Vorlesung

Logik und Diskrete Mathematik

(Mathematik für Informatiker I)

Wintersemester 2012/13

FU Berlin

Institut für Informatik

Klaus Kriegel

Literatur zur Vorlesung:

C. Meinel, M. Mundhenk, Mathematische Grundlagen der Informatik, B.G.Teubner 2000

G. Haggard, J. Schlipf, S. Whitesides, Discrete Mathematics for Computer Science, Brooks Cole

U. Schöning, Logik für Informatiker, Spektrum, 1987

M. Aigner, Diskrete Mathematik, Vieweg & Sohn 1999

D. Hachenberger, Mathematik für Informatiker, Pearson Studium

T. Ihringer, Diskrete Mathematik, Teubner 1994

K. H. Rosen, Discrete Mathematics and its Applications, McGraw-Hill 1991

R. L. Graham, D. E. Knuth, O. Patashnik, Concrete Mathematics, Addison-Wesley 1994

C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill

Contents

1 Aussagenlogik **4**
- 1.1 Aussagen . . . 4
- 1.2 Prädikate und Quantoren . . . 8
- 1.3 Beweistechniken . . . 11
- 1.4 Boolesche Formeln und Boolesche Funktionen . . . 13
- 1.5 Konjunktive und disjunktive Normalformen . . . 18
- 1.6 Vollständige Signaturen . . . 20

2 Grundbegriffe der Mengenlehre **22**
- 2.1 Mengen und Operationen auf Mengen . . . 22
- 2.2 Das Kartesische Produkt und Relationen . . . 26
- 2.3 Äquivalenzrelationen . . . 28
- 2.4 Ordnungsrelationen . . . 30
- 2.5 Funktionen . . . 33
- 2.6 Natürliche Zahlen und vollständige Induktion . . . 36
- 2.7 Endliche Mengen und Kardinalzahlen . . . 38

3 Kombinatorik **42**
- 3.1 Elementare Zählprinzipien . . . 42
- 3.2 Die fundamentalen Zählkoeffizienten . . . 43
- 3.3 Zwölf Arten des Abzählens . . . 46
- 3.4 Rekursionen . . . 47
- 3.5 Lösung von Rekursionen . . . 50
- 3.6 Das Schubfachprinzip . . . 52

4 Diskrete Wahrscheinlichkeitsverteilungen **54**
- 4.1 Wahrscheinlichkeitsräume, Ereignisse und Unabhängigkeit . . . 54
- 4.2 Zufallsvariable und Erwartungswert . . . 59

5 Grundlagen der Graphentheorie **62**
- 5.1 Graphen und ihre Darstellungen . . . 62
- 5.2 Bäume . . . 67
- 5.3 Matchings . . . 71

1 Aussagenlogik

1.1 Aussagen

Die Grundlagen der Aussagenlogik gehen bereits auf die alten Griechen zurück. So beschrieb Aristoteles eine Aussage als einen Satz, von dem es sinnvoll ist zu sagen, dass er wahr oder falsch sei. Diesen Gedanken findet man auch in der heute verwendeten Definition wieder.

Definition: Eine *Aussage* ist ein (formal-)sprachliches Gebilde, das entweder wahr oder falsch ist.

Der Zusatz formalsprachlich weist darauf hin, dass man auch mathematische Symbole und andere Zeichen einer formalen Sprache verwenden kann. Die klassische Aussagenlogik beruht auf zwei Grundprinzipien, dem bereits genannten *Zweiwertigkeitsprinzip*, welches fordert, dass jede Aussage einen eindeutig bestimmten Wahrheitswert hat, der nur *wahr* oder *falsch* sein kann, und dem *Extensionalitätsprinzip*, nach dem der Wahrheitswert einer zusammengesetzten Aussage nur von den Wahrheitswerten ihrer Bestandteile abhängt.

Wir werden im Folgenden (wie in der Informatik üblich) eine 1 für den Wahrheitswert *wahr* und eine 0 für *falsch* verwenden. Das Zusammensetzen von Aussagen erfolgt durch die Verwendung von Verknüpfungswörtern wie *und, oder, nicht, wenn ... dann*, welche auf formal-sprachlicher Ebene durch sogenannte *logische Junktoren* - das sind spezielle Verknüpfungssymbole - dargestellt werden.

Beispiele:

1. Der Satz *"7 ist eine Primzahl."* und der Satz *"7 ist eine ungerade Zahl."* sind wahre Aussagen. Dagegen ist der Satz *"7 ist eine gerade Zahl."* eine falsche Aussage. Genauer gesehen ist der letzte Satz die Negation des zweiten Satzes, denn *nicht ungerade* zu sein, ist (zumindest für ganze Zahlen) das gleiche, wie *gerade* zu sein.

2. Der Satz *"7 ist eine Primzahl und 7 ist ungerade."* sowie der Satz *"7 ist eine Primzahl oder 7 ist gerade."* sind wahre Aussagen. Achtung: Auch der Satz *"7 ist eine Primzahl oder 7 ist ungerade."* ist eine wahre Aussage, denn das logische *oder* ist kein ausschließendes *entweder oder*. Dagegen ist der Satz *"7 ist eine Primzahl und 7 ist gerade."* eine falsche Aussage, denn die zweite Aussage ist falsch.

3. Der Satz *"$\sqrt{2}$ ist eine rationale Zahl."* ist – wie man aus der Schulmathematik weiß – eine falsche Aussage, aber es bedarf schon einiger Überlegungen, um das zu zeigen.

4. Der Satz *"Jede gerade Zahl größer als 2 ist die Summe zweier Primzahlen"* ist eine Aussage, denn entweder gibt es eine gerade Zahl, die sich nicht als Summe zweier Primzahlen darstellen lässt - dann ist die Aussage falsch, oder es gibt keine solche Zahl - dann ist die Aussage wahr. Man nimmt an, dass diese

Aussage wahr ist (Goldbachsche Vermutung), konnte das aber bisher noch nicht beweisen.

5. Der Satz *"Dieser Satz ist falsch."* ist als Russels Paradoxon bekannt. Durch die spezielle Art des Bezugs auf sich selbst kann er weder wahr noch falsch sein und ist deshalb **keine** Aussage.

6. Ein typischer Vertreter für eine ganze Klasse von sprachlichen Gebilden, die keine Aussagen sind, ist der Satz *"Die natürliche Zahl n ist eine Primzahl."*. Setzen wir für n den Wert 7 ein, so entsteht offensichtlich eine wahre Aussage, dagegen für $n = 8$ eine falsche Aussage. Sprachliche Gebilde dieses Typs nennt man auch Aussageformen oder Prädikate - wir werden sie später genauer besprechen.

Nach dem Extensionalitätsprinzip ergibt sich der Wahrheitswert einer zusammengesetzten Aussage ausschließlich aus den Wahrheitswerten der Ausgangskomponenten. Deshalb werden wir uns zuerst damit beschäftigen, welche Operationen zum Zusammensetzen neuer Aussagen verwendet werden sollen und wie diese Operationen auf Wahrheitswerten wirken. Dazu werden Aussagevariable eingeführt und die Wahrheitwerte von zusammengesetzten Aussagen durch *Wahrheitsswerttabellen* (kurz Wahrheitstafeln) beschrieben.

