupel aus Elementarereignissen und der Wahrscheinlichkeitsverteilung besteht.

- a) Werfen von vier fairen Münzen nacheinander.

¹In diesem Beispiel sind wir etwas ungenau mit der Mengennotation. Die Menge $\{z, z\}$ so wie wir sie in Abschnitt 3.1 kennengelernt haben, enthält nur ein Element. Hier verwenden wir $\{z, z\}$ um das Ereignis, dass zweimal Kopf geworfen wird darzustellen.

- b) Würfeln einer gezinkten Münze, bei der Kopf doppelt so wahrscheinlich ist wie Zahl.
- c) Würfeln eines nicht fairen 6-seitigen Würfels, bei dem gerade Zahlen doppelt so wahrscheinlich sind wie ungerade Zahlen.
- d) Würfeln von zwei fairen 6-seitigen Würfeln gleichzeitig.
- e) Würfeln eines fairen 6-seitigen Würfels und eines fairen 4-seitigen Würfels gleichzeitig.

Es gelten die folgenden Fakten, die alle aus den Identitäten für Mengen und Definition 5.3 folgen.

Lemma 5.1. *Es gilt:*

- $\Pr(\Omega) = 1$ und $\Pr(\emptyset) = 0$
- $A \subseteq B \rightarrow \Pr(A) \leq \Pr(B)$
- $\Pr(A) = \sum_{i=1}^k \Pr(A_i)$ für jede Partition $A = \bigcup_{i=1}^k A_i$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(\Omega \setminus A) = 1 - \Pr(A)$
- $\Pr(\bigcup_{i=1}^k A_i) \leq \sum_{i=1}^k \Pr(A_i)$.

Beweis. Wir betrachten den Beweis für den ersten Fakt. Die weiteren folgen dann als Übung. Für $\Pr(\Omega)$ betrachten wir die Wahrscheinlichkeit eines Ereignisses. Mit Definition 5.3 gilt $\Pr(\Omega) = \sum_{\omega \in \Omega} \Pr(\omega)$. Nach der Normierungseigenschaft gilt $\sum_{\omega \in \Omega} \Pr(\omega) = 1$. Ähnlich ist $\Pr(\emptyset)$ eine leere Summe. \square

Selbsttest: Beweisen Sie die weiteren Aussagen durch Rückführen auf Aussagen über Mengen und mit Verwendung der Definitionen.



Definition 5.4. $\Omega \setminus A$ heißt **Komplementärereignis** zu A und wird mit \bar{A} bezeichnet.

Und wir betrachten wieder einige Beispiele zu den neuen Fakten und Begriffen.

Beispiel. Betrachte drei Würfel mit der in Abbildung 5.1 gezeigten Nummerierung. Hierbei haben gegenüberliegende Seiten jeweils den gleichen Wert.

Der Wahrscheinlichkeitsraum für Würfel A ist $\Omega_A = \{2, 6, 7\}$, der Wahrscheinlichkeitsraum für B ist $\Omega_B = \{1, 5, 9\}$ und der für C ist $\Omega_C = \{3, 4, 8\}$. Alle drei Wahrscheinlichkeitsräume haben eine Gleichverteilung als zugrunde liegenden Wahrscheinlichkeitsverteilung.

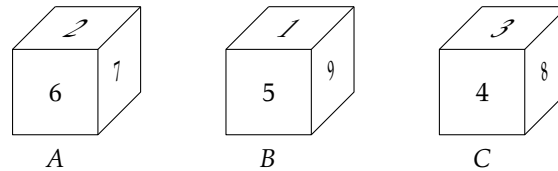


Abbildung 5.1: Würfel für Beispiel, gegenüberliegende Seiten haben den gleichen Wert.

Wir betrachten nun das Experiment, dass jeweils zwei der Würfel gleichzeitig geworfen werden. Wir sagen ein Würfel x gewinnt gegen einen anderen Würfel y , wenn die Wahrscheinlichkeit, dass die mit x geworfene Zahl größer als die Zahl von y ist, größer als $\frac{1}{2}$ ist. Seien G_{AB} , G_{BC} und G_{AC} Ereignisse. Hierbei ist G_{ij} das Ereignis, dass Würfel i gegen Würfel j gewinnt. Wir betrachten nun die Wahrscheinlichkeiten für diese Ereignisse.

Wenn Würfel A und B gegeneinander antreten gilt:

$$\Omega_{AB} = \{(2, 1), (2, 5), (2, 9), (6, 1), (6, 5), (6, 9), (7, 1), (7, 5), (7, 9)\} \text{ und}$$

$$G_{AB} = \{(2, 1), (6, 1), (6, 5), (7, 1), (7, 5)\}$$

Da auf Ω eine Gleichverteilung definiert ist, ist $\Pr(G_{AB}) = \frac{5}{9}$, also gewinnt A gegen B . Analog ist $\Pr(G_{BC}) = \frac{5}{9}$, also gewinnt B gegen C . Allerdings gilt auch $\Pr(G_{AC}) = \frac{4}{9}$, also ist *gewinnt gegen* keine transitive Relation. ◀

Das Beispiel mit den drei Würfeln zeigt, dass vieles in der Wahrscheinlichkeitsrechnung nicht mit Intuition betrachtet werden kann. Daher gilt:

Nichts glauben, immer nachrechnen!

Das nächste Beispiel kombiniert die Grundbegriffe der Wahrscheinlichkeitsrechnung mit den Identitäten und Formeln für Binomialkoeffizienten.

Beispiel. Es werden n unterscheidbare faire Münzen gleichzeitig geworfen. Dann ist $\Omega = \{k, z\}^n$ mit Gleichverteilung, also $\Pr(\omega) = \frac{1}{2^n}$ für alle $\omega \in \Omega$. Sei nun A_j das Ereignis, dass genau j mal Zahl geworfen wird. Es gibt $\binom{n}{j}$ Ereignisse aus Ω , bei denen genau j mal Zahl geworfen wird. Für $i \neq j$ gilt $A_i \cap A_j = \emptyset$ und:

$$\Pr(A_j) = \binom{n}{j} \cdot \frac{1}{2^n}$$

Sei nun B das Ereignis eine gerade Anzahl oft Zahl zu werfen. Es gilt:

$$B = A_0 \cup A_2 \cup \dots \cup A_{2 \cdot \lfloor \frac{n}{2} \rfloor}$$

$$\Rightarrow \Pr(B) = \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \Pr(A_k) = \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \frac{1}{2^n} \binom{n}{k} = \frac{1}{2^n} \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \binom{n}{k}$$

Wir werden nun zeigen, dass $\Pr(B) = \frac{1}{2}$ gilt. Dazu betrachten wir zunächst die folgenden Spezialfälle des binomischen Lehrsatzes:

$$2^n = (1 + 1)^n = \sum_{k=0}^n \binom{n}{k} \quad (5.1)$$

$$0 = 0^n = ((-1) + 1)^n = \sum_{k=0}^n (-1)^k \binom{n}{k} \quad (5.2)$$

Gleichung (5.1) und Gleichung (5.2) können wie folgt aufgeteilt werden:

$$\begin{aligned} 0 &= \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \binom{n}{k} - \sum_{\substack{k=0 \\ k \text{ ungerade}}}^n \binom{n}{k} \\ 2^n &= \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \binom{n}{k} + \sum_{\substack{k=0 \\ k \text{ ungerade}}}^n \binom{n}{k} \end{aligned}$$

Addieren wir beide Gleichungen, erhalten wir

$$2^n = 2 \cdot \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \binom{n}{k}$$

Diese Gleichung können wir nun oben einsetzen und erhalten schlussendlich:

$$\Pr(B) = \frac{1}{2^n} \sum_{\substack{k=0 \\ k \text{ gerade}}}^n \binom{n}{k} = \frac{1}{2^n} \cdot \frac{2^n}{2} = \frac{1}{2}$$

Das nächste Beispiel ist ein sehr bekanntes Beispiel: das sogenannte *Geburtstagsparadoxon*.

Selbsttest: Wie groß ist die Wahrscheinlichkeit ungerade häufig Zahl zu werfen?

Beispiel. Wir treffen zunächst einige Annahmen. Wir gehen davon aus, dass jedes Jahr $d = 365$ Tage hat und dass jedes Datum mit gleicher Wahrscheinlichkeit $\frac{1}{d}$ als Geburtsdatum auftritt. Dies stimmt in der Realität natürlich nicht, Daten die etwa neun Monate nach Silvester oder Karneval liegen sind deutlich wahrscheinlicher.

Wir stellen uns nun die Frage, wie viele Personen in einem Raum sein müssen, damit die Wahrscheinlichkeit, dass mindestens zwei Personen am gleichen Tag Geburtstag haben, mindestens $\frac{1}{2}$ ist. Wir betrachten zunächst die umgekehrte Frage. Wenn k Personen im Raum sind, wie hoch ist die Wahrscheinlichkeit, dass zwei von ihnen am gleichen Tag Geburtstag haben.

Der dem Experiment zugrunde liegenden Wahrscheinlichkeitsraum besteht aus den d^k Elementarereignissen, die jeder möglichen Verteilung der d möglichen Geburtstagen auf die k Personen beinhaltet. Statt dem Ereignis A : *Mindestens zwei Personen haben am gleichen Tag Geburtstag*, das uns eigentlich interessiert, betrachten wir das Komplementärereignis \bar{A} : *keine zwei Personen haben am gleichen Tag Geburtstag*.

Da alle möglichen Zuordnungen die gleiche Wahrscheinlichkeit haben, müssen wir nur Zählen, wie viele der Zuordnungen von Funktionen auf Geburtsdaten in \bar{A} sind. Dies sind genau die k -Permutationen einer d -Menge, da für jede der k Personen ein Geburtsdatum ausgewählt wird und alle möglichen Permutationen der Daten auf die Personen möglich sind. Also gilt

$$\Pr(A) = 1 - \Pr(\bar{A}) = 1 - \frac{d^{\underline{k}}}{d^k}$$

Für $k = 23$ und $d = 365$ ergibt sich $\Pr(A) \geq \frac{1}{2}$. Für eine Gruppe von $k = 60$ Personen (z.B. die Personen in einem Hörsaal) ergibt sich schon $\Pr(A) \geq 0.992$. ◀

Das letzte Beispiel ist eine Anwendung des Inklusions-Exklusionsprinzips auf Wahrscheinlichkeiten.

Beispiel. Wie hoch ist die Wahrscheinlichkeit, dass beim Würfeln von drei unterscheidbaren fairen Würfeln mindestens einer der Würfel eine feste Augenzahl i zeigt? Um das Beispiel besser nachvollziehen zu können, kann man sich zum Beispiel $i = 5$ vorstellen.

Sei A_k für $k \in \{1, 2, 3\}$ das Ereignis, dass der k -te Würfel die Zahl i zeigt. Wir suchen also $\Pr(A_1 \cup A_2 \cup A_3)$. Da die Ereignisse A_k nicht disjunkt sind, kann der Fakt für eine Partition des Ereignisses nicht verwendet werden. Stattdessen verwenden wir das Inklusions-Exklusionsprinzip für 3 Mengen. Es gilt also:

$$\begin{aligned}\Pr(A_1 \cup A_2 \cup A_3) &= \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - \Pr(A_1 \cap A_2) - \Pr(A_1 \cap A_3) - \Pr(A_2 \cap A_3) \\ &\quad + \Pr(A_1 \cap A_2 \cap A_3) \\ &= 3 \cdot \frac{1}{6} - 3 \cdot \frac{1}{36} + 1 \cdot \frac{1}{216} \\ &= \frac{91}{216} \approx 0.42\end{aligned}$$
 ◀

5.2 Bedingte Wahrscheinlichkeiten

Es gibt häufiger den Fall, dass die Wahrscheinlichkeit, dass ein Ereignis A eintritt, wenn bekannt ist, dass ein anderes Ereignis B schon eingetreten ist genauer eingeschätzt werden kann, als die Wahrscheinlichkeit, dass A eintritt, ohne dass mehr bekannt ist. Die Wahrscheinlichkeit, dass A eintritt, wenn bekannt ist, dass B eingetreten ist, nennt man *bedingte Wahrscheinlichkeit*.

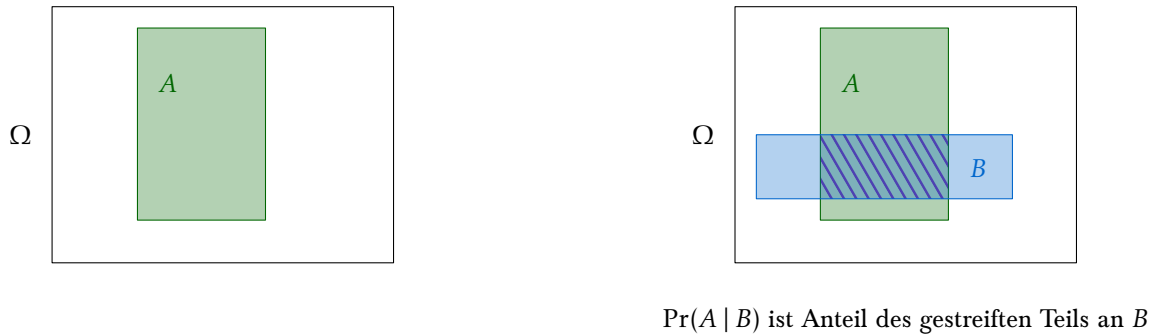


Abbildung 5.2: Visualisierung der bedingten Wahrscheinlichkeit

Definition 5.5. Sei (Ω, \Pr) ein Wahrscheinlichkeitsraum und seien A und B Ereignisse mit $\Pr(B) \neq 0$. Die **bedingte Wahrscheinlichkeit** $\Pr(A | B)$ ist definiert als:

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

Wenn \Pr eine Gleichverteilung ist, kann diese Definition anschaulich wie in Abbildung 5.2 dargestellt werden. Wir betrachten Ω als Rechteck und A als ein Rechteck innerhalb von Ω . Dann ist $\Pr(A)$ die Fläche von Ω geteilt durch die Fläche von A . Wird blind ein Dartpfeil auf Ω geworfen, ist die Wahrscheinlichkeit A zu treffen genau $\Pr(A)$. Für $\Pr(A | B)$ sind A und B zwei (unter Umständen überlappende) Quadrate in Ω . $\Pr(A | B)$ ist nun der Anteil von $A \cap B$ an der Fläche von B . In der Anschauung mit den Dartpfeilen, ist die bedingte Wahrscheinlichkeit $\Pr(A | B)$ die Wahrscheinlichkeit, dass der Dartpfeil in A gelandet ist, unter der Voraussetzung, dass B getroffen wurde.

Eng verwandt mit der Definition der bedingten Wahrscheinlichkeit ist der Satz von Bayes.

Satz 5.2 (Satz von Bayes). Sei (Ω, \Pr) ein Wahrscheinlichkeitsraum und seien A, B Ereignisse mit $\Pr(B) \neq 0$. Dann gilt

$$\Pr(A | B) = \frac{\Pr(A) \cdot \Pr(B | A)}{\Pr(B)}$$

Beweis. Der Beweis ergibt sich durch mehrfaches Anwenden der Definition der bedingten Wahrscheinlichkeit.

$$\Pr(A | B) \cdot \Pr(B) = \Pr(A \cap B)$$

$$\Pr(B | A) \cdot \Pr(A) = \Pr(B \cap A)$$

Gleichsetzen ergibt

$$\Pr(A | B) \cdot \Pr(B) = \Pr(B | A) \cdot \Pr(A)$$

$$\Leftrightarrow \Pr(A | B) = \frac{\Pr(B | A) \cdot \Pr(A)}{\Pr(B)}$$

□

Beispiel. Gegeben seien zwei Boxen (Box 1 und Box 2). In jeder der Boxen sind sieben Bälle. In Box 1 sind 2 orange und 5 blaue Bälle. In Box 2 sind 4 orange und 3 blaue Bälle. In einem Zufallsexperiment wird nun zunächst zufällig gleichverteilt eine Box gewählt und dann aus dieser Box dann zufällig ein Ball. Der Wahrscheinlichkeitsraum ist nun (Ω, \Pr) mit $\Omega = \{1, 2\} \times \{b, o\}$. Anders als bei den bisherigen Beispielen haben wir hier *keine* Gleichverteilung.

Uns interessiert nun die Wahrscheinlichkeit, dass Box 1 gewählt wurde, wenn eine orange Kugel gezogen worden ist. Dafür ist A das Ereignis, dass Box 1 gewählt wurde und B das Ereignis, dass eine orange Kugel gezogen wurde. Dann suchen wir $\Pr(A | B)$. Um den Satz von Bayes anwenden zu können, brauchen wir nun $\Pr(B | A)$, $\Pr(A)$ und $\Pr(B)$. $\Pr(A)$ ergibt sich direkt aus der Beschreibung, genauso wie $\Pr(B | A)$:

$$\Pr(A) = \frac{1}{2} \qquad \Pr(B | A) = \frac{2}{7}$$

Es fehlt nun noch $\Pr(B)$. Hierzu verwenden wir den Fakt, dass $\Pr(B) = \Pr(A \cap B) + \Pr(\bar{A} \cap B)$ gilt. In unserem Fall ist $\Pr(A \cap B) = \frac{1}{2} \cdot \frac{2}{7}$ und $\Pr(\bar{A} \cap B) = \frac{1}{2} \cdot \frac{4}{7}$, insgesamt ist also $\Pr(B) = \frac{6}{14}$. Es ergibt sich:

$$\Pr(A | B) = \frac{\frac{1}{2} \cdot \frac{2}{7}}{\frac{6}{14}} = \frac{1}{3}$$

Alternativ kann der Sachverhalt auch mit einem sogenannten Baumdiagramm dargestellt werden, siehe Abbildung 5.3. Hierbei entspricht der unterste Pfad dem Ereignis, dass erst die zweite Box und dann eine blaue Kugel gezogen wird. Die einzelnen Wahrscheinlichkeiten an den Kanten sind die (bedingten) Wahrscheinlichkeiten, dass das Ereignis an der Kante eintritt.