Die Negation einer Aussage x wird mit $\neg x$ bezeichnet. Diese Operation kehrt den Wahrheitswert von x um, d.h. man kann sie mit $\neg 0 = 1$ und $\neg 1 = 0$ als Funktion auf den Wahrheitswerten beschreiben. Zur Verknüpfung von zwei Aussagen x und y stehen die folgenden Konstrukte zur Verfügung:

- die *Konjunktion* $x \wedge y$, gesprochen "x *und* y";
- die *Disjunktion* $x \vee y$, gesprochen "x *oder* y";
- die *Implikation* $x \to y$, gesprochen "*aus* x $folgt$ y" oder "*wenn* x, *dann* y"
- die *Äquivalenz* $x \leftrightarrow y$, gesprochen "x *genau dann*, *wenn* y",
- die *Antivalenz* $x \oplus y$, gesprochen "*entweder* x *oder* y".

Die dazu korrespondierenden Funktionen auf Wahrheitswerten werden als Operationen (unter Verwendung der gleichen Symbole) in der folgenden Tabelle beschrieben:

x	y	$x \wedge y$	$x \vee y$	$x \to y$	$x \leftrightarrow y$	$x \oplus y$
0	0	0	0	1	1	0
0	1	0	1	1	0	1
1	0	0	1	0	0	1
1	1	1	1	1	1	0

Aus der Tabelle kann man ablesen, dass die Konjunktion $x \wedge y$ dann und nur dann wahr ist, wenn beide Aussagen x und y wahr sind. Die Disjunktion $x \vee y$ ist dann und

nur dann wahr, wenn mindestens eine der Aussagen x und y wahr ist. Die Implikation ist dann und nur dann wahr, wenn x falsch oder y wahr ist. Versuchen Sie selbst, die Äquivalenz und die Antivalenz verbal zu beschreiben!

Ausdrücke, die durch (wiederholtes) Anwenden der Verknüpfungsoperationen aus Variablen gewonnen werden, nennt man *Formeln* (oder *Terme*) der Aussagenlogik. Um eine Formel eindeutig erkennen zu können, müsste man jeweils nach Anwendung einer Verknüpfung die neue Formel durch ein Klammerpaar einschließen. Das führt zur folgenden Definition von Formeln der Aussagenlogik über eine Variablenmenge Var:

1. Alle Variablen aus der Menge Var sowie die Symbole 0 und 1 sind Formeln der Aussagenlogik.
2. Ist t eine Formel der Aussagenlogik, dann ist auch $(\neg t)$ eine Formel der Aussagenlogik.
3. Sind s und t Formeln der Aussagenlogik, dann sind auch die Ausdrücke $(s \wedge t)$, $(s \vee t)$, $(s \to t)$, $(s \leftrightarrow t)$ sowie $(s \oplus t)$ Formeln der Aussagenlogik.
4. Jede Formel der Aussagenlogik kann aus den Variablen und den Symbolen 0 und 1 durch eine endliche Folge von Anwendungen der Regeln 2) und 3) erzeugt werden.

Wir werden zur Abkürzung der Notation (insbesondere an der Tafel) auch häufig die Symbole $\Rightarrow$ und $\Leftrightarrow$ verwenden. Sie drücken auch Implikationen und Äquivalenzen aus, aber nicht auf der Ebene der Formeln, sondern auf der Ebene der Sprache über Aussagen, der sogenannten Metasprache. Der Ausdruck $A \Rightarrow B$ ist also ein Kürzel für den Satz "Wenn Aussage A wahr ist, dann ist auch Aussage B wahr".

Weil die Formeln durch die Klammersetzung sehr unübersichtlich werden können, vereinbart man einige Regeln zur Vereinfachung der Notation (ähnlich wie die bekannte Regel, dass Punktrechnung vor Strichrechnung geht):

- Außenklammern können weggelassen werden.
- In der Reihenfolge $\neg, \wedge, \vee, \to, \leftrightarrow$ trennen die hinteren Junktoren stärker als alle vorangehenden, d.h. die *Bindungsstärke* nimmt in dieser Reihenfolge ab. Alle Klammerungen, die mit dieser Hierarchie der Bindungsstärke in Übereinstimmung stehen, können auch weggelassen werden.

Man kann also für $((\neg x_1) \vee (x_2 \wedge x_3))$ auch $\neg x_1 \vee x_2 \wedge x_3$ schreiben. Dagegen würde das Weglassen der Klammern in der Formel $\neg(x \vee y)$ eine andere Formel erzeugen.

Legt man für alle in einer Formel auftretenden Variablen Wahrheitswerte fest, so induziert eine solche Belegung auch einen Wahrheitswert für die Formel. Man nennt diesen induktiven Prozess auch *Auswertung* der Formel. Die Ergebnisse der Auswertungen einer Formel unter allen möglichen Belegungen werden in einer Wahrheitstafel zusammengefasst.

Definition: Zwei Formeln s und t sind *logisch äquivalent*, wenn jede beliebige Belegung der Variablen für beide Formeln den gleichen Wahrheitswert induziert. Wir schreiben dafür $s \equiv t$.
Wie das folgende Beispiel zeigt, kann die Äquivalenz von zwei Formeln prinzipiell durch Wahrheitstafeln überprüft werden. Zum Nachweis der Äquivalenz der Formeln $s = \neg(x_1 \vee ((x_1 \vee x_2) \wedge x_2)$ und $t = \neg x_1 \wedge \neg x_2$ stellt man folgende Tabelle auf:

x_1	x_2	$x_1 \vee x_2$	$(x_1 \vee x_2) \wedge x_2$	$x_1 \vee ((x_1 \vee x_2) \wedge x_2)$	s
0	0	0	0	0	1
0	1	1	1	1	0
1	0	1	0	1	0
1	1	1	1	1	0
x_1	x_2	$\neg x_1$	$\neg x_2$		t
0	0	1	1		1
0	1	1	0		0
1	0	0	1		0
1	1	0	0		0

Am identischen Wahrheitswerteverlauf für s und t kann man ablesen, dass die Formeln äquivalent sind.

Satz: Für beliebige Formeln s, t, r gelten die folgenden Äquivalenzen:

Assoziativität:	$(s \wedge t) \wedge r$	$\equiv$	$s \wedge (t \wedge r)$
	$(s \vee t) \vee r$	$\equiv$	$s \vee (t \vee r)$
Kommutativität:	$s \wedge t$	$\equiv$	$t \wedge s$
	$s \vee t$	$\equiv$	$t \vee s$
Distributivität:	$s \wedge (t \vee r)$	$\equiv$	$(s \wedge t) \vee (s \wedge r)$
	$s \vee (t \wedge r)$	$\equiv$	$(s \vee t) \wedge (s \vee r)$
Idempotenz:	$s \wedge s$	$\equiv$	s
	$s \vee s$	$\equiv$	s
Dominanz:	$s \wedge 0$	$\equiv$	0
	$s \vee 1$	$\equiv$	1
Neutralität:	$s \wedge 1$	$\equiv$	s
	$s \vee 0$	$\equiv$	s
Absorbtion:	$s \wedge (s \vee t)$	$\equiv$	s
	$s \vee (s \wedge t)$	$\equiv$	s
deMorgansche Regel:	$\neg(s \wedge t)$	$\equiv$	$\neg s \vee \neg t$
	$\neg(s \vee t)$	$\equiv$	$\neg s \wedge \neg t$
Komplementierung:	$s \wedge \neg s$	$\equiv$	0
	$s \vee \neg s$	$\equiv$	1
(*doppelte Negation)*	$\neg\neg s$	$\equiv$	s

Diese Äquivalenzen können leicht mit Wahrheitstafeln bewiesen werden. Der Wahrheitstafelmethode sind jedoch enge Grenzen gesetzt, wenn die Anzahl n der ver-

wendeten Variablen groß wird, denn die entsprechende Wahrheitstafel hat dann 2^n Zeilen.