Aus dem Baudiagramm ergibt sich $\Pr(B) = \frac{6}{14}$ und $\Pr(A \cap B) = \frac{2}{14}$. Mit der Definition der bedingten Wahrscheinlichkeit ist dann wieder $\Pr(A | B) = \frac{2}{6} = \frac{1}{3}$. ◀

Beispiel. Das nächste Beispiel ist ein klassisches Beispiel für die Wichtigkeit von bedingten Wahrscheinlichkeiten. In einer Spielshow bekommt der Gast drei Türen gezeigt. Hinter einer der Türen befindet sich ein Auto, hinter den anderen zwei Türen jeweils eine Ziege. Der Gast wählt eine der drei Türen aus, daraufhin öffnet der Gastgeber immer eine der nicht ausgewählten Türen, hinter der sich eine Ziege befindet. Der Gastgeber bietet nun dem Gast immer an, die Tür zu wechseln. Die Frage ist, ob sich die Gewinnwahrscheinlichkeit steigert, wenn der Gast dieses Angebot annimmt.

Wir modellieren das Experiment als zweistufiges Zufallsexperiment, wobei wir annehmen, dass die Tür, hinter der das Auto ist, zufällig gewählt wurde und der Gast zunächst immer Tür 1 wählt. Die Wahrscheinlichkeiten ergeben sich aus dem Baumdiagramm in Abbildung 5.4.

Sei A_i das Ereignis, dass das Auto hinter Tür i steht und B_j das Ereignis, dass der Gastgeber Tür j öffnet. Dann gilt:

$$\Pr(A_j) = \frac{1}{3} \text{ für } j = 1, 2, 3 \qquad \Pr(B_j) = \frac{1}{2} \text{ für } j = 2, 3$$

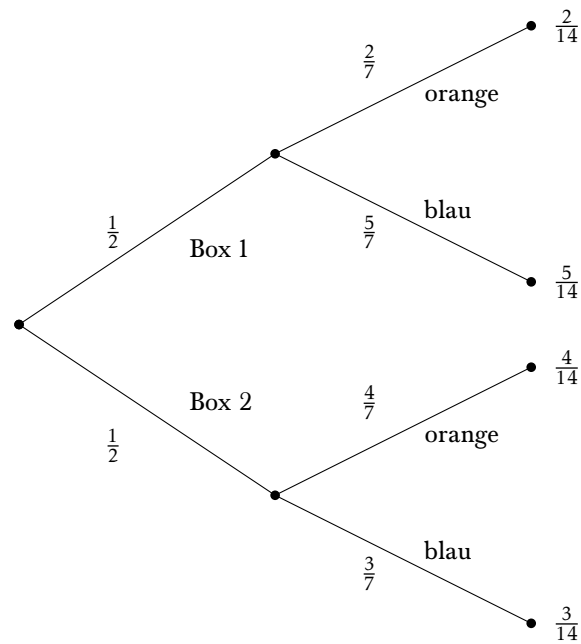


Abbildung 5.3: Baumdiagramm für das Beispiel mit den Boxen und Bällen

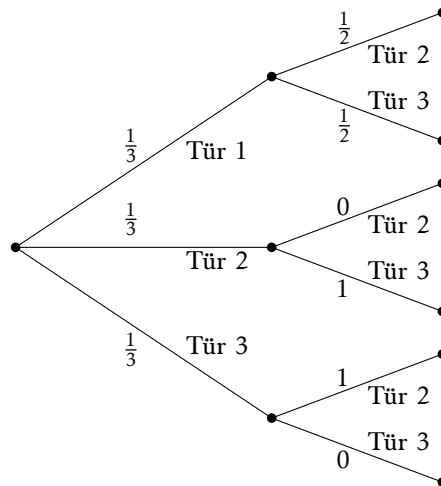


Abbildung 5.4: Baumdiagramm für das Ziegenproblem

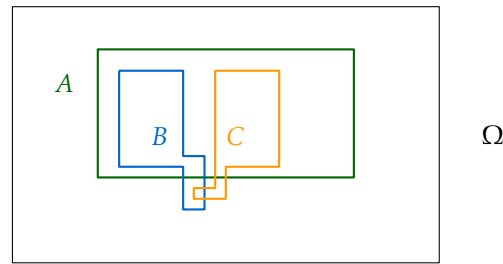


Abbildung 5.5: Gegenbeispiel zu 5.3

$$\Pr(B_2 | A_1) = \frac{1}{2}$$

$$\Pr(B_2 | A_2) = 0$$

$$\Pr(B_2 | A_3) = 1$$

$$\Pr(B_3 | A_1) = \frac{1}{2}$$

$$\Pr(B_3 | A_2) = 1$$

$$\Pr(B_3 | A_3) = 0$$

Mit dem Satz von Bayes ergibt sich dann die Wahrscheinlichkeit, dass das Auto hinter Tür 2 (bzw. 3) ist, wenn Tür 3 (bzw. 2) geöffnet wurde:

$$\Pr(A_2 | B_3) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{2}} = \frac{2}{3}$$

$$\Pr(A_3 | B_2) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{2}} = \frac{2}{3}$$

Damit ist die Gewinnwahrscheinlichkeit höher, wenn die Tür gewechselt wird. ◀

Lemma 5.3. Es gilt

$$\Pr(A \cup B | C) = \Pr(A | C) + \Pr(B | C) - \Pr(A \cap B | C)$$

Beweis. Die Aussage folgt direkt nach dem Inklusions-Exklusionsprinzip. Dies auszurechnen ist dem Leser als Übung überlassen. ◻

Achtung! Umgekehrt gilt die Aussage im Allgemeinen nicht. Es ist also

$$\Pr(A | B \cup C) \neq \Pr(A | B) + \Pr(A | C) - \Pr(A | B \cap C) \quad (5.3)$$

Dies kann man sich direkt an einer Skizze klarmachen. In Abbildung 5.5 sind $\Pr(A | B)$, $\Pr(A | C)$ und $\Pr(A | B \cup C)$ alle fast 1. $\Pr(A | B \cap C)$ hingegen ist 0.

5.2.1 Unabhängigkeit von Ereignissen

Das Konzept der bedingten Wahrscheinlichkeit hat uns gezeigt, dass die Wahrscheinlichkeit, ob ein Ereignis eintritt, davon abhängen kann, ob ein anderes Ereignis eingetreten ist. Ist dies nicht der Fall, spricht man auch von *unabhängigen Ereignissen*.

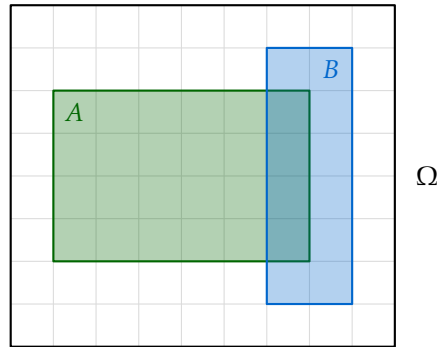


Abbildung 5.6: Visualisierung der Definition der Unabhängigkeit

Definition 5.6. Zwei Ereignisse $A, B \subseteq \Omega$ heißen **unabhängig**, falls gilt $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.

In Abbildung 5.6 gibt es eine visuelle Interpretation der Definition. Wenn eine Gleichverteilung angenommen wird, ist der Anteil von $A \cap B$ an B gleich dem Anteil von A an Ω . (Analog ist der Anteil von $A \cap B$ an A gleich dem Anteil von B an Ω .)

Lemma 5.4. Wenn A und B unabhängig sind, gilt:

$$\Pr(A) = \Pr(A \mid B)$$

Beweis. Dies folgt direkt aus einer Kombination der Definition der Unabhängigkeit und der bedingten Wahrscheinlichkeiten:

$$\Pr(A) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A \mid B) \quad \square$$

Die Definition der Unabhängigkeit lässt sich auf verschiedene Arten auf mehr als zwei Ereignisse verallgemeinern.

Definition 5.7.

- Eine Menge $\{A_1, \dots, A_k\}$ von Ereignissen mit $k > 1$ ist **unabhängig**, falls gilt:

$$\Pr(A_1 \cap A_2 \cap \dots \cap A_k) = \Pr(A_1) \cdot \Pr(A_2) \cdot \dots \cdot \Pr(A_k)$$

- Eine Familie von Ereignissen heißt **unabhängige Familie**, wenn alle Teilmengen unabhängig sind.
- Eine Familie von Ereignissen heißt **paarweise unabhängig**, wenn jedes Paar von Ereignissen unabhängig ist.

Beispiel. Ein fairer Würfel mit 6 Seiten wird gewürfelt. Sei $A = \{1, 2, 3\}$, $B = \{1, 4, 5\}$ und $C = \{1, 2, 3, 4\}$. Wir wollen nun untersuchen, in welchen Konstellationen A, B und C unabhängig sind. Dazu bestimmen wir zunächst alle benötigten Wahrscheinlichkeiten:

$$\begin{aligned} \Pr(A) &= \frac{1}{2} & \Pr(B) &= \frac{1}{2} & \Pr(C) &= \frac{2}{3} \\ \Pr(A \cap B) &= \frac{1}{6} & \Pr(A \cap C) &= \frac{1}{2} & \Pr(B \cap C) &= \frac{1}{3} \\ \Pr(A \cap B \cap C) &= \frac{1}{6} \end{aligned}$$

Wir stellen fest, dass

$$\begin{aligned} \Pr(A \cap B \cap C) &= \frac{1}{6} = \Pr(A) \cdot \Pr(B) \cdot \Pr(C) \\ \Pr(A \cap B) &= \frac{1}{6} \neq \frac{1}{4} = \Pr(A) \cdot \Pr(B) \\ \Pr(A \cap C) &= \frac{1}{2} \neq \frac{1}{3} = \Pr(A) \cdot \Pr(C) \\ \Pr(B \cap C) &= \frac{1}{3} = \Pr(B) \cdot \Pr(C) \end{aligned}$$

Damit sind A, B, C als Tripel unabhängig, genauso B und C als Paar. Die Paare A, B und A, C sind nicht unabhängig. Damit ist $\{A, B, C\}$ keine unabhängige Familie und auch keine paarweise unabhängige Familie. ◀

Das nächste Beispiel zeigt nochmal sehr deutlich, dass Unabhängigkeit eine Eigenschaft ist, die von den Wahrscheinlichkeiten der einzelnen Ereignisse abhängt.

Beispiel. Eine nicht faire Münze wird zweimal geworfen. Dabei ist $\Pr(k) = p$ und $\Pr(z) = (1 - p)$ für $0 \leq p \leq 1$. Das Ereignis A ist, dass beide Würfe das gleiche Ergebnis liefern, Ereignis B ist, dass der erste Wurf Kopf ergibt. Die Frage ist nun, für welche Werte von p sind A und B unabhängig. Wir bestimmen wieder die relevanten Wahrscheinlichkeiten:

$$\Pr(A) = p^2 + (1 - p)^2 \quad \Pr(B) = p \cdot (1 - p) + p^2 = p \quad \Pr(A \cap B) = p^2$$

Die Ereignisse sind also unabhängig, wenn $p^2 = (p^2 + (1 - p)^2) \cdot p$ ist. Dies sind genau die Werte, für die $0 = p(2 \cdot p^2 - 3 \cdot p + 1)$ gilt. Ausrechnen ergibt $p = 0$, $p = \frac{1}{2}$ und $p = 1$. ◀

Zur Unabhängigkeit sollten man sich das Folgende merken:

- Unabhängigkeit ist eine Eigenschaft von mehreren Ereignissen
- Disjunkte Ereignisse mit positiver Wahrscheinlichkeit sind niemals unabhängig
- Unabhängigkeit und paarweise Unabhängigkeit einer Familie von Ereignissen sind verschiedene (aber verwandte) Konzepte

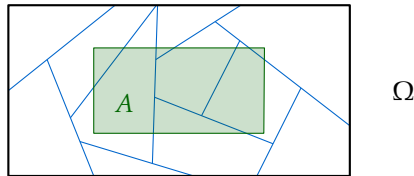


Abbildung 5.7: Illustration zum Satz der totalen Wahrscheinlichkeit

Selbsttest: Was sind die Unterschiede zwischen unabhängigen Ereignissen, paarweiser Unabhängigkeit und einer unabhängigen Familie?



5.2.2 Satz der totalen Wahrscheinlichkeit

Zum Abschluss der grundlegenden Konzepte betrachten wir den Satz der totalen Wahrscheinlichkeit, der häufig praktisch sein kann.

Satz 5.5 (Satz der totalen Wahrscheinlichkeit). Sei $(B_i)_{i \in I}$ eine Partition der Elementarereignisse eines Wahrscheinlichkeitsraums (Ω, Pr) . Dann gilt für ein beliebiges Ereignis A :

$$\text{Pr}(A) = \sum_{i \in I} \text{Pr}(A \mid B_i) \cdot \text{Pr}(B_i)$$

Beweis. Die B_i sind eine Zerlegung von Ω . Dann sind die nichtleeren (und disjunkten) Mengen $(A \cap B_i)_{i \in I}$ eine Partition von A , wie in Abbildung 5.7 skizziert.

Also ist:

$$\text{Pr}(A) = \sum_{i \in I} \text{Pr}(A \cap B_i)$$

und mit der Definition der bedingten Wahrscheinlichkeit

$$\text{Pr}(A) = \sum_{i \in I} \text{Pr}(A \mid B_i) \cdot \text{Pr}(B_i) \quad \square$$

Beispiel. Wir betrachten das in Abbildung 5.8 abgebildete gerichtete Wegenetz². Wir starten in s und gehen in jedem Schritt zufällig gleichverteilt einen der Wege entlang. Wie hoch ist die Wahrscheinlichkeit am Ende in t zu stehen?

Sei B_i die Menge aller Wege, die durch p_i gehen. Dann gilt $\text{Pr}(B_i) = \frac{1}{4}$ für alle $i = 1, 2, 3, 4$. Die Menge der B_i ist eine Partition aller möglicher Wege.

A ist nun das Ereignis, dass der gewählte Weg von s nach t geht. Mit dem Satz der totalen Wahrscheinlichkeit gilt:

$$\text{Pr}(A) = \frac{1}{4} \cdot \left(\frac{1}{3} + \frac{1}{2} + 1 + \left(\frac{1}{4} + \frac{1}{4} \right) \right) = \frac{7}{12}$$

²besteht aus lauter Einbahnstraßen

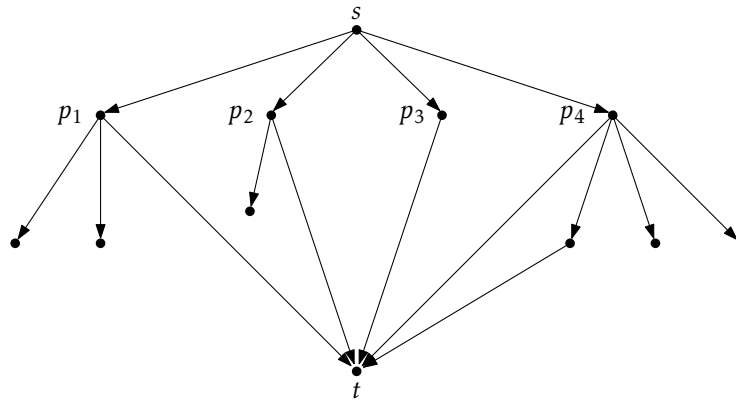


Abbildung 5.8: Wegenetz

5.3 Zufallsvariablen und Erwartungswert

5.3.1 Zufallsvariablen

Häufig interessiert uns ein Zahlenwert, der einem Zufallsexperiment zugeordnet ist. Ein Beispiel kann die Anzahl an Würfeln die Kopf ergeben beim Werfen von 20 Münzen sein. Dieses Konzept wird durch das Konzept einer Zufallsvariablen abgebildet. Hierbei ist der Name „Zufallsvariable“ etwas verwirrend, da es sich eigentlich um eine Funktion handelt.

Definition 5.8. Sei (Ω, \Pr) ein Wahrscheinlichkeitsraum. Eine Zufallsvariable (Zufallsgröße) ist eine Funktion:

$$X : \Omega \rightarrow \mathbb{R}$$

Eine Zufallsvariable weist also einem Elementarereignis einen Zahlenwert aus \mathbb{R} zu.

Beispiel.

- Wir betrachten einen fairen Münzwurf, also $\Omega = \{k, z\}$ mit Gleichverteilung. Wir definieren eine Zufallsvariable mit $X(k) = 0$ und $X(z) = 1$.
- $\Omega = \{k, z\}^n$ und X zählt die Anzahl von Münzen, die Kopf zeigen.
- $\Omega = \{1, \dots, 6\}^2$, also zwei Würfel gleichzeitig werfen. Dann sind zwei Beispiele für Zufallsvariablen $X_{\text{sum}}(i, j) = i + j$, die Zufallsvariable zählt also die Summe der Augen, und $X_{\text{max}}(i, j) = \max\{i, j\}$. ◀

Selbsttest: Finden Sie interessante Zufallsvariablen zum gleichverteilten Wahrscheinlichkeitsraum mit $\Omega = \{1, \dots, 6\}^n$. ?

Häufig interessiert uns die Wahrscheinlichkeit, mit der eine Zufallsvariable einen bestimmten Wert annimmt. Diese Wahrscheinlichkeit hängt von der Wahrscheinlichkeitsverteilung auf Ω ab. Formal können wir für jedes x im Bild von X ein Ereignis

$$(X = x) = \{\omega \in \Omega \mid X(\omega) = x\}$$

definieren. Dieses fasst jeweils alle Elementarereignisse von Ω die mit X auf den gleichen Wert abgebildet werden zusammen. Mithilfe dieser Ereignisse ergibt sich dann ein Wahrscheinlichkeitsraum $(X(\Omega), \Pr_X)$ mit

$$\Pr_X(x) := \Pr(X = x) = \sum_{\omega \in \Omega, X(\omega)=x} \Pr(\omega)$$

Beispiel. Werden zwei unterscheidbare faire Würfel geworfen ist $\Omega = \{1, 2, 3, 4, 5, 6\}^2$. Sei $X_{\max}(i, j) = \max\{i, j\}$ die Zufallsvariable, die einem Wurf das Maximum der Ergebnisse zuweist. Das Bild $X_{\max}(\Omega)$ ist $\{1, 2, 3, 4, 5, 6\}$. Und es ergibt sich die folgende induzierte Wahrscheinlichkeitsverteilung:

$$\Pr(X_{\max} = 1) = \Pr(\{(1, 1)\}) = \frac{1}{36}$$

$$\Pr(X_{\max} = 2) = \Pr(\{(1, 2), (2, 2), (2, 1)\}) = \frac{3}{36}$$

$$\Pr(X_{\max} = 3) = \Pr(\{(1, 3), (2, 3), (3, 3), (3, 1), (3, 2)\}) = \frac{5}{36}$$

$$\Pr(X_{\max} = 4) = \Pr(\{(1, 4), (2, 4), (3, 4), (4, 4), (4, 3), (4, 2), (4, 1)\}) = \frac{7}{36}$$

$$\Pr(X_{\max} = 5) = \Pr(\{(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (5, 4), (5, 3), (5, 2), (5, 1)\}) = \frac{9}{36}$$

$$\Pr(X_{\max} = 6) = \Pr(\{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 5), (6, 4), (6, 3), (6, 2), (6, 1)\}) = \frac{11}{36}$$

Selbsttest: Betrachten Sie den gleichverteilten Wahrscheinlichkeitsraum mit $\Omega = \{k, z\}^3$ und die Zufallsvariable, die einem Elementarereignis die Anzahl von Würfeln mit Wert Kopf zuweist. Geben Sie den Wahrscheinlichkeitsraum $(X(\Omega), \Pr_X)$ an.

5.3.2 Erwartungswert

Häufig wird einer Zufallsvariablen³ ein Wert zugewiesen. Einer der wichtigsten dieser Werte ist der Erwartungswert, der den durchschnittlichen Wert angibt, den eine Zufallsvariable annehmen kann.

Definition 5.9. Sei (Ω, \Pr) ein Wahrscheinlichkeitsraum und X eine Zufallsvariable. Der **Erwartungswert** von X ist definiert als

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr(\omega)$$

³Erinnerung: Das ist eine Funktion!

Bei einer Gleichverteilung gilt nach Definition $E(X) = \frac{\sum_{\omega \in \Omega} X(\omega)}{|\Omega|}$. Wenn über alle Elementarereignisse addiert wird, kann die Summe unter Umständen sehr lang werden. Wenn der Wahrscheinlichkeitsraum $(X(\Omega), \Pr_X)$ bekannt ist, kann mit dem folgenden Lemma der Erwartungswert in vielen Fällen einfacher bestimmt werden.

Lemma 5.6. Sei $X(\Omega)$ das Bild einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$. Es gilt

$$E(X) = \sum_{x \in X(\Omega)} x \cdot \Pr(X = x)$$

Beweis. Der Beweis geschieht durch Umstrukturieren der Summe aus der Definition des Erwartungswertes, sowie der Definition der Wahrscheinlichkeitsverteilung auf $X(\Omega)$. Intuitiv werden alle Summanden $X(\omega) \cdot \Pr(\omega)$ aus der Formel in der Definition die den gleichen Wert $X(\omega)$ haben gruppiert.

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega) \cdot \Pr(\omega) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \in (X=x)} X(\omega) \cdot \Pr(\omega) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \in (X=x)} x \cdot \Pr(\omega) \\ &= \sum_{x \in X(\Omega)} x \cdot \sum_{\omega \in (X=x)} \Pr(\omega) \\ &= \sum_{x \in X(\Omega)} x \cdot \Pr(X = x) \end{aligned}$$

□

Beispiel.

- $\Omega = \{1, 2, \dots, 6\}^2$ gleichverteilt mit $X(i, j) = \max\{i, j\}$. Dann ist

$$\begin{aligned} E(X) &= \sum_{x \in \{1, \dots, 6\}} x \cdot \Pr(X = x) \\ &= 1 \cdot \Pr(X_{\max} = 1) + 2 \cdot \Pr(X_{\max} = 2) + 3 \cdot \Pr(X_{\max} = 3) \\ &\quad + 4 \cdot \Pr(X_{\max} = 4) + 5 \cdot \Pr(X_{\max} = 5) + 6 \cdot \Pr(X_{\max} = 6) \\ &= 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} \\ &= \frac{161}{36} \approx 4,47 \end{aligned}$$

- $\Omega = \{1, 2, \dots, 6\}$ gleichverteilt mit $X(\omega) = \omega$. Der Wert der Zufallsvariable bei Würfelwurf ω ist also ω . Es gilt $X(\Omega) = \{1, \dots, 6\}$ und

$$E(X) = \frac{1}{6} \cdot \sum_{i=1}^6 i = \frac{21}{6} = 3,5$$

Diese beiden Beispiele zeigen, dass der Erwartungswert nicht im Bild der Zufallsvariablen liegen muss. ◀

Nun betrachten wir ein etwas komplexeres Beispiel:

Beispiel. Drei Personen A, B und C wetten auf den Ausgang eines Münzwurfs. Alle drei Personen setzen 2€. Der Einsatz wird zwischen allen Teilnehmenden die richtig geraten haben aufgeteilt. Hat keiner richtig geraten, bekommen alle ihren Einsatz zurück. Wenn wirklich alle nur raten, ist der erwartete Gewinn für jede einzelne Person 0€. ⁴

Nun tun sich B und C zusammen und setzen immer entgegengesetzt. Wir interessieren uns jetzt für den erwarteten Gewinn von A . Formal modellieren wir das Problem als gleichverteilten Wahrscheinlichkeitsraum mit $\Omega = \{+, -\}^3$. Hierbei ist zum Beispiel das Elementarereignis $(+, +, -)$ zu lesen, als A hat richtig getippt, B hat richtig getippt und C hat falsch getippt.

Wir betrachten die Zufallsvariable X_A , die den Reingewinn von A angibt. Der Reingewinn ergibt sich aus dem Gewinn minus dem Einsatz. Die möglichen Reingewinne sind $\{-2, 0, 1, 4\}$. Um die Wahrscheinlichkeiten für die einzelnen Reingewinne zu bestimmen, betrachten wir das in Abbildung 5.9 dargestellte Baumdiagramm und die zugehörige Tabelle. Damit ergibt sich für den Erwartungswert

$$\begin{aligned} E(X_A) &= (-2) \cdot \left(\frac{1}{4} + \frac{1}{4}\right) + 0 \cdot 0 + 1 \cdot \left(\frac{1}{4} + \frac{1}{4}\right) + 4 \cdot 0 \\ &= (-2) \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \\ &= -\frac{1}{2} \end{aligned}$$

Damit wird Person A über mehrere Spiele hinweg Verlust machen. ◀

Selbsttest: Rechnen Sie nach, dass der erwartete Gewinn beim *fairen* Spiel 0€ ist. ?

Häufig zählen Zufallsvariablen etwas, sind also Funktionen $X : \Omega \rightarrow \mathbb{N}$. Für diese kann der Erwartungswert noch einmal anders aufgeschrieben werden:

Lemma 5.7. Für eine Zufallsvariable $X : \Omega \rightarrow \mathbb{N}$ gilt:

$$E(X) = \sum_{i=1}^{\infty} i \cdot \Pr(X = i)$$

Beweis. Die Aussage folgt aus $E(X) = \sum_{x \in X(\Omega)} i \cdot \Pr(X = i)$. ◻

Selbsttest: Machen Sie sich klar, warum der (sehr kurze) Beweis für Lemma 5.7 gilt. ?

⁴Dies nachzurechnen ist eine einfache Aufgabe

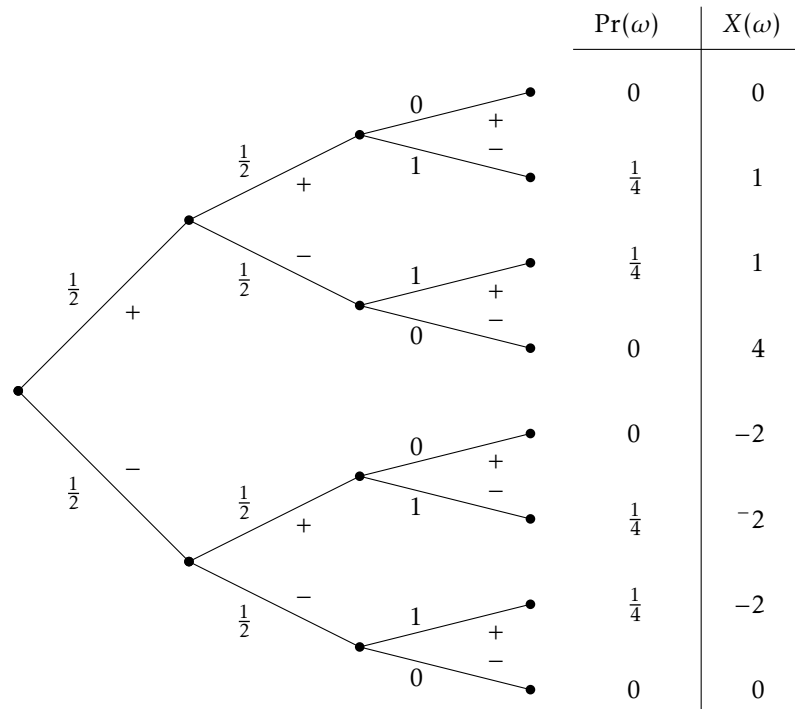


Abbildung 5.9: Visualisierung des Münzspiels

Lemma 5.8. Für $X : \Omega \rightarrow \mathbb{N}$ gilt:

$$E(X) = \sum_{i=0}^{\infty} \Pr(X > i)$$

Beweis. Es gilt $(X > i) = \bigcup_{j=i+1}^{\infty} (X = j)$. Also gilt

$$\Pr(X > i) = \sum_{j=i+1}^{\infty} \Pr(X = j) \quad \text{und}$$

$$E(X) = \sum_{i=1}^{\infty} i \cdot \Pr(X = i)$$

Wir schreiben die einzelnen Wahrscheinlichkeiten $\Pr(X > i)$ als Tabelle auf.

$\Pr(X = 1)$	+	$\Pr(X = 2)$	+	$\Pr(X = 3)$	+	$\Pr(X = 4)$	+	...	$\Pr(X > 0)$
		$\Pr(X = 2)$	+	$\Pr(X = 3)$	+	$\Pr(X = 4)$	+	...	$\Pr(X > 1)$
				$\Pr(X = 3)$	+	$\Pr(X = 4)$	+	...	$\Pr(X > 2)$
					+	$\Pr(X = 4)$	+	...	$\Pr(X > 3)$
<hr/>									
$1 \cdot \Pr(X = 1)$	+	$2 \cdot \Pr(X = 2)$	+	$3 \cdot \Pr(X = 3)$	+	$4 \cdot \Pr(X = 4)$	+	...	$E(X)$

Mit dem Prinzip des doppelten Abzählens folgt die Aussage. □

Wir werden uns in Abschnitt 5.4.3 eine Anwendung dieser Gleichung anschauen.

Zum Bestimmen von Erwartungswerten ist die *Linearität des Erwartungswerts* eine wichtige Eigenschaft. Um diese formal betrachten zu können, definieren wir zunächst was es bedeutet, zwei Zufallsvariable zu addieren oder eine Zufallsvariable mit einer Konstanten zu multiplizieren.

Seien $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariablen und sei c eine reelle Zahl. Wir definieren neuen Zufallsvariablen $(X_1 + X_2)$ und (cX) durch

$$\begin{aligned}(X_1 + X_2)(\omega) &= X_1(\omega) + X_2(\omega) \\ (c \cdot X_1)(\omega) &= c \cdot X_1(\omega)\end{aligned}$$

Hierbei ist $+$ die neu definierte Addition von Zufallsvariablen und $+$ die bekannte Addition auf den reellen Zahlen, ähnlich für \cdot und \cdot .

Anmerkung: Die Operatoren $+$ und \cdot definieren Operationen auf *Funktionen*. Ein Ähnliches Konzept ist eventuell aus der Vorlesung *Konzepte der Programmierung* bekannt. $+$ kann so interpretiert werden, dass es zwei Funktionen (X_1, X_2) als Eingabe bekommt und als Ausgabe die neue Funktion $(X_1 + X_2)$ hat.

Diese ist dann so definiert, dass der Eingabewert ω zunächst in X_1 und dann in X_2 eingesetzt wird. Die Ergebnisse sind reelle Zahlen. Diese können dann mit der bekannten Addition $+$ auf den reellen Zahlen addiert werden.

Aus der Definition des Erwartungswerts folgt direkt das folgende Lemma:

Lemma 5.9 (Linearität des Erwartungswerts). Für Zufallsvariablen $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ und $c \in \mathbb{R}$ gilt:

- $E(X_1 + X_2) = E(X_1) + E(X_2)$
- $E(c \cdot X_1) = c \cdot E(X_1)$

Allgemein gilt für $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ und $c_1, \dots, c_n \in \mathbb{R}$

$$E(c_1 \cdot X_1 + c_2 \cdot X_2 + \dots + c_n \cdot X_n) = c_1 \cdot E(X_1) + c_2 \cdot E(X_2) + \dots + c_n \cdot E(X_n)$$

Beweis. Die ersten beiden Aussagen folgen aus der Definition des Erwartungswertes. Die dritte Aussage kann dann mit vollständiger Induktion gezeigt werden. \square

Selbsttest: Führen Sie die Beweise von Lemma 5.9.

Beispiel. Vor der Vorlesung haben n Studierende ihre Jacken an eine Garderobe gehängt. In der Vorlesung geht ein Feueralarm los und jeder greift beim Verlassen des Raumes eine zufällige Jacke. Was ist die erwartete Anzahl an Studierenden, die ihre eigene Jacke greifen?

Auf dem Wahrscheinlichkeitsraum ist eine Gleichverteilung definiert mit Ω als die Menge aller Permutationen von $\{1, \dots, n\}$. Wir definieren zunächst sogenannte

Indikatorzufallsvariablen $X_i : \Omega \rightarrow \{0, 1\}$ für $i = 1, \dots, n$ und stellen die Anzahl der Studierenden mit eigener Jacke dann als Summe dieser Zufallsvariablen dar.

$$X_i = \begin{cases} 0 & \text{Person } i \text{ hat eine fremde Jacke} \\ 1 & \text{sonst} \end{cases}$$

Ist X nun die Anzahl der Studierenden, die ihre eigene Jacke haben gilt $X = \sum_{i=1}^n X_i$. Wir bestimmen nun $E(X_i)$. Es gilt

$$\begin{aligned} \Pr(X_i = 1) &= \frac{(n-1)!}{n!} = \frac{1}{n} \\ \Pr(X_i = 0) &= 1 - \Pr(X_i = 1) = \frac{n-1}{n} \\ E(X_i) &= 0 \cdot \frac{n-1}{n} + 1 \cdot \frac{1}{n} = \frac{1}{n} \end{aligned}$$

Mit der Linearität des Erwartungswertes ergibt sich dann

$$E(X) = \sum_{i=1}^n E(X_i) = 1$$

Im Erwartungswert wird also nur genau ein Studierender die eigene Jacke bekommen.

Beispiel.

Hier noch das Münzen und stehenbleiben Beispiel aus der VL einfügen

5.4 Besondere Verteilungen

Einige Zufallsverteilungen treten immer wieder auf und haben daher besondere Namen. Für diese kann dann einmal die Wahrscheinlichkeit für eine bestimmte Art von Ereignis und eine allgemeine Formel für den Erwartungswert bestimmt werden. Folgt ein Zufallsexperiment dann dieser Verteilung, können die Formeln einfach verwendet werden und die Werte müssen nicht jedes Mal neu bestimmt werden.

5.4.1 Bernoulli-Verteilung

Wir haben nun schon mehrfach Zufallsexperimente mit genau zwei möglichen Ereignissen gesehen. Diese werden häufig mit Erfolg oder Misserfolg bezeichnet. Im Folgenden gehen wir davon aus, dass ein Erfolg mit Wahrscheinlichkeit p und ein Misserfolg mit Wahrscheinlichkeit $1-p$ eintritt, dabei ist $0 \leq p \leq 1$ eine für das Experiment feste Zahl.

Sei nun X die Zufallsvariable, die einen Erfolg auf 1 und einen Misserfolg auf 0 abbildet. Für den Erwartungswert von X gilt.

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Wir nennen diesen Erwartungswert auch den Erwartungswert einer Bernoulli-verteilten Zufallsvariable. ■

5.4.2 Binomialverteilung

Um die Binomialverteilung für $n > 0$ zu definieren, betrachten wir ein Zufallsexperiment, in dem ein Bernoulli-Experiment n mal unabhängig ausgeführt wird. Sei nun X die Zufallsvariable, die die Anzahl k von Erfolgen zählt. Es ist also $X(\Omega) = \{0, 1, 2, \dots, n\}$. Die von X induzierte Wahrscheinlichkeitsverteilung auf dem Bild von X heißt **Binomialverteilung**. Genauer, sei $b(k, n, p)$ die Wahrscheinlichkeit, dass genau k Erfolge bei den Bernoulli-Experimenten mit Parameter p eintreten.

Satz 5.10. Es gilt:

$$(a) \quad b(k, n, p) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

$$(b) \quad b(k, n, p) \text{ beschreibt eine Wahrscheinlichkeitsverteilung auf } \{0, 1, \dots, n\}$$

$$(c) \quad E(X) = n \cdot p$$

Beweis.

- (a) Wir suchen also die Wahrscheinlichkeit bei n Experimenten genau k mal einen Erfolg zu haben. Wird die Reihenfolge außer Acht gelassen, ist die Wahrscheinlichkeit k Erfolge und $n - k$ Misserfolge zu haben genau $p^k \cdot (1 - p)^{n-k}$. Es gibt $\binom{n}{k}$ Möglichkeiten, um die k Erfolge auf die n Versuche zu verteilen. Damit gilt $b(k, n, p) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$.
- (b) Wir überprüfen nun, ob $b(k, n, p)$ tatsächlich eine Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$, also ob die Summe der Wahrscheinlichkeiten tatsächlich 1 ist. Aus dem binomischen Lehrsatz folgt:

$$\sum_{k=0}^n \binom{n}{k} \cdot p^k (1 - p)^{n-k} = (p + 1 - p)^n = 1^n = 1$$

- (c) Nun betrachten wir den Erwartungswert von X . Wir benutzen wieder den Trick eine Indikatorzufallsvariable zu benutzen. X lässt sich schreiben als

$$X = X_1 + X_2 + \dots + X_n$$

wobei X_i eine Indikatorzufallsvariable ist, die angibt, ob Experiment i ein Erfolg war. Da die X_i Bernoulli-verteilt sind, ist $E(X_i) = p$. Wegen der Linearität des Erwartungswertes ist dann:

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_n) = n \cdot p \quad \square$$

5.4.3 Geometrische Verteilung

Es gibt in der echten Welt häufig Prozesse, die so lange wiederholt werden, bis ein bestimmtes Ergebnis eintritt. Ein Beispiel ist es, dass so lange die Telefonnummer einer Arztpraxis gewählt wird, bis die Leitung nicht mehr besetzt ist. Allgemeiner gesagt, werden so lange unabhängige Bernoulli-Experimente mit Erfolgswahrscheinlichkeit p durchgeführt, bis zu ersten Mal ein Erfolg eintritt. Ein Elementarereignis ω eines solchen Prozesses kann dann also eine Folge der Form $(0, 0, \dots, 0, 0, 1)$ bestehend aus $k-1$ Nullen und einer Eins dargestellt werden. Sei X die Zufallsvariable, die der Folge $(0, \dots, 0, 1)$ die Zahl k zuordnet. Anders als bei den vorherigen Zufallsvariablen gilt

$X(\Omega) = \mathbb{N} \setminus \{0\}$, das Bild ist also (abzählbar) unendlich.