Beispiel: Der Beweis der folgenden Äquivalenz mit Wahrheitstafeln würde 16 Zeilen erfordern. Verwendet man dagegen die Absorbtion und die doppelte Negation zur Ersetzung von Subformeln, so erhält man einen einfachen und kurzen Beweis.

$$\begin{aligned} x_1 \vee ((x_2 \vee x_3) \wedge \neg(\neg x_1 \wedge (\neg x_1 \vee x_4))) &\equiv x_1 \vee ((x_2 \vee x_3) \wedge \neg\neg x_1) \\ &\equiv x_1 \vee ((x_2 \vee x_3) \wedge x_1) \\ &\equiv x_1 \end{aligned}$$

Die folgende Liste enthält weitere Äquivalenzen, welche zum Beweis der Äquivalenz von komplexen Formeln häufig angewendet werden:

$$\begin{array}{lrcl} (1) & s \to t & \equiv & \neg s \vee t \\ (2) & s \leftrightarrow t & \equiv & s \wedge t \vee \neg s \wedge \neg t \\ (3) & s \to t \wedge r & \equiv & (s \to t) \wedge (s \to r) \\ (4) & s \to t \vee r & \equiv & (s \to t) \vee (s \to r) \\ (5) & s \wedge t \to r & \equiv & (s \to r) \vee (t \to r) \\ (6) & s \vee t \to r & \equiv & (s \to r) \wedge (t \to r) \end{array}$$

Definition: Eine Formel s wird *erfüllbar* genannt, wenn es eine Belegung der Variablen von s gibt, die für s den Wert 1 induziert. Die Formel s wird *allgemeingültig, logisch gültig* oder eine *Tautologie* genannt, wenn sie für jede Belegung den Wert 1 annimmt. Eine Formel, die unerfüllbar ist, wird *Kontradiktion* genannt.

1.2 Prädikate und Quantoren

Definition: Ein *Prädikat* ist eine Aussageform, die eine (oder mehrere) Variable enthält, so dass bei Ersetzung der Variablen durch Elemente aus einem gegebenen Individuenbereich U eine Aussage mit eindeutig bestimmtem Wahrheitswert entsteht, z.B. $P(x)$: "$x = 0$" oder $Q(x)$: "$x + 0 = x$" oder $R(x, y)$: "$x \leq x + y$" für den Bereich der ganzen Zahlen.

Die Belegung der Variablen durch konkrete Objekte ermöglicht somit (durch Betrachtung eines Spezialfalls), ein Prädikat in eine Aussage umzuwandeln. So sind für die genannten Beispiele die Aussagen $P(0)$, $Q(5)$ und $R(3, 4)$ wahr, dagegen sind $P(3)$ und $R(3, -2)$ falsche Aussagen.

Die sogenannten *Quantoren* erlauben es, aus diesen Spezialfällen allgemeinere Aussagen abzuleiten: Durch das Hinzufügen der Wendungen "für alle ...", symbolisch ausgedrückt durch den *Allquantor* $\forall$, oder "es gibt ein ...", symbolisch ausgedrückt durch den *Existenzquantor* $\exists$, werden die Variablen in einem Prädikat gebunden. Sind alle Variablen eines Prädikats gebunden, entsteht eine Aussage, also ein Satz, der wahr oder falsch ist.

Die Ausssage "$\forall x \in U \quad P(x)$" ist wahr, wenn für jedes Element $a \in U$ die Aussage $P(a)$ wahr ist. Dagegen ist "$\exists x \in U \quad P(x)$" eine wahre Aussage, wenn (mindestens) ein Element $a \in U$ existiert, so dass die Aussage $P(a)$ wahr ist.

Beispiele:

- Die Aussagen "$\forall x \in \mathbb{N} \quad x + 0 = x$" und "$\exists x \in \mathbb{N} \quad x^2 = x$" sind wahr, aber die Aussagen "$\exists x \in \mathbb{N} \quad x + 1 = x$" und "$\forall x \in \mathbb{N} \quad x^2 = x$" sind falsch.

- Die Aussage "$\forall x \in \mathbb{N}\ \exists y \in \mathbb{N} \quad y \leq x$" ist wahr, denn für einen beliebigen Wert $x = a$ erfüllt der Wert $y = a$ die Ungleichung $y \leq x$. Dagegen ist die Aussage "$\forall x \in \mathbb{N}\ \exists y \in \mathbb{N} \quad y < x$" falsch, denn für $x = 0$ gibt es keine kleinere natürliche Zahl.

- Die falsche Aussage im letzten Punkt ist ein typisches Beispiel dafür, dass der Bereich, über dem die Aussage gemacht wird, von entscheidender Bedeutung sein kann: Wenn man den Bereich $\mathbb{N}$ der natürlichen Zahlen gegen die Bereiche $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ der ganzen, rationalen bzw. reellen Zahlen austauscht, entstehen offensichtlich wahre Aussagen wie "$\forall x \in \mathbb{Z}\ \exists y \in \mathbb{Z} \quad y < x$".

Allgemein ist die Frage, ob eine durch Quantoren gebildete Aussage wahr oder falsch ist, algorithmsch nicht entscheidbar. Die Goldbachsche Vermutung ist ein Beispiel einer Aussage, deren Wahrheitswert nicht bekannt ist. In vielen anderen Fällen kann man die Frage aber durch genauere Überlegungen beantworten. Wie kann man in solchen Fällen sich selbst und andere von der Richtigkeit seiner Überlegungen überzeugen? Der typische Beweis dafür, dass eine quantisierte Aussage wahr ist, erfolgt in drei Stufen. Zuerst wird die Aussage durch Anwendung von äquivalenten Umformungen aus der Aussagenlogik und aus dem nachfolgenden Satz in eine Standardform gebracht, bei der alle auftretenden Quantoren am Anfang stehen (man nennt dies eine *Pränexform*). Danach erfolgt die Belegung der Variablen in Form eines Spiels zwischen zwei Parteien: Einem *Beweiser* und seinem *Gegenspieler*, der nachzuweisen versucht, dass die Ausage falsch ist. Dabei darf der Gegenspieler bei jedem Allquantor die entsprechende Variable x durch ein beliebiges Objekt a aus dem Individuenbereich belegen. Sollte die Aussage doch falsch sein (also nicht für alle Objekte gelten), würde der Gegenspieler gerade ein solche Objekt wählen. Ist die Aussage wahr, dann ist es (für den Beweiser) egal, welches Objekt a der Gegenspieler gewählt hat. Der Beweiser ist bei allen Existenzquantoren am Zuge und muss ein passendes Objekt (in Abhängigkeit von den vorher vom Gegenspieler gwählten Objekten) finden, für welches die nachfolgende Ausssage wahr ist. Nachdem alle Variablen belegt sind, haben wir eine (variablenfreie) Aussage. Im letzten Schritt muss diese Aussage verifiziert (als wahr bewiesen) werden.

Beginnen wir mit den Umformungsregeln.

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

$$
\begin{array}{lrcl}
(1) & \neg\forall x\ P(x) & \equiv & \exists x \neg P(x) \\
(2) & \neg\exists x\ P(x) & \equiv & \forall x \neg P(x) \\
(3) & \forall x\ P(x) \wedge \forall x\ Q(x) & \equiv & \forall x\ (P(x) \wedge Q(x)) \\
(4) & \exists x\ P(x) \vee \exists x\ Q(x) & \equiv & \exists x\ (P(x) \vee Q(x)) \\
(5) & \forall x \forall y\ R(x, y) & \equiv & \forall y \forall x\ R(x, y) \\
(6) & \exists x \exists y\ R(x, y) & \equiv & \exists y \exists x\ R(x, y)
\end{array}
$$

Achtung: Die folgenden Formelpaare sind im allgemeinen nicht äquivalent:

$$\begin{array}{lll} \forall x\ P(x) \vee \forall x\ Q(x) & \textit{und} & \forall x\ (P(x) \vee Q(x)) \\ \exists x\ P(x) \wedge \exists x\ Q(x) & \textit{und} & \exists x\ (P(x) \wedge Q(x)) \\ \forall x\quad (\exists y\ R(x,y)) & \textit{und} & \exists y\quad (\forall x\ R(x,y)) \end{array}$$

Konkrete Gegenbeispiele für das erste und zweite Paar erhält man für den Bereich der ganzen Zahlen, wenn $P(x)$ (bzw. $Q(x)$) aussagt, dass x eine gerade (bzw. ungerade) Zahl ist. Für das dritte Paar kann man das Prädikat $R(x,y)$:"$x \leq y$" über den reellen Zahlen verwenden.

Wir wollen die drei Stufen eines solchen Beweises an einem einfachen Beispiel demonstrieren und die folgende Aussage für den Bereich der rationalen Zahlen beweisen.

Satz: Für beliebige $x, y \in \mathbb{Q}$ mit $x < y$ gibt es eine von x und y verschiedene Zahl $z \in \mathbb{Q}$, die zwischen x und y liegt.

Die Beweisidee liegt klar auf der Hand: man setzt für z den Mittelwert aus x und y und kann dann die Behauptung nachrechnen. Wir stellen die Aussage als Formel mit Quantoren dar. Der Bereich $\mathbb{Q}$ ist durch Verabredung festgelegt und wird nicht explizit genannt:

$$\forall x\ \forall y\ [(x < y) \rightarrow \exists z\ (x < z \wedge z < y)]$$

In diesem Fall könnte man die erste Stufe überspringen und gleich mit dem Spiel der Festlegung der Werte beginnen, bei dem der Gegner x und y vorgibt und wir im Fall $x < y$ (als Beweiser) $z = \frac{x+y}{2}$ setzen. Um das dreistufige Verfahren formal korrekt umzusetzen, würden wir aber zuerst die Implikation hinter den beiden Allquatoren äquivalent umformen:

$$\begin{array}{ll} (x < y) \rightarrow \exists z\ (x < z \wedge z < y) & \Leftrightarrow \neg(x < y) \vee \exists z\ (x < z \wedge z < y) \\ & \Leftrightarrow \exists z\,(x \geq y) \vee \exists z\ (x < z \wedge z < y) \\ & \Leftrightarrow \exists z\,(x \geq y \vee\ \ (x < z \wedge z < y)) \end{array}$$

Im ersten Schritt wurde die Regel $p \rightarrow q \equiv \neg p \vee q$ angewendet. Im zweiten Schritt wurde die Negation von $x < y$ in $x \geq y$ umgewandelt und der Existenzquantor $\exists z$ davorgesetzt - das kann man machen, weil $x \geq y$ in keiner Weise von z abhängt. Im letzten Schritt wurde die vierte Regel aus obigem Satz verwendet. Jetzt liegt die Aussage in Pränexform vor:

$$\forall x\ \forall y\ \exists z\ [x \geq y \vee (x < z \wedge z < y)].$$

Das Spiel beginnt mit der Vorgabe von zwei Werten $x = a$ und $y = b$ durch den Gegenspieler. Der Beweiser setzt $z = (a + b)/2$.
Nun erfolgt die Verifikation der Aussage $a \geq b \vee (a < (a+b)/2 \wedge (a+b)/2 < y)$ durch Betrachtung von zwei Fällen: Entweder es gilt $a \geq b$, dann wird die Aussage durch den linken Term der Disjunktion wahr oder es gilt $a < b$. In diesem Fall müssen beide Ungleichungen auf der rechten Seite der Disjunktion erfüllt sein, aber beides folgt über den Zwischenschritt $a/2 < b/2$ jeweils kombiniert mit den Ungleichungen $a/2 \leq a/2$ bzw. $b/2 \leq b/2$.

Bleibt die Frage, warum man diese Argumentation nicht für den Bereich $\mathbb{N}$ der natürlichen Zahlen wiederholen kann. Die Antwort ist wieder nicht schwer: Der Ausdruck $\frac{a+b}{2}$ führt in vielen Fällen aus dem Bereich $\mathbb{N}$ heraus. Das ist aber noch kein Beweis dafür, dass die Aussagge über diesem Bereich falsch ist. Um das zu zeigen, beweist man, dass die negierte Aussage wahr ist. Wir bilden die Negation mit Hilfe der Regeln (1) und (2) aus obigem Satz und der deMorganschen Regel:

$$\exists x\ \exists y\ \forall z\ [(x < y) \wedge (x \geq z \vee z \geq y)].$$

Da die Aussage bereits in Pränexform vorliegt, kann das Spiel der Variablenbelegung beginnen. Wir setzen als Beweiser $x = 0$ und $y = 1$. Der Gegenspieler belegt z mit einem Wert c. Wir müssen jetzt die Aussage $0 < 1 \wedge (0 \geq c \vee c \geq 1)$ verifizieren. Offensichtlich ist $0 < 1$ wahr und wir müssen nur noch zeigen, dass mindestens eine der Ungleichungen aus $(0 \geq c \vee c \geq 1)$ wahr ist. Wenn aber die erste Ungleichung nicht erfüllt ist, muss c eine natürliche Zahl größer als 0, also mindestens 1 sein und damit ist die zweite Ungleichung erfüllt.

1.3 Beweistechniken

In diesem Abschnitt geht es darum, einige grundlegende Beweisstrategien kennenzulernen. Da das prinzipielle Verständnis im Mittelpunkt stehen soll, sind die in den Beispielen bewiesenen Aussagen sehr einfach gewählt. Grundsätzlich muss man sich bei Beweisen auch immer danach richten, für welchen Leser der Beweis aufgeschrieben wird. Bei einem Beweis für Experten kann man mehr Vorwissen voraussetzen und deshalb auch mehrere Gedanken zu einem Schritt zusammenfassen. Dagegen sollte bei Beweisen für Nicht-Experten möglichst jeder Schritt elementar und einfach nachvollziehbar sein.
Gerade bei den scheinbaren Selbstverständlichkeiten wird ein weiteres Problem deutlich: Bevor man an einen Beweis geht, muss man sich klarmachen, was man schon als Basiswissen voraussetzen kann, und was man noch beweisen muss. Oft ist als erster hilfreicher Schritt eine geeignete Formalisierung der Aussage notwendig.