Satz 5.11. *Es gilt:*

- a) $\Pr(X = k) = (1 - p)^{k-1} \cdot p$
- b) $\Pr(X = k)$ beschreibt eine Wahrscheinlichkeitsverteilung auf $\mathbb{N} \setminus \{0\}$
- c) $E(X) = \frac{1}{p}$

Beweis.

- a) Jedes Ereignis $(X = x)$ beinhaltet genau ein Elementarereignis ω . Aus der Unabhängigkeit der Bernoulli-Experimente folgt direkt

$$\Pr(\underbrace{0, \dots, 0}_{k-1}, 1) = (1 - p)^{k-1} \cdot p$$

- b) Mit der Formel für die geometrische Reihe⁵ gilt:

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} \cdot p = p \cdot \sum_{k=0}^{\infty} (1 - p)^k = p \cdot \frac{1}{1 - (1 - p)} = 1$$

- c) Nun bestimmen wir den Erwartungswert von X . Da im Bild von X die Zahlen $\{1, 2, 3, \dots\}$ sind, können wir Lemma 5.8 verwenden. Es gilt

$$E(X) = \sum_{k=0}^{\infty} \Pr(X > k)$$

Die Wahrscheinlichkeit $\Pr(X > k)$ entspricht der Wahrscheinlichkeit, dass die ersten k Experimente Misserfolge sind. Diese Wahrscheinlichkeit ist $(1 - p)^k$. Damit ergibt sich wieder mit der Formel für die geometrische Reihe:

$$E(X) = \sum_{k=0}^{\infty} (1 - p)^k$$

⁵Für $|x| < 1$ gilt $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$

$$\begin{aligned}
 &= \frac{1}{1 - (1 - p)} \\
 &= \frac{1}{p}
 \end{aligned}
 \quad \square$$

Beispiel. Ein technisches Gerät startet sich mit Wahrscheinlichkeit $\frac{2}{3}$ nach dem Hochfahren direkt wieder neu. Wie häufig wird das Gerät erwartet hochfahren, bis es an bleibt?

Dies kann als geometrische Verteilung mit $p = \frac{1}{3}$ beschrieben werden. Im Erwartungswert wird das Gerät also dreimal hochfahren, bevor es an bleibt. ◀

Beispiel. Ein bekanntes Beispiel in dem geometrisch verteilten Zufallsvariablen auftauchen ist das sogenannte *Coupon Collector Problem*. In diesem gibt es n Sammelkarten, die alle gleichverteilt in Müsliboxen enthalten sind. Die Frage ist, wie viele Boxen im Erwartungswert gekauft werden müssen, bis alle Karten gefunden wurden.

Sei m_i die Anzahl der gekauften Boxen, bis i verschiedene Karten gefunden wurde. Es gilt $m_0 = 0$ und $m_1 = 1$. Sei X die Zufallsvariable, welche die Anzahl an Boxen angibt, die gekauft werden müssen, um alle Karten zu bekommen. Wir unterteilen X in n Zufallsvariablen X_i , welche jeweils die Anzahl der Boxen zählen, die zwischen Zeitpunkt $m_{i-1} + 1$ und m_i gekauft wurden. Es gilt $X = \sum_{i=1}^n X_i$. Jedes X_i ist eine geometrisch verteilte Zufallsvariable mit $p = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$. und $E(X_i) = \frac{n}{n-i+1}$. Mit der Linearität des Erwartungswertes ergibt sich:

$$\begin{aligned}
 E(X) &= \sum_{i=1}^n \frac{n}{n-i+1} \\
 &= n \sum_{i=1}^n \frac{1}{n-i+1} \\
 &= n \sum_{i=1}^n \frac{1}{i} \\
 &= n \cdot H_n \approx n \ln n
 \end{aligned}$$

Dabei ist H_n eine Bezeichnung für $\sum_{i=1}^n \frac{1}{i}$. Diese Summe wird die n -te harmonische Zahl genannt und wächst ungefähr genauso schnell wie der natürliche Logarithmus $\ln n$. ◀

Entwurf

III

Graphentheorie

KAPITEL 6

Graphen

Voraussetzung:

- Mengen Kapitel 3
- Relationen Abschnitt 3.2
- Doppeltes Abzählen Abschnitt 4.4
- Binomialkoeffizienten Abschnitt 4.3



Lernziele

Die Studierenden ...

- ... zeichnen abstrakt gegebene Graphen
- ... zeigen Aussagen zu Graphen
- ... führen Algorithmen auf Graphen durch



In diesem Kapitel beschäftigen wir uns mit Graphen. Graphen sind diskrete Strukturen, welche in der Informatik häufig zum Modellieren von bestimmten Zusammenhängen eingesetzt werden. Ein sehr anschauliches, teilweise aber auch irreführendes Beispiel ist die Modellierung eines Straßennetzes, in dem Kreuzungen mit Straßen verbunden sind.

6.1 Grundlegende Definitionen

Definition 6.1.

- Ein (endlicher) **ungerichteter Graph** ist ein Paar (V, E) bestehend aus einer (endlichen) Knotenmenge V und einer Kantenmenge E von Paaren $e = \{u, v\}$ mit $u, v \in V$.^a

- Ein (endlicher) **gerichteter Graph** ist ein Paar (V, E) bestehend aus einer (endlichen) Knotenmenge V und einer Kantenmenge $E \subseteq V \times V$. Die Kante $e = (u, v)$ ist dann gerichtet von u nach v .

^aManchmal wird für schlichte Graphen auch $E \subseteq \binom{V}{2}$ geschrieben

Zu dieser Definition gibt es einige Anmerkungen:

1. In einem ungerichteten Graph sind die Kante $\{u, v\}$ und $\{v, u\}$ die gleichen Kanten. In einem gerichteten Graph gilt $(u, v) \neq (v, u)$. Dies ist wieder der Unterschied zwischen einer Menge und einem geordneten Paar.
2. Eine Kante von v zu sich selber wird als **Schleife** bezeichnet.
3. Wir betrachten im folgenden Schleifenlose Graphen ohne Mehrfachkanten. Diese werden auch **schlichte Graphen** genannt.

Wir betrachten nun die möglichen Beziehungen der Knoten und Kanten zueinander:

Definition 6.2.

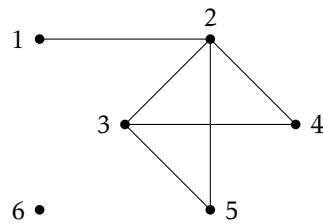
- $u \in V$ ist **adjazent** zu $v \in V$, wenn $\{u, v\} \in E$ in einem ungerichteten Graphen gilt, bzw. wenn $(u, v) \in E$ in einem gerichteten Graphen gilt.
- $e \in E$ ist **inzident** zu $v \in V$, wenn $v \in e$ ist. $v \in V$ ist inzident zu $e \in E$, wenn $v \in e$ gilt. Diese Definition gilt so nur für ungerichtete Graphen

Es gibt verschiedene Standarddarstellungen von Graphen.

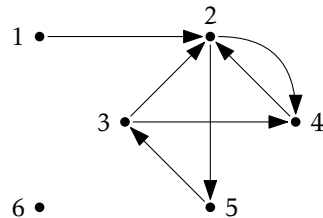
1. Bei der *grafischen* Darstellung werden die Knoten als Punkte in der Ebene dargestellt. Eine Kante zwischen zwei Knoten wird als Strecke oder Kurve dargestellt, siehe Abbildung 6.1 für ein Beispiel.
2. Bei der *Adjazenzliste* wird für jeder der Knoten eine Liste aller adjazenten Knoten angegeben. Die Adjazenzliste verbraucht proportional zu $|V| + |E|$ viel Platz.
3. Die *Adjazenzmatrix* eines Graphen ist eine $|V| \times |V|$ Matrix, bei der im Eintrag an Stelle i, j eine 1 steht, wenn i und j adjazent sind, und sonst eine 0. Die Adjazenzmatrix verbraucht immer $|V| \times |V|$ viel Platz.

Beispiel. Die Adjazenzlistendarstellung für den gerichteten Graphen aus Abbildung 6.1 ist:

1:[2]
2:[4,5]
3:[2,4]
4:[2]



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



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

Abbildung 6.1: Grafische Darstellung eines Graphen

5:[3]

6:[]

Die Adjazenzmatrix für den gleichen Graphen ist:

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	0	1	1	0
3	0	1	0	1	0	0
4	0	1	0	0	0	0
5	0	0	1	0	0	0
6	0	0	0	0	0	0

Selbsttest: Zeichnen Sie eine grafische Darstellung für verschiedene gerichtete und ungerichtete Graphen. Geben Sie dann die Adjazenzlistendarstellung und die Adjazenzmatrixdarstellung für den Graph an.

Tun Sie das Gleiche in dem Sie eine Adjazenzliste oder Adjazenzmatrix angeben und den passenden Graph zeichnen.

Wie können Sie anhand der Adjazenzliste und Adjazenzmatrix erkennen, ob ein Graph ungerichtet ist?

Bevor wir nun tiefer in die weiteren Begriffe einsteigen, nennen wir einige Beispiele für Problemen, die im Zusammenhang mit Graphen definiert werden können.

- Das erste Problem ist das Problem des *Graphenfärben*. Dieses hat als Eingabe einen ungerichteten Graphen. Die Frage ist nun: *Was ist die minimale Anzahl*

an Farben, die nötig ist, damit jeder Knoten gefärbt ist, aber keine zwei adjazenten Knoten die gleiche Farbe haben?

Direkte Anwendungen dieses Problems sind Graphen, welche Funknetzwerke modellieren. Jeder Sender ist ein Knoten, und zwei Knoten sind mit einer Kante verbunden, wenn die Sender sich gegenseitig stören können. Eine Färbung dieses Graphen entspricht nun einer Zuweisung von störungsfreien Funkkanälen zu den Sendern.

Ein sehr bekanntes Problem in diesem Kontext ist der sogenannte *Vierfarbensatz*. Dieser sagt, dass jede Landkarte mit maximal 4 Farben gefärbt werden kann, sodass benachbarte Länder verschiedene Farben haben¹. Der Beweis für diesen Satz wurde 1987 von Appel und Haken mithilfe eines Computers geführt, ein Beweis ohne Computerunterstützung ist bis heute nicht bekannt.

- Eine ganze Klasse von Problemen in Graphen sind *kürzeste Wege Probleme*. Im einfachsten Fall wird nach der kleinsten Anzahl von Kanten gefragt, die durchlaufen werden müssen, um von einem Startknoten zu einem Zielknoten zu kommen. In komplexeren Varianten, haben die Kanten eine Längenangabe, die dann aufsummiert wird, oder es wird nach den kürzesten Wegen zwischen allen Paaren von Knoten gefragt.
- Das *Problem des Handelsreisenden* fragt in einem Graph mit Kantengewichten nach dem kürzesten Weg, der alle Knoten genau einmal besucht. Im Gegensatz zu den meisten kürzesten Wege Problemen, für die es effiziente Algorithmen gibt, ist es nicht bekannt, ob das Problem des Handelsreisenden effizient gelöst werden kann.

Nun betrachten wir weitere Begriffe, die wir im Folgenden benutzen werden.

Definition 6.3. Der **Grad** $\deg_G(v)$ eines Knotens $v \in V$ in einem ungerichteten Graph $G = (V, E)$ ist die Anzahl der adjazenten Knoten von v .

In einem gerichteten Graph $G = (V, E)$, ist der **Ingrad** $\text{indeg}_G(v)$ eines Knoten $v \in V$ die Anzahl der in v ankommenden Kanten (u, v) und der **Ausgrad** $\text{outdeg}_G(v)$ die Anzahl der ausgehenden Kanten (v, u)

Beispiel. In Abbildung 6.1 gilt im ungerichteten Graph $\deg(2) = 4$ und $\deg(6) = 0$. Im gerichteten Graph gilt $\text{indeg}(2) = 3$ und $\text{outdeg}(2) = 2$. ◀

Selbsttest: Bestimmen Sie die Knotengrade für die anderen Knoten im Graph. ?

Definition 6.4. Ein Graph $G' = (V', E')$ ist ein **Untergraph** von $G = (V, E)$, falls $V' \subseteq V$ und $E' \subseteq E$.

¹Benachbarte Länder haben in diesem Fall eine Grenze, die aus mehr als einem Punkt besteht

G' heißt **induzierter Untergraph**, falls außerdem gilt, dass $E' = E \cap \{\{u, v\} \mid u, v \in V'\}$.

Ein induzierter Untergraph ist also ein Untergraph, bei dem alle Kanten, die in G zwischen den Knoten in V' vorhanden sind, in E' sind. Insbesondere ist ein induzierter Untergraph eindeutig durch die Knotenmenge V' bestimmt, da der induzierte Untergraph dann alle Kanten aus E zwischen zwei Knoten in V' enthält.

Um einen ‚normaler‘ Untergraph anzugeben, müssen sowohl die Knotenmenge V' als auch die Kantenmenge E' explizit angegeben werden.

Beispiel. In Abbildung 6.2 ist G' ein Untergraph, aber kein induzierter Untergraph von G . G'' ist der von den Knoten $\{v_2, v_3, v_4, v_5\}$ induzierte Untergraph. ◀

Selbsttest: Zeichnen Sie einen Graph mit mindestens fünf Knoten. Wählen Sie drei Knoten aus, was ist der induzierte Untergraph dieser Knotenmenge. Hat Ihr Graph einen Untergraph auf den gleichen drei Knoten, der kein induzierter Untergraph ist?

Die Summe der Knotengrade in einem ungerichteten Graph lässt sich wie folgt beschreiben:

Lemma 6.1 (Handschlaglemma). In einem schlichten ungerichteten Graph $G = (V, E)$ gilt:

$$\sum_{v \in V} \deg_G(v) = 2|E|$$

Beweis. Sei $G = (V, E)$ mit $V = \{v_1, \dots, v_n\}$ und $E = \{e_1, \dots, e_m\}$ ein schlichter ungerichteter Graph. Betrachte die *Inzidenzmatrix* von G . Dies ist eine Matrix mit m Zeilen und n Spalten. Die Inzidenzmatrix hat eine 1 an Position i, j , wenn $v_j \in e_i$ ist, oder verbal, wenn v_j inzident zu e_i ist.

Die Summe jeder Zeile ist 2, da jede Kante inzident zu genau 2 Knoten ist. Also gibt es $2 \cdot m$ Einsen in der Matrix. In Spalte j ist die Summe der Einsen $\deg_G(v_j)$, da der Knoten v_j inzident zu $\deg_G(v_j)$ Kanten ist. Mit doppeltem Abzählen ergibt sich dann $\sum_{v \in V} \deg_G(v) = 2 \cdot |E|$. ◻

Korollar 6.2. In jedem ungerichteten Graph ist die Anzahl der Knoten mit ungeradem Grad gerade.

Beweis. Dies folgt direkt aus dem Handschlaglemma. Es gilt $\sum_{v \in V} \deg_G(v) = 2|E|$ nach dem Handschlaglemma, also ist die Summe aller Grade gerade. Die Summe der Knotengrade kann aufgeteilt werden in die geraden und die ungeraden Grade. Die Summe der geraden Grade ist gerade, das es sich um eine Summe von geraden Zahlen handelt. Damit die Gesamtsumme gerade sein kann, muss also auch die Summe der ungeraden Grade gerade sein. ◻

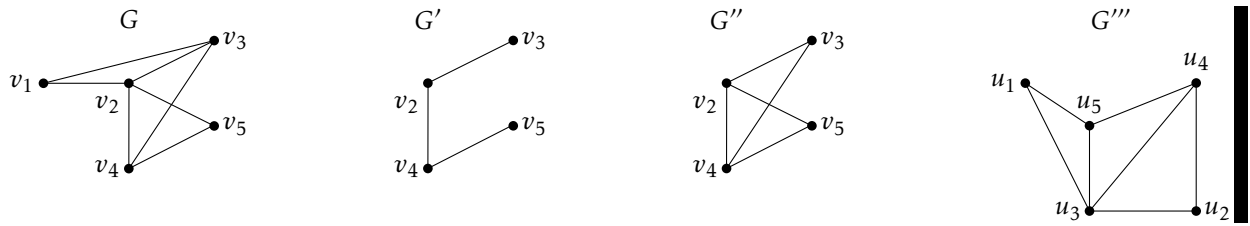


Abbildung 6.2: G' ist Untergraph von G , aber kein induzierter Untergraph. G'' ist induzierter Untergraph von G . G''' ist isomorph zu G .

Definition 6.5. Zwei ungerichtete Graphen $G = (V, E)$ und $G' = (V', E')$ heißen **isomorph**, wenn es eine Bijektion $\Phi : V \rightarrow V'$ gibt, mit der Eigenschaft:

$$\forall u, v \in V : \{u, v\} \in E \Leftrightarrow \{\Phi(u), \Phi(v)\} \in E'$$

Intuitiv betrachtet sind zwei isomorphe Graphen strukturell gleich. Die Funktion Φ kann man sich als eine *Umbenennung* der Knoten aus V vorstellen, sodass diese danach die gleichen Namen haben wie Knoten aus V' . Die Zusatzbedingung gibt an, dass die im umbenannten Graph genau die gleichen Kanten wie in G' vorhanden sein müssen. Es ist nicht bekannt, ob effizient entschieden werden kann, ob zwei Graphen isomorph sind.