Viele mathematische Sätze haben die Form einer Implikation, sie sagen, dass aus einer bestimmten Voraussetzung in Form einer Aussage p eine Behauptung in Form einer Aussage q folgt. Wir wollen uns zuerst mit den verschiedenen Techniken zum Beweis von solchen Implikationen beschäftigen. Basis für die Gültigkeit solcher Beweise sind einige einfache Äquivalenzen und Implikationen, die man leicht mit der Wahrheitstafelmethode nachweisen kann.

Direkte Beweise

Der *direkte Beweis* beruht darauf, die Implikation $p \to q$ in mehrere elementare Teilschritte zu zerlegen, wobei man die folgende Tautologie nutzt:

$$((p \to r) \wedge (r \to q)) \to (p \to q).$$

Natürlich kann man die zwei Teilschritte auf der linken Seite weiter unterteilen, bis man bei einer Kette elementarer Implikationen angekommen ist. Wie das folgende

Beispiel zeigt, bewegt man sich bei der Begründung der Elementarschritte in einem System, das sich auf einigen Axiomen (Grundannahmen) aufbaut und in dem man auf bereits bewiesene Tatsachen zurückgreifen kann.

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

Beweis: Die Idee ist offensichtlich – ist n durch 6 teilbar, so kann man den Faktor 6 und damit auch den Faktor 3 von n abspalten. Folglich kann man den Faktor 3 mindestens zweimal von n^2 abspalten. Wenn wir diese Idee etwas formaler umsetzen wollen, müssen wir mit der Definition von Teilbarkeit beginnen:

$$n \in \mathbb{N} \text{ ist durch } k \in \mathbb{N} \text{ teilbar, falls ein } l \in \mathbb{N} \text{ existiert, so dass } n = k \cdot l.$$

Die Tatsache, dass n durch k teilbar ist (also k ein Teiler von n ist) wird symbolisch durch $k \,|\, n$ ausgedrückt. Damit kann man die Voraussetzung des Satzes durch eine einfache Formel ausdrücken und die folgende Beweiskette bilden:

$6 \,\|\, n$		Hypothese
$\exists l \in \mathbb{N}$	$n = 6 \cdot l$	Teilbarkeitsdefinition
$\exists l \in \mathbb{N}$	$n = (3 \cdot 2) \cdot l$	$6 = 3 \cdot 2$
$\exists l \in \mathbb{N}$	$n^2 = ((3 \cdot 2) \cdot l)((3 \cdot 2) \cdot l)$	Quadrieren
$\exists l \in \mathbb{N}$	$n^2 = (3 \cdot 3)((2 \cdot 2) \cdot (l \cdot l))$	Muliplikation ist assoziativ und kommutativ
$\exists l \in \mathbb{N}$	$n^2 = 9 \cdot (4 \cdot l^2)$	$3 \cdot 3 = 9$ und $2 \cdot 2 = 4$
$\exists l' \in \mathbb{N}$	$n^2 = 9 \cdot l'$	$l' = 4l^2$
$9 \,\|\, n^2$		Teilbarkeitsdefinition

Genau betrachtet haben wir beim Schritt von der vierten zur fünften Zeile sogar mehrere Elementarschritte zu einem Schritt zusammengefasst.

Indirekte Beweise

Manchmal ist es schwierig, den Beweis direkt zu führen. Als Alternativen bieten sich indirekte Beweise durch Kontraposition oder in der Form von Widerspruchs-Beweisen an. Beim *Beweis durch Kontraposition* wird anstelle von $p \to q$ die logisch äquivalente Aussage $\neg q \to \neg p$ bewiesen. Beim Widerspruchs-Beweis wird an Stelle von $p \to q$ die logisch äquivalente Aussage $(p \wedge \neg q) \to 0$ bewiesen. Wir demonstrieren beide Beweisverfahren an einfachen Beispielen.

Satz: Für jede natürliche Zahl n gilt: Ist n^2 ungerade, so ist auch n ungerade.

Beweis durch Kontraposition: Da die Negation von *"ungerade sein"* die Eigenschaft *"gerade sein"* ist, lautet die Kontraposition *"Ist n gerade, so ist auch n^2 gerade"*. und dafür gibt es einen einfachen direkten Beweis:
Ist n gerade, so gibt es eine ganze Zahl k mit $n = 2k$. Folglich ist $n^2 = (2k)^2 = 2 \cdot (2k^2)$ und somit ist n^2 gerade.

Satz: Für jede natürliche Zahl n gilt: Ist $\sqrt{n}$ keine ganze Zahl, dann ist $\sqrt{n}$ auch nicht rational.

Beweis durch Widerspruch: Man geht von der Annahme aus, dass $\sqrt{n}$ keine natürliche Zahl, aber eine rationale Zahl ist, und muss daraus einen Widerspruch

ableiten. Sei $n = p_1^{k_1} \cdot \ldots \cdot p_l^{k_l}$ die eindeutige Primzahlzerlegung von n, wobei $p_1, \ldots, p_l$ die paarweise verschiedenen Primfaktoren von n sind und $k_1, \ldots, k_l$ die Vielfachheiten dieser Primfaktoren. Wären alle Werte $k_1, \ldots, k_l$ gerade Zahlen, dann wäre $\sqrt{n}$ eine ganze Zahl, nämlich $p_1^{k_1/2} \cdot \ldots \cdot p_l^{k_l/2}$. Nach unserer Annahme muss also mindestens einer dieser Werte ungerade sein, o.B.d.A. (d.h. ohne Beschränkung der Allgemeinheit) sei das k_1.
Darüber hinaus soll $\sqrt{n}$ rational, also als Quotient aus zwei natürlichen Zahlen m und m' darstellbar sein. Damit ist $n = \left(\frac{m}{m'}\right)^2$ und $n \cdot m'^2 = m^2$. Kombiniert man diese Gleichung mit den Primzahlzerlegungen von $m = q_1^{i_1} \cdot \ldots \cdot q_j^{i_j}$ und $m' = r_1^{i'_1} \cdot \ldots \cdot r_{j'}^{i'_{j'}}$, ergibt sich:

$$p_1^{k_1} \cdot \ldots \cdot p_l^{k_l} \cdot r_1^{2i'_1} \cdot \ldots \cdot r_{j'}^{2i'_{j'}} = q_1^{2i_1} \cdot \ldots \cdot q_j^{2i_j}.$$

Folglich tritt der Primfaktor p_1 auf der linken Seite in ungerader Vielfachheit auf und auf der rechten Seite in gerader Vielfachheit (unabhängig davon, ob p_1 überhaupt in m oder m' vorkommt). Das ist aber ein Widerspruch zur eindeutigen Primzahlzerlegung von natürlichen Zahlen.

Beweise durch Fallunterscheidung

Häufig ist es notwendig, verschiedene Fälle zu analysieren. Das dabei verwendete logische Prinzip ist Äquivalenz der Aussagen $p \to q$ und $(p \wedge r \to q) \wedge (p \wedge \neg r \to q)$, wir unterscheiden also die Fälle r und $\neg r$.

Beispiel: Wir beweisen durch Fallunterscheidung, daß für jede Primzahl $p \geq 5$ die Zahl $p^2 - 1$ durch 24 teilbar ist.
Zuerst formen wir $p^2 - 1$ in $(p+1)(p-1)$ um und beobachten, dass von drei aufeinanderfolgenden ganzen Zahlen genau eine durch 3 teilbar ist. Da $p > 3$ und Primzahl ist, muss $p-1$ oder $p+1$ und damit auch p^2-1 durch 3 teilbar sein. Bleibt zu zeigen, dass $p^2 - 1$ durch 8 teilbar ist. Da p ungerade ist, sind sowohl $p - 1$ als auch $p + 1$ gerade und damit ist $p^2 - 1$ durch 4 teilbar. Den noch fehlenden Faktor 2 kann man durch Fallunterscheidung nachweisen:
1. Fall: Ist $p - 1$ durch 4 teilbar, so ist $p - 1 = 4k$ und $p + 1 = 4k + 2 = 2(2k + 1)$ und damit $p^2 - 1 = 8k(2k + 1)$ für eine natürliche Zahl k.
2. Fall: Ist $p - 1$ nicht durch 4 teilbar, so hat es die Form $4m + 2 = 2(2m + 1)$ für eine natürliche Zahl m und folglich ist $p + 1 = 4m + 4 = 4(m + 1)$. Damit erhalten wir $p^2 - 1 = 8(2m + 1)(m + 1)$.

1.4 Boolesche Formeln und Boolesche Funktionen

Der englische Mathematiker George Boole (1815-1864) hat sich um die mathematische Formalisierung der Aussagenlogik besonders verdient gemacht. Deshalb werden eine Reihe von Begriffen, die wir im Folgenden kennenlernen werden, mit seinem Namen verknüpft.

In diesem Abschnitt wird ein exakter und damit etwas formalerer Zugang zur Aussagenlogik besprochen. Insbesondere soll deutlich unterschieden werden zwischen der *Syntax*, d.h. der äußeren Gestalt einer Formel und ihrer *Semantik*, d.h. ihrer inneren Bedeutung. Wir beginnen mit der syntaktischen Seite, also mit der Beschreibung, welche Gestalt ein Ausdruck haben muss, um eine Boolesche Formel zu repräsentieren. Durch die zusätzliche Einführung eines Rangs wird es möglich, sogenannte Induktionsbeweise (sie werden im zweiten Kapitel genauer behandelt) über der Menge aller Booleschen Terme führen zu können.

Definition: *Primformeln* oder auch *atomare Formeln* sind die Booleschen Konstanten 0 und 1 sowie die Variablen $\{x, y, z, x_1, x_2, x_3, \ldots\}$. *Boolesche Formeln* (oder auch *Boolesche Terme*) und ihr *Rang* werden durch den folgenden induktiven Prozess definiert:

1. Alle Primformeln sind Formeln vom Rang 0, d.h. $rg(0) = rg(1) = rg(x_i) = 0$.
2. Für jede Formel t ist $\neg t$ eine Formel mit $rg(\neg t) = rg(t) + 1$.
3. Für alle Formeln s und t sind auch $(s \wedge t)$ und $(s \vee t)$ Formeln und man definiert $rg(s \wedge t) = rg(s \vee t) = max\{rg(s), rg(t)\} + 1$.
4. Alle Booleschen Formeln können aus Primformeln durch wiederholte Anwendungen der Regeln (2) und (3) erzeugt werden.

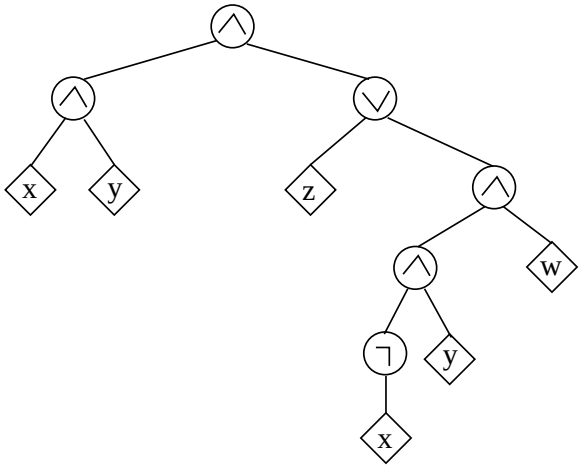
Anmerkungen:
1) Bei aussagenlogischen Formeln haben wir die Junktoren $\rightarrow, \leftrightarrow, \oplus$ kennengelernt. Spricht man über Boolesche Formeln, ist die Verwendung dieser Junktoren nicht üblich. Wir haben aber bereits gesehen, wie man sie logisch äquivalent durch $\vee, \wedge$ und $\neg$ ersetzen kann.

2) In der Literatur findet man häufig einen leicht modifizierten Ansatz, bei dem man auf die Booleschen Konstanten verzichtet und nur Variable als Primterme zulässt. Der Unterschied ist marginal, wir werden später noch genauer darauf eingehen.

3) Die Regeln zum Weglassen von Klammern wurden bereits besprochen. Zusätzlich wird vereinbart, dass bei gleichartigen Operationen jeder nichtgeklammerte Ausdruck als linksassoziativ geklammert angesehen wird.

4) Eine anschauliche Erklärung für den Aufbau von Termen und ihrem Rang bekommt man durch den sogenannten Syntaxbaum des Terms. Der Syntaxbaum einer Primformel besteht nur aus einem einzelnen Knoten, der mit der entsprechenden Variable markiert wird. Der Syntaxbaum eines Terms, der durch Regel 2 als $t = (\neg t')$ entsteht, hat einen mit $\neg$ markierten Wurzelknoten, unter dem der Syntaxbaum von t' steht. Für Terme, die nach Regel 3 entstehen, wird der Wurzelknoten mit dem entsprechenden Junktor markiert und zu den Syntaxbäumen der zwei verwendeten Teilterme verbunden. Knoten, die mit Primformeln markiert sind, nennt man Blätter des Syntaxbaums. Der längste Weg vom Wurzelknoten zu einem Blatt beschreibt den Rang der Formel.

Beispiel: Der Term $((x \wedge y) \wedge (z \vee (((\neg x) \wedge y) \wedge w)))$ hat den folgenden Syntaxbaum:



Induktive Beweise können nun über den Rang geführt werden: Ist P eine Eigenschaft von Zeichenfolgen derart, dass

1) alle Primformeln p diese Eigenschaft haben und
2) für alle Formeln s, t, die die Eigenschaft P haben, auch die Formeln $s \wedge t$, $s \vee t$ und $\neg s$ die Eigenschaft P haben,

dann haben alle Formeln die Eigenschaft P.

Auch die bereits besprochene Fortsetzung einer Wahrheitswertbelegung der Variablen auf zusammengesetzte Formeln, die sogenannte Auswertung der Formel wird induktiv definiert. Durch diese Auswertung erschließt sich die Bedeutung der Formel, also ihre semantische Seite. Im Folgenden bezeichen wir mit $\mathbb{B} = \{0, 1\}$ die Menge der Wahrheitswerte (Boolesche Werte).