Beispiel. Der Graph G''' in Abbildung 6.2 ist isomorph zu G mit der Funktion:

$$\Phi : \{v_1, v_2, v_3, v_4, v_5\} \rightarrow \{u_1, u_2, u_3, u_4, u_5\}$$

$$\Phi(v_1) = u_1$$

$$\Phi(v_2) = u_3$$

$$\Phi(v_3) = u_5$$

$$\Phi(v_4) = u_4$$

$$\Phi(v_5) = u_2$$

Aus einem oder mehreren Graphen können durch Vereinigung oder Schnitt neue Graphen erzeugt werden.

Definition 6.6.

- Für $G_1 = (V_1, E_1)$ und $G_2 = (V_2, E_2)$ ist $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$ und $G_1 \cap G_2 = (V_1 \cap V_2, E_1 \cap E_2)$.
- Der Graph $\overline{G} = (V, \overline{E})$ mit

$$\{u, v\} \in \overline{E} \text{ genau dann, wenn } \{u, v\} \notin E$$

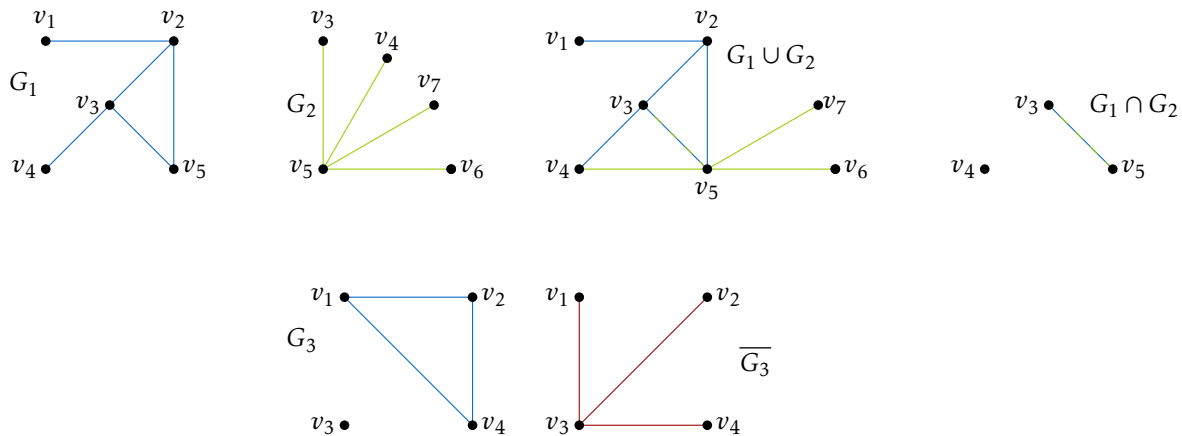


Abbildung 6.3: Darstellung von Schnitt, Vereinigung und Komplement von Graphen

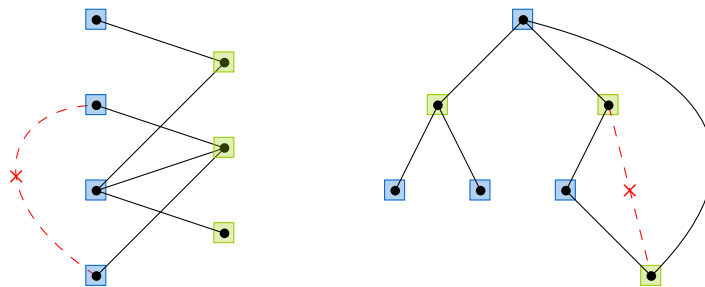


Abbildung 6.4: In Schwarz: zwei Beispiele für bipartite Graphen. Die roten gestrichelten Kanten verletzen die Eigenschaften für bipartite Graphen.

ist der **Komplementärgraph** zu $G = (V, E)$.

Beispiele zu den Definitionen finden sich in Abbildung 6.3.

Selbsttest: Zeichnen Sie zwei Graphen, deren Knotenmengen nicht disjunkt sind. Bestimmen Sie die Vereinigung und den Schnitt der Graphen. Bilden Sie das Komplement von beiden Graphen.

Eine besondere Graphklasse sind die sogenannten *bipartiten Graphen*:

Definition 6.7. Ein ungerichteter Graph $G = (V, E)$ ist **bipartit**, wenn die Knoten in zwei Mengen A und B partitioniert werden können, sodass für alle Kanten $\{u, v\}$ gilt, dass $u \in A$ und $v \in B$, oder $u \in B$ und $v \in A$ gilt.

Beispiele zu bipartiten Graphen finden sich in Abbildung 6.4.

Im Folgenden schauen wir uns einige ungerichtete Graphen an, die eigene Namen haben.

1. Der **vollständige Graph** K_n besteht aus $n \geq 1$ Knoten und allen $\binom{n}{2}$ möglichen Kanten. Jeder Graph mit n Knoten ist Untergraph von K_n .

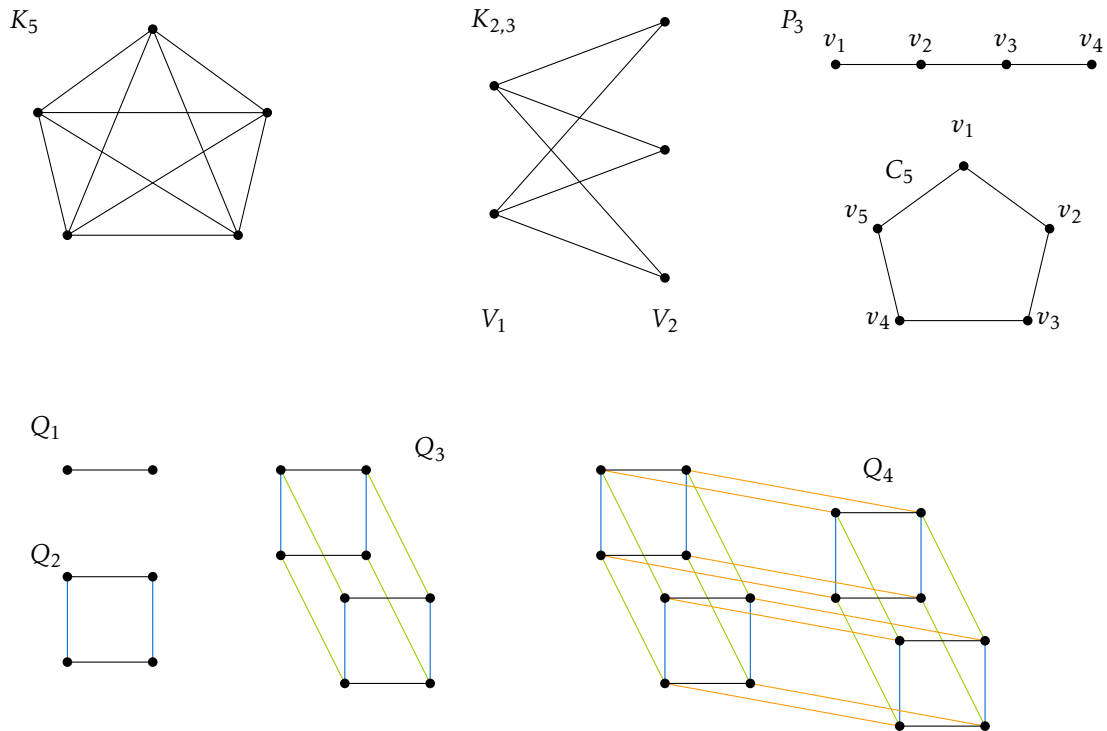


Abbildung 6.5: Beispiele zu besonderen Graphen

2. Der **vollständige bipartite Graph** $K_{n,m}$ besteht aus $n + m \geq 1$ Knoten in zwei disjunkten Mengen V_1 und V_2 und allen möglichen Kanten zwischen V_1 und V_2 . Jeder bipartite Graph ist Untergraph von einem $K_{n,m}$.
3. Der **Hyperwürfel** Q_n halt alle möglichen 0 – 1-Folgen der Länge n als Knoten. Zwei solcher 0 – 1-Folgen sind mit einer Kante verbunden, wenn sich die Folgen an genau einer Stelle unterscheiden.
4. Der **Weg (Pfad)** P_n für $n \geq 0$ besteht aus $n + 1$ Knoten v_1, \dots, v_{n+1} und den n Kanten $\{v_i, v_{i+1}\}$ für $1 \leq i \leq n$. Wir nennen n auch die **Länge** des Wegs.
5. Der **Kreis** C_n für $n \geq 3$ besteht aus n Knoten und den Kanten aus P_{n-1} , zusammen mit der Kante $\{v_n, v_1\}$.

Beispiele für diese Graphen finden sich in Abbildung 6.5.

Definition 6.8. Ein ungerichteter Graph heißt **k-regulär**, wenn alle Knoten Grad k haben.

Beispiel. C_n ist 2-regulär für jedes $n \geq 3$. Genauso gilt, dass K_n $n - 1$ -regulär und Q_n ist ebenfalls n -regulär. Mit den Handschlaglemma ergibt sich dann direkt, dass Q_n genau $n \cdot 2^{n-1}$ Kanten hat. ◀

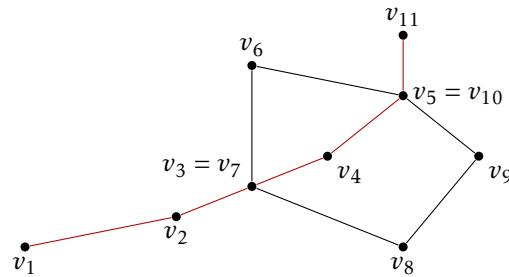


Abbildung 6.6: Ein Kantenzug und eine Unterfolge, die einen Weg bildet (rot)

Wege und Kreise erfüllen eine besondere Funktion als Untergraphen von größeren Graphen, wie wir später sehen werden (Sätze 6.5 und 6.8). Wir definieren die folgenden Verallgemeinerungen, die später in Beweisen nützlich sein werden.

Definition 6.9. Sei $G = (V, E)$ ein ungerichteter Graph. Eine Folge von k Knoten $v_1 v_2 \dots v_k$ heißt **Kantenzug**, falls zwischen zwei aufeinanderfolgenden Knoten immer eine Kante existiert. Es gilt also $\{v_i, v_{i+1}\} \in E$ für alle $i \in \{1, 2, \dots, |V| - 1\}$. Die **Länge** des Kantenzuges ist $k - 1$. Der Kantenzug heißt **geschlossen** falls $v_1 = v_k$.

Ein nicht geschlossener Kantenzug, indem kein Knoten mehrfach vorkommt, heißt **Weg**. Ein geschlossener Kantenzug, indem kein Knoten außer dem ersten/letzten mehrfach vorkommt, heißt **Kreis**.

Wege und Kreise nach Definition 6.9 entsprechen genau Untegraphen P_k bzw. C_k , daher der Name.

Falls in einem Kantenzug $v_1 v_2 \dots v_k$ ein Knoten $v_i = v_j$ mit $i < j$ doppelt vorkommt, dann können wir den Kantenzug „abkürzen“, indem wir v_i und alle Knoten zwischen v_i und v_j löschen. Das Ergebnis ist offensichtlich immer noch ein Kantenzug (siehe Abbildung 6.6). Mehrfaches Wiederholen erzeugt irgendwann einen Weg. Es gilt also

Beobachtung 6.3. Jeder Kantenzug zwischen Knoten u und v hat eine Unterfolge von Knoten, die einen Weg zwischen u und v bildet. Jeder geschlossener Kantenzug hat eine Unterfolge, die einen Kreis bildet.

Speziell ist jeder kürzeste Kantenzug zwischen zwei Knoten u und v ein Weg.

Definition 6.10. Seien u und v Knoten in einem ungerichteten Graphen $G = (V, E)$. Dann ist u von v **erreichbar** in G , wenn es in G einen Weg gibt, der u und v verbindet.

Beobachtung 6.4. Die Relation $R = \{(u, v) \mid u \text{ erreicht } v\}$ ist eine Äquivalenzrelation.

Selbsttest: Beweisen Sie Beobachtung 6.4.



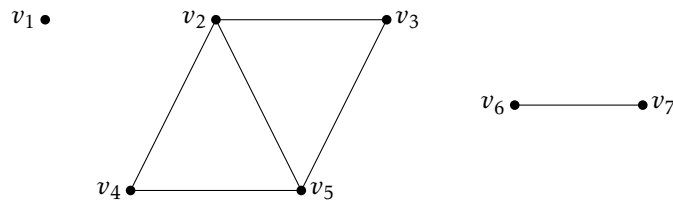


Abbildung 6.7: Ein Graph mit drei Zusammenhangskomponenten

Definition 6.11. Die Äquivalenzklassen von $R = \{(u, v) \mid u \text{ erreicht } v\}$ heißen **Zusammenhangskomponenten**. Ein ungerichteter Graph $G = (V, E)$ heißt **zusammenhängend**, wenn er genau eine Zusammenhangskomponente hat.

Beispiel. Der in Abbildung 6.7 dargestellte Graph hat drei Zusammenhangskomponenten. Der Knoten v_1 kann v_3 nicht erreichen, v_2 erreicht jedoch v_3, v_4 und v_5 .

Definition 6.12. Seien u, v Knoten eines ungerichteten Graphs $G = (V, E)$. Wenn u und v in der gleichen Zusammenhangskomponente liegen, definieren wir den **Abstand** $d_G(u, v)$ als die Länge des kürzesten Weges zwischen u und v . Sind u, v in verschiedenen Zusammenhangskomponenten, ist $d_G(u, v) = \infty$. Der **Durchmesser** eines Graphen ist das Maximum aller paarweisen Abstände.

Beispiel. Im Graph in Abbildung 6.8 gilt $d(v_1, v_3) = 3$ und der Durchmesser des Graphen ist 3.

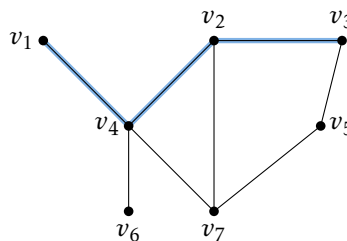


Abbildung 6.8: Im Graph haben v_1 und v_3 Abstand 3. Der Durchmesser des Graphen ist auch 3

Selbsttest: Bestimmen Sie alle paarweise Abstände in Abbildung 6.8.

Nun betrachten wir Eigenschaften von bestimmten Graphklassen, beginnend mit einer Charakterisierung von bipartiten Graphen:

Satz 6.5. Ein ungerichteter Graph ist genau dann bipartit, wenn alle als Untergraph enthaltenen Kreise gerade Länge haben.

Um Satz 6.5 zu beweisen, benötigen wir das folgende Lemma.

Lemma 6.6. Sei $G = (V, E)$ ein ungerichteter Graph und sei $\pi = v_0 v_1 \dots v_m$ ein kürzester Weg (der Länge m) zwischen Knoten v_0 und v_m . Dann ist $v_0 v_1 \dots v_k$ ein kürzester Weg zwischen v_0 und v_k , für alle $k \in \{1, 2, \dots, m\}$.

Beweis. Wir führen einen Beweis durch Widerspruch. Nehmen wir an, es gibt einen kürzeren Weg $v_0 u_1 u_2 \dots u_{\ell-1} v_k$ der Länge $\ell < k$ zwischen v_0 und v_k . Dann hat der Kantenzug $\pi' = v_0 u_1 u_2 \dots u_{\ell-1} v_k v_{k+1} \dots v_m$ Länge $m - k + \ell < m$ und verbindet v_0 und v_m . Laut Beobachtung 6.3 ist also $d_G(v_0, v_m) < m$, ein Widerspruch. \square

Beweis von Satz 6.5. Wir gehen im folgenden davon aus, dass der Graph zusammenhängend ist. Ist dies nicht der Fall, kann die Aussage mit dem folgenden Beweis für jede Zusammenhangskomponente einzeln gezeigt werden. Die zu zeigende Aussage hat die Form $X \Leftrightarrow Y$; wir zeigen, wie häufig, beide Richtungen der Äquivalenz einzeln.

\Rightarrow Sei $G = (V, E)$ bipartit, es gilt also $V = A \cup B$ mit $A \cap B = \emptyset$ und Kanten gehen nur von A nach B . Sei C ein Kreis in G . C benutzt abwechseln Knoten aus A und B , hat also gerade Länge.

\Leftarrow Nun wollen wir zeigen, dass wenn alle Kreise gerade Länge haben, der Graph bipartit ist. Dafür müssen wir (nach Definition) eine Partition von V in Knotenmengen A und B finden, sodass keine Kanten mit beiden Endpunkten in A oder mit beiden Endpunkten in B existieren.

Sei $v \in V$ ein beliebiger Knoten. Wir definieren $A = \{u \in V \mid d_G(v, u) \text{ gerade}\}$ und $B = V \setminus A$. Wir zeigen zuerst mit Hilfe eines Widerspruchsbeweises, dass es keine Kanten zwischen Knoten in A gibt.

Angenommen, es gäbe eine Kante $\{w_1, w_2\}$ mit $w_1, w_2 \in A$. Sei π_1 der kürzeste Weg von v zu w_1 und π_2 der kürzeste Weg von v zu w_2 . Da w_1 und w_2 beide aus A sind, haben π_1 und π_2 beide gerade Länge.

Da es eine Kante zwischen w_1 und w_2 gibt, gilt $|d_G(v, w_1) - d_G(v, w_2)| \leq 1$. Da die beiden Pfade gerade Länge haben, kann nicht gelten, dass $|d_G(v, w_1) - d_G(v, w_2)| = 1$, also ist $d_G(v, w_1) = d_G(v, w_2)$ und damit $|\pi_1| = |\pi_2|$.

Sei z der letzte gemeinsame Knoten von π_1 und π_2 , siehe Abbildung 6.9. Wir zerlegen π_1 in zwei Pfade $\pi'_1 = v \dots z$ und $\pi''_1 = z \dots w_1$, und zerlegen analog π_2 in π'_2 und π''_2 . Nach Lemma 6.6 sind jetzt π'_1 und π'_2 beides kürzeste Wege zwischen v und z , also gilt insbesondere $|\pi'_1| = |\pi'_2|$. Da $|\pi_1| = |\pi_2|$, muss auch $|\pi''_1| = |\pi''_2|$ gelten.