Definition: Für eine Variablenmenge $Var_n = \{x_1, x_2, \ldots, x_n\}$ wird eine Abbildung $\omega : Var_n \longrightarrow \mathbb{B}$ eine *aussagenlogische Belegung* (auch kurz *Belegung, Realisierung*) genannt. Da man solche Abbildungen auch als n–Tupel von Nullen und Einsen darstellen kann, lässt sich die Menge Ω_n aller Belegungen der Variablenmenge $Var_n = \{x_1, x_2 \ldots, x_n\}$ auch durch $\mathbb{B}^n$ darstellen (die umkehrbare Abbildung ψ_n ordnet jeder Belegung ω das n-Tupel $(\omega(x_1), \omega(x_2), \ldots, \omega(x_n))$ zu).

Wir bezeichnen mit $\mathcal{F}_n$ die Menge aller aus Var_n erzeugbaren Formeln und definieren induktiv eine Funktion $\Phi_n : \Omega_n \times \mathcal{F}_n \longrightarrow \{0, 1\}$ durch

1. Für die Formeln vom Rang 0 und Belegungen $\omega \in \Omega_n$ ist $\Phi_n(\omega, 0) = 0$, $\Phi_n(\omega, 1) = 1$ und $\Phi_n(\omega, x_i) = \omega(x_i)$ für alle $x_i \in Var_n$.

2. Sei $\Phi_n(\omega, s)$ und $\Phi_n(\omega, t)$ für alle Formeln $s, t \in \mathcal{F}_n$ vom Rang $\leq k-1$ und alle Belegungen $\omega \in \Omega_n$ definiert, dann ist
   $\Phi_n(\omega, (s \wedge t)) = \Phi_n(\omega, s) \wedge \Phi_n(\omega, t)$,
   $\Phi_n(\omega, (s \vee t)) = \Phi_n(\omega, s) \vee \Phi_n(\omega, t)$ und
   $\Phi_n(\omega, \neg s) = \neg \Phi_n(\omega, s)$,
   womit $\Phi_n(\omega, u)$ für alle Formeln u von Grad $\leq k$ definiert ist.

Natürlich kann man diese Definition auch auf die Menge Ω der Belegungen aller Variablen $Var = \{x_1, x_2, \ldots\}$ und die Menge aller Formeln $\mathcal{F}$ über Var ausweiten. Die dabei definierte Funktion $\Phi : \Omega \times \mathcal{F} \longrightarrow \{0, 1\}$ stimmt auf den eingeschränkten Bereichen mit den Funktionen Φ_n überein.

Für jede Belegung $\omega_0 \in \Omega_n$ repräsentiert die Funktion $\Phi(\omega_0, \quad) : \mathcal{F}_n \longrightarrow \mathbb{B}$ die Fortsetzung dieser Belegung auf alle Formeln. Dagegen stellt für jede Formel $t \in \mathcal{F}_n$ die Funktion $\Phi(\quad, t) : \Omega_n \longrightarrow \mathbb{B}$ den Wahrheitswertverlauf der Formel dar. Benutzt man vorher die Bijektion ψ^{-1}, um n-Tupel in Belegungen umzuwandeln, so erhalten wir auf diese Weise die von t repäsentierte Boolesche Funktion f_t, d.h.

$$f_t = \Phi(\psi^{-1}(\quad), t) : \mathbb{B}^n \longrightarrow \mathbb{B}.$$

Diese Interpretation von Formeln unter allen möglichen Belegungen nennt man die *Semantik der Aussagenlogik*.

Definition: Funktionen von $\mathbb{B}^n$ nach $\mathbb{B}$ werden *n–stellige Boolesche Funktionen* genannt. Die Menge dieser Funktionen wird mit $\boldsymbol{B_n}$ bezeichnet. Die Menge aller Booleschen Funktionen ergibt sich durch $\boldsymbol{B} = \bigcup_{i \in \mathbb{N}} \boldsymbol{B_i}$, wobei man unter $\boldsymbol{B_0}$ die Menge $\mathbb{B}$ selbst versteht.

Obwohl eine Aussagenverknüpfung und die Operation auf den Booleschen Werten nicht gleichzusetzen sind, werden für beide die gleichen Symbole verwendet. Der Unterschied wird aus dem Kontext klar. Sind s und t Boolesche Formeln, dann bezeichnet $s \vee t$ die Disjunktion von s und t, also eine neue Formel (auf der linken Seite in der Definition der Interpretation Φ_n). Sind a und b Boolesche Werte, dann bezeichnet $a \vee b$ die Disjunktion der Wahrheitswerte, also keinen neuen Ausdruck, sondern einen konkreten Wert (auf der rechten Seite).

Definition: Zwei Formeln $s, t \in \mathcal{F}_n$ heißen *logisch äquivalent* (auch *semantisch äquivalent, wertverlaufsgleich* oder kurz *äquivalent*), falls $\Phi_n(\omega, s) = \Phi_n(\omega, t)$ für alle Belegungen $\omega \in \Omega_n$ gilt. Man schreibt dafür $s \equiv t$.

Offensichtlich gilt $s \equiv t$ genau dann, wenn s und t dieselbe Boolesche Funktion repräsentieren, d.h. wenn $f_s = f_t$. Jede Boolesche Funktion f, die durch einen Booleschen Term t dargestellt werden kann, hat unendlich viele Repräsentationen, nämlich $t \vee (x \wedge \neg x)$, $t \vee (x \wedge \neg x) \vee (x \wedge \neg x)$, usw.
Die Frage, ob jede Boolesche Funktion durch einen Booleschen Term repräsentiert werden kann, ist dagegen noch ungeklärt - wir werden im nächsten Abschnitt den Nachweis führen, dass man sie positiv beantworten kann.

Bereits mit wenigstelligen Booleschen Funktionen kann man relativ komplexe Zusammenhänge ausdrücken. Ein deutliches Indiz dafür ist die doppelt exponentiell wachsende Anzahl der n–stelligen Booleschen Funktionen. Die Bestimmung dieser Anzahl ist relativ einfach:
Allgemein beschreibt die Potenz k^m die Anzahl der Funktionen, die einen bestimmten m–elementigen Definitionsbereich A in einen k–elementigen Wertebereich B abbilden, denn man hat für jedes der m Elementes aus A genau k Möglichkeiten der Zuordnung ihres Bildes, also insgesamt $\underbrace{k \cdot k \cdot \ldots \cdot k}_{m \text{ mal}} = k^m$ Möglichkeiten.

Im konkreten Fall ist $A = \mathbb{B}^n$ eine 2^n–elementige Menge und $B = \mathbb{B}$ ist 2–elementig. Folglich gibt es $2^{(2^n)}$ n–stellige Boolesche Funktionen. Das enorme Wachstum dieser Größe wird schon an einem kleinen Beispiel deutlich:
Während für $n = 3$ die Anzahl der 3–stelligen Booleschen Funktionen mit $2^8 = 256$ noch recht übersichtlich ist, ergibt sich für $n = 5$ (und $2^n = 2^5 = 32$), dass die Anzahl der 5–stelligen Booleschen mit 2^{32} bereits über 4 Milliarden liegt.