Schließlich beobachten wir, dass π''_1 , π''_2 und die Kante $\{w_1, w_2\}$ zusammen einen Kreis bilden (per Definition von z kann kein Knoten doppelt vorkommen). Die

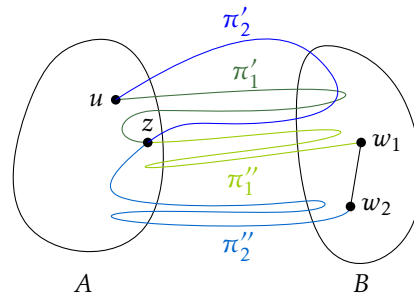


Abbildung 6.9: Visualisierung für die Rückrichtung der Implikation

Länge des Kreises ist $|\pi'_1| + |\pi'_2| + 1 = 2|\pi'_1| + 1$, was eine ungerade Zahl ist; ein Widerspruch.

Das Argument dafür, dass keine zwei Knoten aus B mit einer Kante verbunden sind, ist analog. \square

Selbsttest: Zeigen Sie, dass jeder Baum ein bipartiter Graph ist und dass die Mengen A und B bis auf vertauschung äquivalent sind.



6.2 Bäume

Zusammenhängende Graphen, welche keine Kreise als Untergraph haben spielen eine besondere Rolle in der Graphentheorie.

Definition 6.13. Ein **Baum** ist ein kreisfreier ungerichteter zusammenhängender Graph. Einen kreisfreier ungerichteter Graph, der nicht unbedingt zusammenhängend ist, nennt man einen **Wald**.

Untergraphen die Bäume sind, haben einen besonderen Namen:

Definition 6.14. Sei $G = (V, E)$ ein ungerichteter zusammenhängender Graph mit $|V| = n$. Ein Untergraph T auf allen n Knoten, der ein Baum ist, heißt **aufspannender Baum**.

Analog zu aufspannenden Bäumen in zusammenhängenden Graphen gibt es **aufspannende Wälder** für nicht zusammenhängenden Graphen, die aus einem aufspannenden Baum für jede Zusammenhangskomponente bestehen.

Beispiele zu den Definitionen finden sich in Abbildung 6.10. Aus diesen Definitionen ergibt sich die folgende Beobachtung:

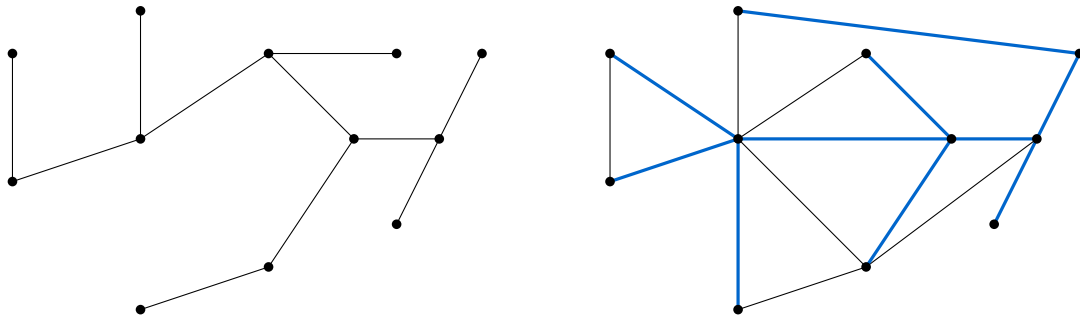


Abbildung 6.10: Ein Baum (links) und ein Graph mit einem blau markierten aufspannenden Baum (rechts)

Beobachtung 6.7. Jeder zusammenhängende Graph hat einen aufspannenden Baum. Dieser Baum ist nur eindeutig, wenn der Graph selbst ein Baum ist. Dann ist der aufspannende Baum der Graph selbst.

Selbsttest: Zeichnen Sie verschiedene Graphen und entscheiden Sie jeweils, ob diese Bäume, Wälder oder nichts von beidem sind. Zeichnen Sie mindestens einen Graph der weder Baum noch Wald ist. Finden Sie verschiedene aufspannende Bäume in diesem Graph.



Der folgende Satz gibt mehrere äquivalente Charakterisierungen von Bäumen.

Satz 6.8. Die folgenden Aussagen sind äquivalent für einen ungerichteten Graphen $G = (V, E)$.

- (1) $G = (V, E)$ ist ein Baum.
- (2) Je zwei Knoten sind durch genau einen Weg verbunden.
- (3) G ist zusammenhängend und es gilt $|E| = |V| - 1$.

Beweis. Wie zeigen die Äquivalenz der Aussagen in dem wir zunächst $(1) \Rightarrow (2)$ und $(2) \Rightarrow (1)$ zeigen und dann $(1) \Rightarrow (3)$ und $(3) \Rightarrow (1)$. Da Äquivalenz eine Äquivalenzrelation ist, folgen die anderen Äquivalenzen dann aus der Transitivität.

(1) \Rightarrow (2) Wir zeigen den ersten Schritt durch einen Widerspruchsbeweis. Angenommen u und v sind nicht durch genau einen Weg verbunden. Dann gibt es zwei Fälle:

1. u und v sind durch keinen Weg verbunden. In diesem Fall ist G nicht zusammenhängend. Dies ist ein Widerspruch zur Annahme, dass G ein Baum ist, da Bäume zusammenhängend sind.
2. u und v sind durch zwei (oder mehr) Wegen verbunden. Fixiere zwei dieser Wege π_1 und π_2 und benenne die Knoten als $\pi_1 = up_1p_2\dots v$ und $\pi_2 =$

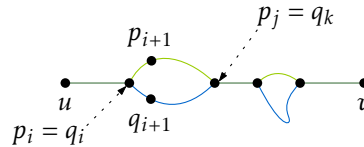


Abbildung 6.11: Darstellung der Bezeichner für Fall 2 in (1) \Rightarrow (2)

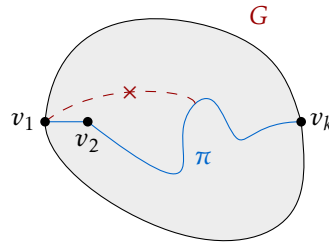


Abbildung 6.12: Darstellung des Arguments, dass v_1 ein Blatt ist

$uq_1q_2\dots v$. Sei i der kleinste Index, sodass $p_i = q_i$ aber $p_{i+1} \neq q_{i+1}$ gilt und $j > i$ und $k > i$ die kleinsten Indizes größer als i für die dann wieder $p_j = p_k$ gilt, siehe Abbildung 6.11 für eine Illustration. Dann ist $p_ip_{i+1}\dots p_j = q_kq_{k-1}\dots q_i$ ein Kreis in G . Dies ist wieder ein Widerspruch zur Annahme, dass G ein Baum ist, da Bäume kreisfrei sind.

(2) \Rightarrow (1) Wir führen wieder einen Widerspruchsbeweis. Angenommen G ist kein Baum, dann ist G nicht zusammenhängend oder nicht kreisfrei.

1. Wenn G nicht zusammenhängend ist, dann gibt es $u, v \in V$, die nicht durch einen Pfad verbunden sind, ein Widerspruch zu (2).
2. Wenn G nicht kreisfrei ist, gibt es einen Kreis $C = v_1v_2\dots v_k$. Dann sind v_1 und v_k einmal durch die Kante $\{v_1, v_k\}$ und einmal durch den Pfad $v_1v_2\dots v_k$ verbunden. Ein Widerspruch zur Annahme, dass jedes Paar von Knoten nur durch einen Pfad verbunden ist.

(1) \Rightarrow (3) Wir wollen zeigen, dass jeder Baum zusammenhängend ist und es gilt $|V| = |E| - 1$. Der erste Teil folgt direkt aus der Definition eines Baumes. Betrachten wir nun also den zweiten Teil.

Als Zwischenschritt zeigen wir zunächst, dass jeder Baum mit $|V| \geq 2$ mindestens einen Knoten mit Grad 1 hat. Einen solchen Knoten nennen wir ein **Blatt**. Das folgende Argument ist in Abbildung 6.12 visualisiert. Sei $\pi = v_1v_2\dots v_k$ ein maximaler Weg in G . Dies meint, dass π ein Weg ist, dem kein Knoten hinzugefügt werden kann. Wir wollen nun argumentieren, dass v_1 ein Blatt ist. Betrachte die adjazenten Knoten von v_1 in G . Alle adjazenten Knoten liegen auf π , weil π sonst kein maximaler Weg wäre. Auf π kann nur v_2 ein Nachbar von v_1 sein, da sonst ein Kreis geschlossen werden würde, also ist v_1 ein Blatt.

Wir entfernen nun v_1 sowie die Kante $\{v_1, v_2\}$ aus dem Baum. Das Ergebnis ist ein neuer Baum $T' = (V', E')$ mit $|V'| = |V| - 1$ und $|E'| = |E| - 1$. Dieses Verfahren

kann wiederholt werden, bis nach $|V| - 1$ Schritten ein Graph mit 1 Knoten und 0 Kanten übrig bleibt. Es wurden also $|V| - 1$ viele Knoten und Kanten entfernt, damit hatte G genau $|V|$ Knoten und $|V| - 1$ Kanten.

(3) \Rightarrow (1) Sei $T = (V, E')$ ein aufspannender Baum von G . Dann ist $|E'| = |V| - 1$. Dies folgt aus (1) \Rightarrow (3). Aber es gilt auch $|E| = |V| - 1$ aus der Bedingung der aktuell betrachteten Falles. Also ist $T = G$. \square

6.3 Graphentraversierung

In diesem Abschnitt betrachten wir *Graphtraversierungen*, also verschiedene Möglichkeiten systematisch alle Knoten und Kanten in einem Graph abzulaufen.

6.3.1 Breitensuche

Die Breitensuche startet die Traversierung an einem Startknoten. Dann werden zunächst alle direkten Nachbarn des Startknotens betrachtet. Danach alle noch nicht betrachteten Nachbarn dieser Nachbarn und so weiter. Um die Knoten in der richtigen Reihenfolge vorzuhalten, werden diese in einer sogenannten Warteschlange (Queue)² zwischengespeichert. Eine Queue ist eine Datenstruktur, in die Elemente eingefügt und aus der Elemente entfernt werden können. Dabei werden die Elemente in einer first-in-first-out (FIFO) Reihenfolge betrachtet. Es kann also jeweils nur das Element, das von allen gespeicherten Elementen zuerst eingefügt wurde, entfernt werden.³

Jeder Knoten hat drei mögliche Zustände: *nicht entdeckt*, *entdeckt* und *abgeschlossen*. Am Anfang sind alle Knoten bis auf den Startknoten *nicht entdeckt*, der Startknoten ist *entdeckt*. Zusätzlich zum Zustand wird für jeden Knoten ein Wert d und ein Wert π gespeichert. Wir werden später (Satz 6.9) sehen, dass der Wert $d[v]$ eines Knoten $v \in V$ nach Abschluss des Algorithmus genau dem Abstand von s zu v entspricht und dann der kürzeste Weg durch die π Werte rekonstruiert werden kann.

Die Breitensuche wiederholt nun in einer Schleife die folgenden Schritte so lange, bis die Warteschlange leer ist. Zunächst wird der erste Knoten u aus der Warteschlange entfernt. Dann werden alle Knoten v mit $\{u, v\} \in E$ bzw. für gerichtete Graphen $(u, v) \in E$ nach und nach betrachtet. Wenn diese als nicht entdeckt markiert sind, werden sie zur Warteschlange hinzugefügt und als entdeckt markiert. Zudem wird $d[v] = d[u] + 1$ und $\pi[v] = u$ gesetzt. Wurden alle benachbarten Knoten von u in dieser Art behandelt, wird u als abgeschlossen markiert. Der Pseudocode findet sich in Algorithmus 1 und ein Beispiel in Abbildung 6.13.

Die Breitensuche kann verwendet werden, um verschiedene algorithmische Fragen zu beantworten.

²Bekannt aus *Konzepte der Programmierung*

³Wie bei einer Warteschlange an einer Kasse im Supermarkt: die Person, die sich zuerst angestellt hat, ist auch als Erstes dran.

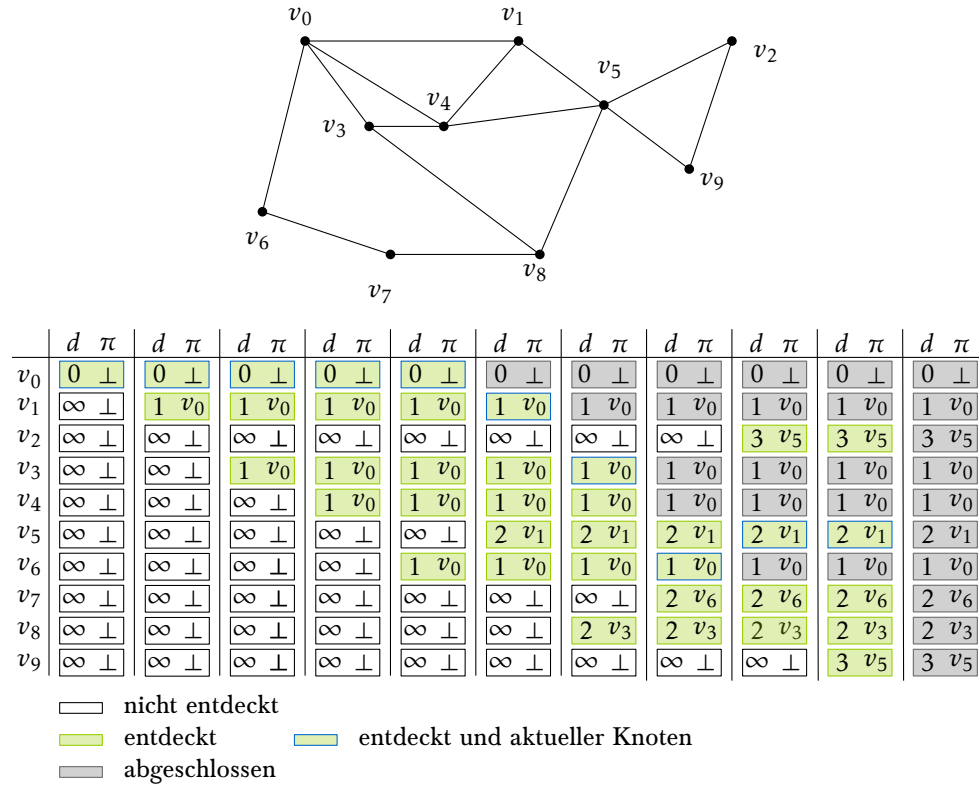


Abbildung 6.13: Beispiel für den Verlauf der d und π Werte bei einer Breitensuche in einem ungerichteten Graph.

Der aktuelle Knoten, wenn sich ein Eintrag ändert, ist markiert. Alle Änderungen von *entdeckt* zu *abgeschlossen*, die zwischen zwei Durchläufen der for-Schleife passieren werden zu einem Schritt zusammengefasst. Am Ende sind alle Knoten *abgeschlossen*.

Algorithm 1 Breitensuche

```
function BREITENSUCHE( $G = (V, E), s \in V$ )  
  ▸ Initialisierung:  
  markiere alle Knoten als nicht entdeckt  
   $Q \leftarrow$  leere Warteschlange  
   $d[v] \leftarrow \infty; \pi[v] \leftarrow \perp$   
   $d[s] \leftarrow 0$   
  markiere  $s$  als entdeckt  
  füge  $s$  in  $Q$  ein.  
  ▸ Hauptschleife:  
  while  $Q \neq \emptyset$  do  
     $u \leftarrow$  entferne ersten Knoten aus  $Q$   
    for  $v$  mit  $\{u, v\} \in E$  bzw.  $(u, v) \in E$  do  
      if  $v$  nicht entdeckt then  
        füge  $v$  zu  $Q$  hinzu  
        markiere  $v$  als entdeckt  
         $d[v] \leftarrow d[u] + 1$   
         $\pi[v] = u$   
    markiere  $u$  als abgeschlossen
```

- In einem ungerichteten zusammenhängenden Graph gibt die Breitensuche einen aufspannenden Baum, $T = (V, E')$ mit $E' = \{\{u, v\} \mid v \in V, u = \pi[v]\}$. Dieser Baum wir **BFS-Baum** genannt.
- Wenn der aufspannende Baum gegeben ist, kann einfach geprüft werden, ob der Graph bipartit ist. Der Baum kann direkt in zwei Mengen A und B unterteilt werden, indem entlang jedes Pfades von s zu einem Knoten die Knoten abwechselnd in die Menge A und B eingefügt werden. Nun kann für alle Kanten, die nicht im Baum sind, geprüft werden, ob diese zwischen zwei Knoten aus der gleichen Menge verlaufen. Ist dies der Fall, kann der Graph nicht bipartit sein.
- Wie der Pseudocode andeutet, kann der Algorithmus auch auf gerichteten Graphen angewendet werden.

Selbsttest: Zeichnen Sie Graphen mit mindestens 8 Knoten und 12 Kanten und führen Sie die Breitensuche auf diesen aus. Betrachten Sie gerichtete und ungerichtete Graphen.



Für den folgenden Satz erweitern wir zunächst noch die Begriffe des Weges und des Abstands auf gerichtete Graphen.

Definition 6.15. Ein **gerichteter Weg** ist ein gerichteter Graph $G = (V, E)$ mit $V = v_1, \dots, v_{n+1}$ und $E = \{(v_i, v_{i+1}) \mid 1 \leq i \leq n\}$.

Definition 6.16. Sei $G = (V, E)$ ein gerichteter Graph und $u, v \in V$. Der Abstand $d_G(u, v)$ ist die Länge des kürzesten gerichteten Weges zwischen u und v .

Es ist zu beachten, dass in gerichteten Graphen gelten kann, dass $d_G(u, v) \neq d_G(v, u)$. Ein einfaches Beispiel dafür ist ein gerichteter Kreis mit 3 Knoten.

Selbsttest: Wie muss die Breitensuche angepasst werden, so dass *alle* Knoten durchlaufen werden und nicht nur solche, die vom Startknoten mit einem (gerichteten) Weg erreicht werden können?

Wir zeigen nun eine algorithmisch wichtige Eigenschaft der Breitensuche:

Satz 6.9. Nach Ende der Breitensuche mit Startknoten s gilt $d[v] = d_G(s, v)$ für alle $v \in V$.

Beweis. Zunächst einmal kann festgestellt werden, dass sich der d -Wert jedes Knoten genau einmal ändert, und zwar genau an dem Zeitpunkt wo er von ∞ auf einen Wert gesetzt wird und der Knoten danach in Q eingefügt wird. Das ist genau der Zeitpunkt, an dem der Knoten *entdeckt* wird.

Wir führen den weiteren Beweis in drei Schritten. Der erste Schritt zeigt, dass die Reihenfolge, in der die Knoten in Q eingefügt werden, mit ihren d -Werten am Ende des Algorithmus zusammenhängen. Dann wird gezeigt, dass $d[v] \geq d_G(s, v)$ gilt und abschließend dass $d[v] \leq d_G(s, v)$ gilt. Zusammen ergibt sich dann $d[v] = d_G(s, v)$, wie gefordert.

1. Wir zeigen: Wenn u vor v in Q steht, gilt $d[v] \leq d[u] + 1$. Wir zeigen diesen Zusammenhang mittels vollständiger Induktion über die Anzahl der Einfügeoperationen in Q . Für den Anker haben wir $n = 1$ Einfügeoperationen. In dem Fall ist in Q nur s enthalten und die Aussagen stimmen. Die Induktionsvoraussetzung ist, dass die Aussage nach n Einfügeoperationen gilt. Für den Schritt gehen wir von n nach $n + 1$.

Wir betrachten nun drei besondere Knoten: Sei v der Knoten, der in der $n + 1$ -sten Einfügeoperation eingefügt wird. Dann ist u der Knoten, der in diesem Durchlauf der Hauptschleife aus Q entfernt wurde und h das vorderste Element der Hauptschleife, nachdem v entfernt wurde. Da v wegen u eingefügt wird, gilt $d[v] = d[u] + 1$. Da h und u gleichzeitig in der Warteschlange waren, gilt nach Induktionsvoraussetzung $d[u] \leq d[h] + 1$. Wir betrachten nun zwei Fälle.

- a) Ist $d[h] = d[u]$, dann gilt auch $d[v] = d[h] + 1$. Da die restlichen Knoten der Warteschlange nach Induktionsvoraussetzung auch alle nur Werte $d[h]$ oder $d[h] + 1$ haben können, gilt die Aussage in diesem Fall.

- b) Ist $d[h] = d[u] + 1$, dann gilt auch $d[v] = d[u] + 1 = d[h]$. Da u gerade erst entfernt wurde haben die restlichen Knoten in der Warteschlange nach der Induktionsvoraussetzung Wert $d[h]$ und auch in diesem Fall gilt die Aussage.
2. Nun zeigen wir, dass $d[v] \geq d_G(s, v)$ gilt, also dass der d -Wert den Abstand nicht unterschätzt. Der Beweis erfolgt wieder mittels vollständiger Induktion über die Anzahl der Einfügeoperationen in Q . Für den Anker betrachten wir wieder die erste Einfügeoperation. Hier gilt $d[s] = 0 = d_G(s, s)$ und die Aussage ist wahr. Die Induktionsvoraussetzung ist, dass für die Knoten, die in allen Einfügeoperationen bis einschließlich Operation n eingefügt wurden die Aussage gilt. Im Schritt folgern wir nun, dass die Aussage dann auch für Operation $n + 1$ gilt.

Sei v der $n + 1$ -ste Knoten, der zu Q hinzugefügt wurde und sei u der Knoten von dem aus v entdeckt wurde. Nach Induktionsvoraussetzung gilt nun $d[u] = d_G(s, u)$ und nach dem Algorithmus gilt $d[v] = d[u] + 1$. Da v von u aus entdeckt wurde, gibt es einen Weg von s nach v mit Länge $d[u] + 1$, also ist $d_G(s, v) \leq d[v]$.

3. Nun zeigen wir, dass auch $d[v] \leq d_G(s, v)$ gilt, also dass die d -Werte die tatsächlichen Abstände nicht überschätzen. Hier führen wir die Induktion über $d_G(s, v)$, also über den tatsächlichen Abstand eines Knotens von s . Für den Anker betrachten wir alle Knoten mit Abstand 0 zu s . Dies ist nur der Knoten s selbst und die Aussage folgt aus der Initialisierung. Die Induktionsvoraussetzung ist, dass die Aussage für alle Knoten mit Abstand n zu s gilt. Im Schritt folgern wir daraus, dass Aussage auch für alle Knoten mit Abstand $n + 1$ gilt.

Sei v ein Knoten mit $d_G(s, v) = n + 1$ und sei z der Vorgänger von v auf einem kürzesten Weg in G . Dann gilt:

$$d_G(s, v) = d_G(s, z) + 1 = d[z] + 1$$

wobei die letzte Gleichung aus der Induktionsvoraussetzung und 2. folgt. Wir betrachten nun zwei Fälle. Im ersten Fall war v schon in Q , als z bearbeitet wurde. In diesem Fall gilt mit 1.

$$d[v] \leq d[z] + 1 = d_G(s, v)$$

Im zweiten Fall wurde v wegen z in die Warteschlange eingefügt und es gilt wegen des Algorithmus, dass

$$d[v] = d[u] + 1 = d_G(s, v). \quad \square$$

6.3.2 Tiefensuche

Der zweite Traversierungsalgorithmus den wir betrachten ist die *Tiefensuche*. Hier ist die Idee, dass einem Weg im Graph so lange gefolgt wird, bis keine neuen Knoten

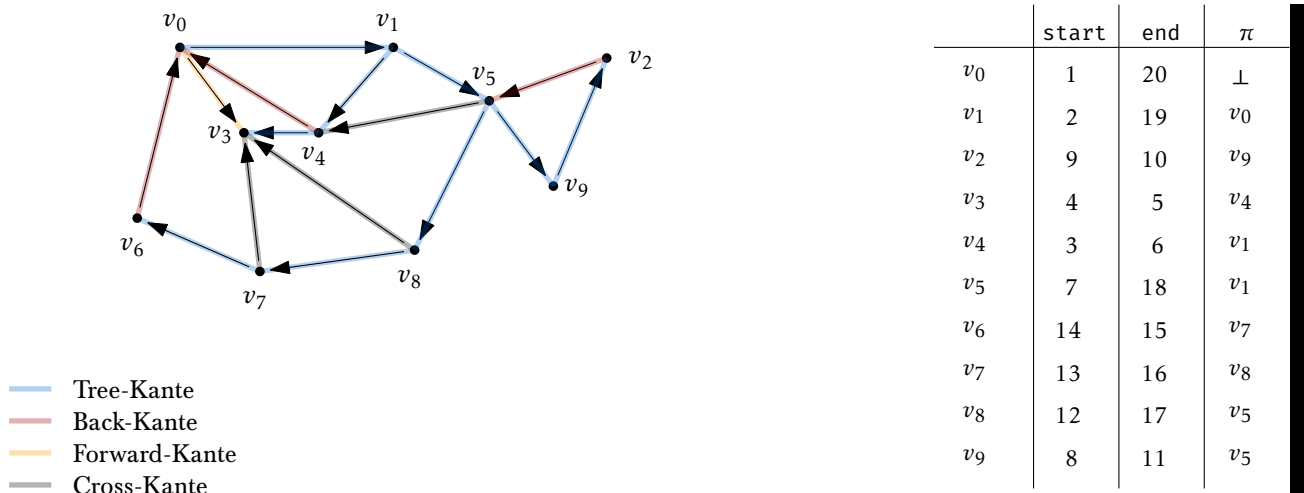


Abbildung 6.14: Beispiel Tiefensuche

mehr entdeckt werden. Dann wird der Weg so lange zurückverfolgt, bis es wieder nicht besuchte Nachbarn gibt. Für die Breitensuche hatten wir explizit eine Queue in der wir entdeckte Nachbarn, die jedoch noch nicht abschließend besucht wurden, gespeichert haben. Für die Tiefensuche benutzen wir Rekursion und damit implizit einen Stack für die gleiche Aufgabe. Ein Stack ist eine Datenstruktur, aus der immer nur das zuletzt hinzugefügte Element entfernt werden kann.⁴ Wir benutzen den Stack nur implizit, weil die Reihenfolge durch rekursive Aufrufe und nicht durch die Verwendung eines Stacks definiert ist.

Die Tiefensuche besteht so aus zwei Methoden, die erste führt eine Initialisierung durch und die zweite übernimmt mit rekursiven Aufrufen eine ähnliche Aufgabe wie die Hauptschleife in der Breitensuche. In dieser Initialisierung werden Werte $start$, $ende$ und π für jeden Knoten mit ∞ bzw. \perp initialisiert und ein globaler Zähler $zeit$ mit 0 angelegt. Dann wird eine rekursive Funktion `DFS-visite` mit dem Startknoten aufgerufen. In `DFS-visite` wird der aufrufende Knoten u als entdeckt markiert und $zeit$ wird um eins erhöht. Dann wird $start[u] = zeit$ gesetzt und alle zu u benachbarten Knoten v werden betrachtet. Sind diese noch nicht entdeckt, wird $\pi[v] = u$ gesetzt und `DFS-visite` wird mit v aufgerufen. Nachdem alle Nachbarn rekursiv abgearbeitet wurden, wird $zeit$ um eins erhöht und $ende$ auf $zeit$ gesetzt. Der Pseudocode findet sich in Algorithmus 2 und ein Beispiel mit $start$ und $ende$ für jeden Knoten in Abbildung 6.14.

Die Tiefensuche kann für ähnliche Anwendungen wie die Breitensuche verwendet werden:

- In einem ungerichteten zusammenhängenden Graph gibt die Tiefensuche einen aufspannenden Baum, $T = (V, E')$ mit $E' = \{\{u, v\} \mid v \in V, u = \pi[v]\}$. Dieser Baum wird **DFS-Baum** genannt.

⁴Beide Konzepte, also Stacks und Rekursion werden in *Konzepte der Programmierung* behandelt.

Algorithm 2 Pseudocode für die Tiefensuche

```
function TIEFENSUCHE( $G = (V, E), s \in V$ )
    markiere alle Knoten als nicht entdeckt
     $\text{start}[v] \leftarrow \infty; \text{ende}[v] \leftarrow \infty; \pi[v] \leftarrow \perp$ 
     $\text{zeit} \leftarrow 0$ 
    DFS-VISIT( $s$ )

function DFS-VISIT( $u$ )
    markiere  $u$  als entdeckt
     $\text{zeit} = \text{zeit} + 1$ 
     $\text{start} = \text{zeit}$ 
    for  $v$  mit  $\{u, v\} \in E$  bzw.  $(u, v) \in E$  do
        if  $v$  nicht entdeckt then
             $\pi[v] = u$  DFS-VISIT( $v$ )
    markiere  $u$  als abgeschlossen
     $\text{zeit} = \text{zeit} + 1$ 
     $\text{ende} = \text{zeit}$ 
```

- Wenn der aufspannende Baum gegeben ist, kann einfach geprüft werden, ob der Graph bipartit ist. Der Baum kann direkt in zwei Mengen A und B unterteilt werden, indem entlang jedes Pfades von s zu einem Knoten die Knoten abwechselnd in die Menge A und B eingefügt werden. Nun kann für alle Kanten, die nicht im Baum sind geprüft werden, ob diese zwischen zwei Knoten aus der gleichen Menge verlaufen. Ist dies der Fall, kann der Graph nicht bipartit sein.
- Wie der Pseudocode andeutet, kann der Algorithmus auch auf gerichteten Graphen angewendet werden.

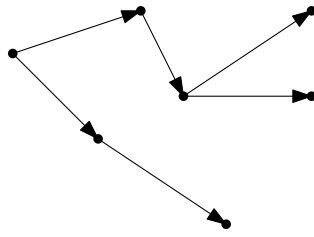
Abstände können mit der Tiefensuche jedoch nicht bestimmt werden.

Wir klassifizieren die Kanten nach dem Zustand der Endpunkte, wenn eine Kante zum ersten Mal betrachtet wird. Ist der Startknoten schon entdeckt, der Zielknoten aber noch nicht, ist die Kante eine **Tree-Kante**. Dies sind auch genau die Kanten, die im DFS-Baum enthalten sind. Geht die Kante zwischen zwei Knoten, die schon entdeckt aber noch nicht abgeschlossen sind, sprechen wir von einer **Back-Kante**. Geht die Kante von einem entdeckten zu einem abgeschlossenen Knoten, wobei der entdeckte Knoten ein Vorfahr des abgeschlossenen Knotens im DFS-Baum ist die Kante eine **Forward-Kante**, alle anderen Kanten heißen **Cross-Kanten**. Diese Konzepte sind in Abbildung 6.14 visualisiert.

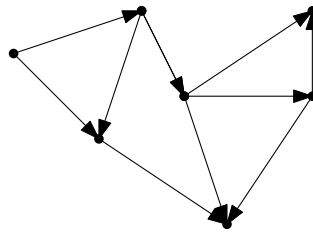
Selbsttest: Beantworten Sie die gleichen Frage wie für die Breitensuche.



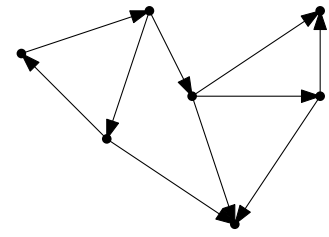
Es gilt jedoch noch die folgende Beobachtung, welche uns im weiteren Verlauf noch helfen wird:



DAG



DAG



kein DAG

Abbildung 6.15: Beispiel für Graphen die DAGs sind und für Graphen, die keine DAGs sind

Beobachtung 6.10. Für jedes Paar von Knoten mit $start[v] \leq start[u]$ gilt:

- $[start[u], ende[u]] \subseteq [start[v], ende[v]]$ oder
- $[start[u], ende[u]] \cap [start[v], ende[v]] = \emptyset$.

Die erste Aussage gilt, wenn v Vorgänger von u im DFS-Baum ist. Sonst gilt die zweite Aussage.

6.4 Gerichtete azyklische Graphen

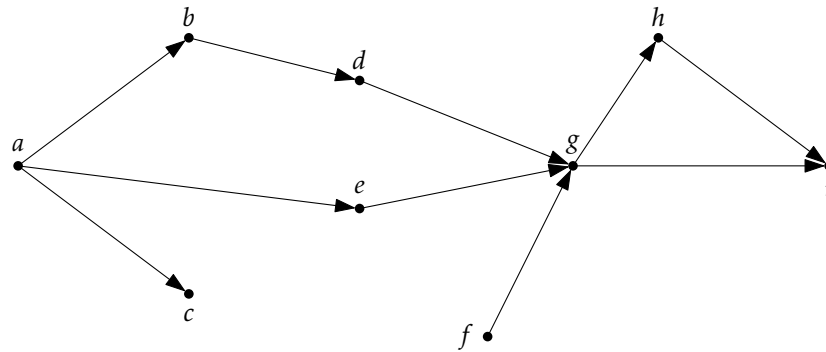
Wir hatten einen Baum als einen zusammenhängenden ungerichteten kreisfreien Graph definiert. Nun betrachten wir kreisfreie gerichtete Graphen.

Definition 6.17. Ein **gerichteter azyklischer Graph** (DAG) ist ein gerichteter Graph ohne gerichtete Kreise.

Beispiele für Graphen, die DAGs sind und für Graphen die keine DAGs sind finden sich in [Abbildung 6.15](#).

DAGs werden häufig dafür verwendet, um Abhängigkeiten zum Beispiel in Projekten oder Prozessen zu modellieren. Dabei sind dann die einzelnen Aufgaben, die zu erledigen sind die Knoten im Graph und es gibt eine gerichtete Kante von Aufgabe a zu Aufgabe b , wenn Aufgabe a vor Aufgabe b erledigt werden muss. Eine natürliche Fragestellung ist dann eine Reihenfolge, in der die Aufgaben abgearbeitet werden können, ohne dass die Abhängigkeiten verletzt werden.

Der Bezug zur Informatik wird klar, wenn die Prozesse als Prozesse im Computer betrachtet werden. Bilden diese einen gerichteten Kreis, gibt es einen sogenannten Deadlock und kein Prozess kann ausgeführt werden.



Mögliche topologische Sortierung: $a \leq f \leq c \leq e \leq b \leq d \leq g \leq h \leq i$

Abbildung 6.16: Ein gerichteter Graph und eine zugehörige topologische Sortierung

Definition 6.18. Eine **topologische Sortierung** eines gerichteten Graphen ist eine totale Ordnung (V, \leq) , sodass $(u, v) \in E \Rightarrow u \leq v$.

Um bei der Analogie zu bleiben ist eine topologische Sortierung eine Reihenfolge in der die Prozesse ausgeführt werden können, ohne dass die Abhängigkeiten verletzt werden. In in Abbildung 6.16 findet sich ein Graph mit einer möglichen zugehörigen topologischen Sortierung.

Es gilt die folgende Beobachtung, welche direkt an einem gerichteten Kreis mit drei Knoten gesehen werden kann.

Beobachtung 6.11. Wenn G gerichtete Kreise hat, gibt es keine topologische Sortierung für G .

Umgekehrt werden wir nun zeigen, dass für jeden DAG eine topologische Sortierung gefunden werden kann. Genauer zeigen wir zunächst das folgende:

Satz 6.12. Jeder DAG hat mindestens eine Quelle (Knoten mit Ingrad 0) und eine Senke (Knoten mit Ausgrad 0).

Beweis. Wähle einen Knoten v_0 in V . Ist v_0 bereits eine Senke haben wir eine Senke gefunden und sind fertig. Sonst gibt es einen Knoten v_1 mit $(v_0, v_1) \in E$. Dieser Prozess kann wiederholt werden. Da es keine Kreise gibt, kann es nicht passieren, dass wir irgendwann zu einem Knoten zurückkehren, also endet der Prozess nach maximal $|V|$ Schritten in einer Senke. Mit einem analogen Argument kann auch eine Quelle gefunden werden. \square

Dies gibt auch direkt einen Algorithmus zum Finden einer topologischen Sortierung: Wähle eine beliebige Quelle, gebe dieser Quelle den nächsten Wert, lösche die Quelle und alle anliegenden Kanten und wiederhole bis alle Knoten bearbeitet wurden.

Selbsttest: Vollziehen Sie den Beweis von Satz 6.12 am Beispiel aus Abbildung 6.16 nach.



Dies ist allerdings kein besonders effizientes Verfahren. Schneller kann eine topologische Sortierung mit Verwendung der Tiefensuche gefunden werden. Zunächst beweisen wir die folgende Eigenschaft.

Lemma 6.13. *Ein Graph G ist ein DAG, genau dann, wenn es keine Back-Kanten beim Ausführen der Tiefensuche gibt.*

Beweis. Wir unterteilen den Beweis der Äquivalenz in zwei Implikationen. Zunächst zeigen wir, dass ein DAG keine Back-Kanten hat durch einen Widerspruchsbeweis. Angenommen, G hat eine Back-Kante (u, v) . Dann wurden u und v beide schon entdeckt, aber keiner der Knoten ist abgeschlossen. Es gilt $\text{start}[u] > \text{start}[v]$ und wir befinden uns im ersten Fall von Beobachtung 6.10. Damit ist v Vorfahr von u im DFS Baum und die Kante (u, v) schließt einen gerichteten Kreis.

Nun zeigen wir, dass wenn G keine Back-Kanten hat, G ein DAG ist. Wir führen wieder einen Beweis durch Widerspruch. Wir nehmen also an, dass G kein DAG ist, dann gibt es einen gerichteten Kreis $C = v_1, \dots, v_k$ in G . Sei v_i der Knoten in C mit kleinstem start -Wert. Dann ist $\pi = v_{i+1}, \dots, v_k, v_1, \dots, v_{i-1}$ ein Pfad in G der zum Zeitpunkt an dem v_i entdeckt wird aus nicht entdeckten Knoten besteht. Während der Tiefensuche werden alle Knoten auf π entdeckt, bevor v_i abgeschlossen wird. Hierbei muss π jedoch kein Teilpfad des DFS-Baums sein, die Knoten können auch in einer anderen Reihenfolge abgearbeitet werden. In jedem Fall wird v_{i-1} entdeckt, bevor v_i abgeschlossen ist und damit ist (v_{i-1}, v_i) eine Back-Kante. \square

Wir können nun Lemma 6.13 benutzen, um eine alternative Regel anzugeben um eine topologische Sortierung zu finden:

Satz 6.14. *Wenn die Tiefensuche so lange mit einem neuen, nicht entdeckten Knoten wiederholt wird, bis alle Knoten entdeckt wurden und dabei eine gemeinsame Zeit verwendet, dann ist die Reihenfolge der Knoten nach ihren end-Werten sortiert eine umgekehrte topologische Sortierung in einem DAG G .*

Beweis. Um zu zeigen, dass diese Reihenfolge eine umgekehrte topologische Sortierung gibt, müssen wir nach der Definition der topologischen Sortierung also zeigen, dass $(u, v) \in E \Rightarrow \text{end}[v] \leq \text{end}[u]$ gilt. Wir betrachten den Zeitpunkt, an dem die Kante (u, v) bearbeitet wird. Nach der Definition der Tiefensuche ist u zu diesem Zeitpunkt schon entdeckt. Wir unterteilen drei Fälle je nachdem welchen Zustand v hat.

1. v ist auch entdeckt, aber noch nicht abgeschlossen, dann wäre (u, v) eine Back-Kante. Da der Graph aber ein DAG ist, ist dies nach Lemma 6.13 nicht möglich.
2. v ist nicht entdeckt, dann wird $\text{end}[v]$ im rekursiven Aufruf vor $\text{end}[u]$ gesetzt.
3. v ist abgeschlossen. Dann ist $\text{end}[v] \leq \text{end}[u]$ nach Definition. \square

Anmerkung: Satz 6.12 hat starke Ähnlichkeiten zur Struktur von Satz 3.6. Dies ist kein Zufall, es gibt tatsächlich starke Parallelen zwischen DAGs und Halbordnungen. Es gelten die folgenden Beobachtungen:

1. Ist (M, \leq) eine Halbordnungsrelation, dann ist $G = (M, E)$, $E = \{(u, v) \mid u \leq v\}$ ein DAG. Eine lineare Erweiterung von (M, \leq) ist dann eine topologische Sortierung.
2. Andersrum definiert nicht jeder DAG eine Halbordnung. Allerdings ist die gerichtete Erreichbarkeit eine Halbordnungsrelation. Das bedeutet, dass in einem DAG die Relation $R = \{(u, v) \mid \exists \text{ gerichteter Weg von } u \text{ nach } v\}$ eine Halbordnung ist. Eine lineare Erweiterung von R ist dann eine topologische Sortierung von G .
Verbunden mit dieser Überlegung ist der Begriff des **transitiven Abschlusses**. Der transitive Abschluss eines DAG $G = (V, E)$ ist der Graph $G^* = (V, E^*)$ mit $E^* = \{(u, v) \mid \exists \text{ gerichteter Weg von } u \text{ nach } v\}$. Jede topologische Sortierung von G ist auch eine topologische Sortierung von G^* .

Selbsttest: Was sind Unterschiede und Gemeinsamkeiten der Traversierungsalgorithmen? Welche algorithmischen Fragestellungen lassen sich mit beiden Strategien lösen und welche nur mit einer der Strategien?

6.5 Planare Graphen

In diesem Abschnitt beschäftigen wir uns mit einer besonderen Klasse von Graphen, den sogenannten *planaren Graphen*. Dies sind Graphen, die so auf einem Blatt Papier gezeichnet werden können, ohne dass sich die Kanten schneiden.

Definition 6.19. Ein ungerichteter Graph heißt **planar**, wenn er sich so in der Ebene (\mathbb{R}^2) zeichnen lässt, dass sich Kanten nur in gemeinsamen Endknoten berühren.

Eine **Zeichnung** ist die Zuordnung von Knoten zu Punkten und Kanten zu stetigen Kurven ohne Selbstschnitte, die die entsprechenden Punkte verbinden.

Eine **Region** oder **Facette** einer Zeichnung ist eine maximale Menge an Punkten, die voneinander erreichbar sind, ohne eine Kante zu kreuzen.

Ein Beispiel für einen planaren Graphen mit einer zugehörigen Zeichnung und den von der Zeichnung definierten Regionen findet sich in Abbildung 6.17. Hierbei ist Region 1 eine unbeschränkte Region.

Mit diesen Definitionen lässt sich nun ein Zusammenhang zwischen der Anzahl der Knoten, Kanten und Facetten eines zusammenhängenden planaren Graphen zeigen.

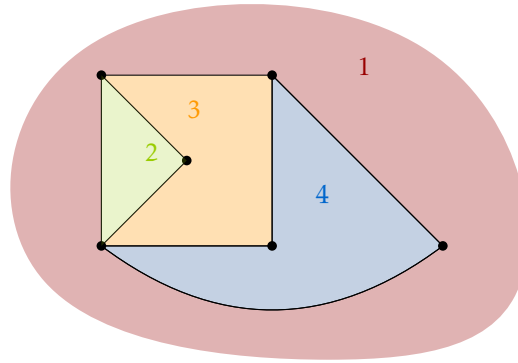


Abbildung 6.17: Ein planarer Graph mit einer planaren Zeichnung und den vier von der Zeichnung definierten Regionen.

Satz 6.15 (Eulerscher Polyedersatz). Sei $G = (V, E)$ ein zusammenhängender planarer Graph mit $V \neq \emptyset$. Für jede Zeichnung von G gilt:

$$v - e + f = 2$$

Wobei $v = |V|$, $e = |E|$ und f die Anzahl der Facetten angibt.

Beweis. Wir zeigen die Aussage mit vollständiger Induktion über die Anzahl e der Kanten. Für den Anker ist $e = 0$. Da G zusammenhängend ist bedeutet dies, dass G aus nur einem Knoten besteht und die Zeichnung nur die unbegrenzte Facette hat. Es gilt also $v - e + f = 1 - 0 + 1 = 2$.

Die Induktionsvoraussetzung ist, dass die Aussage für beliebige Graphen mit maximal e Kanten gilt. Im Schritte gehen wir von e nach $e + 1$ Kanten und betrachten eine Fallunterscheidung.

1. G ist ein Baum. Dann hat G genau $e + 2$ Knoten und eine Facette⁵. Es gilt also $e + 2 - e + 1 + 1 = 2$.
2. G ist kein Baum, also hat G mindestens einen Kreis C als Untergraph. Sei T ein aufspannender Baum von G und k eine Kante von C , die nicht in T ist. Wird nun k aus der Zeichnung von G entfernt, entsteht eine Zeichnung eines Graphen G' mit $f - 1$ Facetten und e Kanten. Es gilt also mit der Induktionsvoraussetzung für die letzte Gleichung:

$$v - e + (f - 1) = 2$$

Mit Umstellen der linken Seite ergibt sich genau die Eigenschaft, die wir zeigen wollen:

$$v - (e + 1) + f = 2$$

Hierbei haben wir den sogenannten *Jordanschen Kurvensatz* verwendet, der besagt, dass jede doppelpunktfreie geschlossene Kurve in der Ebene eine Region in

⁵Dies nehmen wir an dieser Stelle ohne Beweis hin

zwei Facetten unterteilt. Diese Kurve ist an der Stelle die Einbettung des Kreises C , welche durch das Entfernen von k unterbrochen wird. \square

Selbsttest: Geben Sie eine (abstrakte) Zeichnung, die den Sachverhalt aus Satz 6.15 visualisiert. Machen Sie sich insbesondere klar, welche Werte von e , f und v im Induktionsschritt verwendet werden.



Aus dem Eulerschen Polyedersatz folgt direkt eine lineare obere Schranke für die Anzahl der Kanten in einem planaren Graphen. Ein planarer Graph hat also eine Anzahl an Kanten, die linear von der Anzahl der Knoten beschränkt ist.

Korollar 6.16. *Ein planarer Graph $G = (V, E)$ mit $|V| \geq 3$ hat höchstens $3 \cdot |V| - 6$ Kanten.*

Beweis. Wir nehmen zunächst an, dass G zusammenhängend ist. Mit dem Eulerschen Polyedersatz gilt $|E| = |V| + f - 2$, wobei f die Anzahl der Facetten in einer Zeichnung von G ist. Für eine feste Anzahl an Knoten, ist die Anzahl der Kanten also maximiert, wenn es möglichst viele Facetten gibt. Wenn es eine Facette gibt, die nur von mehr als drei Kanten begrenzt ist, kann diese durch das Einfügen einer weiteren Kante in zwei Facetten unterteilt werden. Daher haben wir die maximale Anzahl an Facetten, wenn jede Facette von drei Kanten begrenzt ist. Mit doppeltem Abzählen ergibt sich dann:

$$\begin{aligned} 2 \cdot |E| &= 3 \cdot f & \text{also} \\ f &= \frac{2 \cdot |E|}{3} \end{aligned}$$

Es ergibt sich:

$$\begin{aligned} |E| &= |V| + \frac{2 \cdot |E|}{3} - 2 \\ \Leftrightarrow 3 \cdot |E| &= 3 \cdot |V| + 2 \cdot |E| - 6 \\ \Leftrightarrow |E| &= 3 \cdot |V| - 6 \end{aligned}$$

Nun betrachten wir den Fall, dass G unzusammenhängend ist. Für jede Zusammenhangskomponente $C = (V', E')$ von G gilt $|E'| \leq 3 \cdot |V'| - 6$ falls $|V'| \geq 3$, und trivialerweise $|E'| \leq |V'| \leq 3 \cdot |V'|$ falls $|V'| \in \{1, 2\}$. Falls G also mindestens eine Zusammenhangskomponente mit mindestens drei Knoten hat, dann können wir die Knoten- und Kantenanzahlen einfach aufsummieren, um $|E| \leq 3 \cdot |V| - 6$ zu erhalten. Andernfalls liefert aufsummieren $|E| \leq |V|$, woraus $|E| \leq 3 \cdot |V| - 6$ folgt, da $|V| \geq 3$. \square

Der weiter oben erwähnte Vierfarbensatz hat eine direkte Beziehung zu planaren Graphen. Wenn in jedes Land ein Knoten gezeichnet wird und zwei Knoten mit einer Kante verbunden werden, wenn die Länder eine gemeinsame Grenze haben ergibt sich ein planarer Graph. Der Vierfarbensatz sagt allgemeiner also aus, dass jeder planare Graph mit maximal vier Farben so gefärbt werden kann, dass adjazente Knoten verschiedene Farben haben.

Hier noch historischen Exkurs einfügen

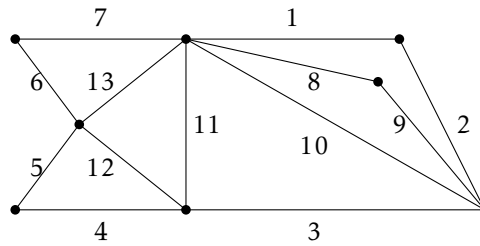


Abbildung 6.18: Ein Graph mit einem Eulerkreis

6.6 Eulerkreise und Eulerpfade

Im letzten Abschnitt zum Thema Graphen beschäftigen wir uns mit der Fragestellung, welche heutzutage als der Ursprung der Graphentheorie gilt. Die Problemstellung stammt aus dem frühen 18. Jahrhundert und wurde damals durch Brücken in der Stadt Königsberg motiviert. In der Stadt gibt es mehrere Teile die durch Flüsse getrennt sind, über die Flüsse gibt es Brücken. Die Frage ist nun, ob es eine Möglichkeit gibt, einen Spaziergang so zu planen, sodass jede Brücke genau einmal verwendet wird. Das **Königsberger Brückenproblem** kann durch einen Graphen modelliert werden. Hierbei werden die durch die Flüsse abgetrennte Bereiche als Knoten modelliert und die Brücken durch Kanten. Eine allgemeine Regel um zu erkennen, ob ein solcher Rundgang möglich ist, hat Leonhard Euler 1736 aufgestellt. Betrachtet man den Stadtplan von Königsberg aus dem 18. Jahrhundert sieht man, dass für die Beantwortung des ursprünglichen Problems mehrere Kanten zwischen zwei Knoten erlaubt sein müssten. Im Folgenden betrachten wir eine Verallgemeinerung des Problems auf Graphen, wobei wir weiter von schlichten Graphen ausgehen. Werden die entsprechenden Begriffe jedoch passend auf Graphen mit Mehrfachkanten verallgemeinert gelten die gezeigten Aussagen auch für solche Graphen.

Definition 6.20. In einem ungerichteten Graph $G = (V, E)$ ist ein **Eulerpfad** ein Kantenzug, der jede Kante einmal durchquert. Ein **Eulerkreis** ist ein geschlossener Eulerpfad.

Entgegen ihrem Namen sind Eulerpfade und Eulerkreise also nicht notwendigerweise Pfade oder Kreise, sondern lediglich Kantenzüge. Ein Beispiel zu einem Eulerkreis findet sich in Abbildung 6.18.

Der folgende Satz gibt eine einfach zu überprüfende Charakterisierung von Graphen die einen Eulerkreis haben.

Satz 6.17. Ein zusammenhängender ungerichteter Graph hat einen Eulerkreis genau dann, wenn alle Knoten geraden Grad haben.

Beweis. Zunächst einmal zeigen wir, dass wenn ein Eulerkreis existiert alle Knoten geraden Grad haben müssen. Seien v_1, \dots, v_m die Knoten, die auf dem Eulerkreis besucht werden, in dieser Reihenfolge. Zur Erinnerung: Ein Knoten kann mehrfach in

dieser Reihenfolge auftreten. Dann wird v_i durch die Kante $\{v_{i-1}, v_i\}$ „betreten“ und durch $\{v_i, v_{i+1}\}$ wieder „verlassen“. Jedes Mal, wenn ein Knoten besucht wird, wird also zwei zum Grad beigetragen und insgesamt ergibt sich ein gerader Grad für jeden Knoten.

Für die andere Richtung geben wir einen konstruktiven Beweis an. Also wir zeigen nicht nur, dass es einen Eulerkreis gibt, wenn alle Knoten geraden Grad haben, der Beweis gibt auch eine Vorschrift an, wie dieser gefunden werden kann. Hat der Graph nur einen Knoten, sind wir fertig. Ansonsten hat G mindestens 3 Knoten, da jeder Knoten geraden Grad hat.

Da jeder Knoten geraden Grad hat, gibt es keine Blätter⁶ in G . Aus der Kontraposition der Aussage, dass wenn G ein Baum ist, ein Blatt existiert, folgt dann also, dass G kein Baum ist und daher ein einfacher Kreis C in G existiert. Dieser kann mit Hilfe von Breitensuche oder Tiefensuche gefunden werden. Werden nun alle Kanten von C aus G entfernt ergibt sich ein neuer Graph, in dem alle Knoten geraden Grad haben. Dieser Graph ist unter Umständen jedoch nicht mehr zusammenhängend. Jede Zusammenhangskomponente hat jedoch mindestens einen Knoten der auf C liegt. Wir wählen einen diesen Knoten als Repräsentant der Komponente. In den einzelnen Zusammenhangskomponenten kann nun rekursiv wieder ein Eulerkreis gefunden werden, wobei die Rekursion stoppt, wenn die Zusammenhangskomponente nur aus einem einzelnen Knoten besteht. Diese Eulerkreise werden dann am jeweiligen Repräsentanten in C eingefügt und es ergibt sich ein Eulerkreis des gesamten Graphen. \square

Eine ähnliche Charakterisierung wie in Satz 6.17 gibt es auch für Eulerpfade. Hier wird gefordert, dass entweder keine oder genau zwei Knoten einen ungeraden Grad haben, und die restlichen Knoten geraden Grad. Der Beweis ist analog zum Beweis von Satz 6.17.

Selbsttest: Führen Sie den Algorithmus aus dem Beweis von Satz 6.17 auf dem Graph in Abbildung 6.18 aus.



Selbsttest: Führen Sie den Beweis für die Charakterisierung von Eulerpfaden.



Wird nun gefordert, dass jeder Knoten genau einmal auf einem Kreis bzw. Pfad besucht wird, ist keine Charakterisierung bekannt, welche leicht geprüft werden kann. Das Problem nennt sich dann **Hamiltonkreis** bzw. **Hamiltonpfad** und es ist nicht bekannt, ob es einen effizienten Algorithmus gibt, welcher entscheidet, ob ein gegebener Graph einen solchen Kreis bzw. Pfad als Untergraph hat.

⁶Also Knoten mit Grad 1

Entwurf

Anhang

Entwurf

Todo list

Klost: muss noch geschrieben werden	i
Klost: Der Teil zum Resolutionskalkül ist noch nicht im Skript und wird voraussichtlich auch nicht in der Vorlesung behandelt werden.	10
Klost: In diesem Abschnitt werden eventuell nach und nach noch weitere Beispiele ergänzt.	37
Max: Vereinheitlichung der Namensgebung?	96
Hier noch das Münzen und stehenbleiben Beispiel aus der VL einfügen	121■
Hier noch historischen Exkurs einfügen	152■