Wie wir bereits wissen, kann für sehr kleine Werte von n die Äquivalenz von Formeln aus $\mathcal{F}_n$ leicht durch *Wahrheitstafeln* überprüft werden. Diese Methode ist für große n nicht geeignet, denn die entsprechende Wahrheitstafel hat dann 2^n Zeilen. Eine wichtige Alternative bietet die äquivalente Umformung von Termen unter Verwendung der Gesetze der Booleschen Algebra (Assoziativgesetz, Kommutativgesetz usw.). Formale Grundlage für die Anwendung dieser Gesetze in verschiedenen Situationen ist das folgende Ersetzungstheorem.

Definition: Die Menge der *Subformeln* einer Formel t wird mit $Sf(p)$ bezeichnet und induktiv definiert:

- $Sf(p) = \{p\}$ für alle Primformeln p,
- $Sf(\neg t) = Sf(t) \cup \{\neg t\}$ für alle Formeln t
- $Sf(s \circ t) = Sf(s) \cup Sf(t) \cup \{s \circ t\}$ für alle Formeln s, t und alle zweistelligen Junktoren $\circ$.

Definition: Mit $s[\,t/x]$ bezeichnet man den aus s entstehenden Term, wenn jedes Auftreten der Variable x in s durch den Term t ersetzt wird.

Beispiel: Für den Term $s = x \wedge (y \vee x)$ ergibt sich durch Ersetzung von x durch den Term $t = x \wedge z$ der folgende Term:

$$s[\,t/x] = s[(x \wedge z)/x] = (x \wedge (y \vee x))[(x \wedge z)/x] = (x \wedge z) \wedge (y \vee (x \wedge z))$$

Eine besonders anschauliche Deutung dieser Definition erhält man durch Betrachtung der Syntaxbäume: Der Syntaxbaum des Terms $s[\,t/x]$ entsteht aus dem Syntaxbaum von s, wenn jedes mit x markierte Blatt durch den Syntaxbaum von t ersetzt wird.

Substitutionstheorem: Seien s_1 und s_2 zwei semantisch äquivalente Terme und x eine Variable. Dann gelten für jeden Term t die folgenden Äquivalenzen:

$$t[\,s_1/x] \equiv t[\,s_2/x] \qquad \text{und} \qquad s_1[\,t/x] \equiv s_2[\,t/x].$$

Zum besseren Verständnis sollte man sich noch einmal das Beispiel der Termvereinfachung am Ende von Abschnitt 1.1 ansehen und sich klar machen, in welchem Schritt welche Variante des Substitutionstheorems angewendet wurde.

1.5 Konjunktive und disjunktive Normalformen

Nachdem besprochen wurde, wie man Boolesche Terme unter einer Belegung der Variablen interpretiert und dass somit jeder Boolesche Term eine Boolesche Funktion repräsentiert, wollen wir jetzt den umgekehrten Weg gehen und für eine gegebene Boolesche Funktion einen passenden Term finden, der diese Funktion repräsentiert. Es wird sich zeigen, dass solche Terme nicht eindeutig sind, aber in jedem Fall kann man aus der Vielzahl der Repräsentanten zwei Standardvertreter hervorheben - die *kanonische konjunktive Normalform* und die *kanonische disjunktive Normalform*.

Definition (Literale, Maxterme, Minterme und Normalformen):

- Ein *Literal* ist eine Variable $x_i \in V$ oder deren Negation.
- Ein *Maxterm* ist eine Disjunktion von Literalen (dazu zählt man auch ein einzelnes Literal) und ein *Minterm* ist eine Konjunktion von Literalen (dazu zählt man auch ein einzelnes Literal).
- Wenn man vereinbart, dass $V = V_n$ die endliche Variablenmenge $\{x_1, x_2, \ldots, x_n\}$ bezeichnet, dann nennt man einen Minterm (bzw. Maxterm) *vollständig*, wenn er für jedes $1 \leq i \leq n$ entweder das Literal x_i oder das Literal $\neg x_i$ enthält.
- Eine Disjunktion von Mintermen wird *disjunktive Normalform* (kurz DNF) genannt.
- Eine Konjunktion von Maxtermen wird *konjunktive Normalform* (kurz KNF) genannt.

Beispiele: $(x_1 \wedge x_2) \vee (\neg x_1 \wedge x_3) \vee x_2$ 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 DNF's als auch KNF's.

Zur Vereinfachung der Beschreibung vereinbaren wir die folgende Notation: $x_i^1 = x_i$ und $x_i^0 = \neg x_i$ für alle $x_i \in V$.

Satz: Jede Boolesche Funktion $f : \mathbb{B}^n \longrightarrow \mathbb{B}$ ist durch eine DNF d_f und durch eine KNF k_f repräsentierbar, wobei:

$$\begin{aligned} d_f &= \bigvee_{(b_1,\ldots,b_n)\in f^{-1}(1)} x_1^{b_1} \wedge \ldots \wedge x_n^{b_n} \\ k_f &= \bigwedge_{(b_1,\ldots,b_n)\in f^{-1}(0)} x_1^{\neg b_1} \vee \ldots \vee x_n^{\neg b_n} \end{aligned}$$

Man nennt d_f die kanonische DNF und k_f die kanonische KNF von f. Im Spezialfall $f^{-1}(1) = \emptyset$ setzen wir $d_f = 0$ und im Spezialfall $f^{-1}(0) = \emptyset$ setzen wir $k_f = 1$.

Beweis: An dieser Stelle müssen wir noch einmal auf die Bijektion zwischen der Menge aller Belegungen $\omega : V_n \longrightarrow \mathbb{B}$ und der Menge der n-Tupel $\mathbb{B}^n$ eingehen, welche mit ψ_n bezeichnet wurde, d.h. $\psi_n(\omega)$ ist das n–Tupel $(\omega(x_1), \ldots, \omega(x_n))$ und für das

Tupel $(b_1, \ldots, b_n)$ bezeichnen wir mit $\psi_n^{-1}(b_1, \ldots, b_n)$ die Belegung $\omega : V_n \longrightarrow \mathbb{B}$ mit $\omega(x_i) = b_i$ für alle i zwischen 1 und n.

Zuerst überzeugen wir uns davon, dass der Minterm $x_1^{b_1} \wedge \ldots \wedge x_n^{b_n}$ für die konkrete Belegung $\psi_n^{-1}(b_1, \ldots, b_n)$ wahr (1) und für alle anderen Belegungen falsch (0) ist. In der Tat wird eine Konjunktion von Literalen genau dann 1, wenn jedes Literal 1 ist und diese Situation wird nur bei der Belegung $\psi_n^{-1}(b_1, \ldots, b_n)$ erreicht. Die Disjunktion über alle $(b_1, \ldots, b_n) \in f^{-1}(1)$ führt dazu, dass alle Belegungen, die aus dieser Menge kommen, durch "ihren" Minterm akzeptiert (auf 1 bewertet) werden, während Tupel aus $f^{-1}(0)$ von allen Mintermen aus d_f verworfen (auf 0 bewertet) werden.

Der Beweis für k_f ist analog: Jede Disjunktion $x_1^{\neg b_1} \vee \ldots \vee x_n^{\neg b_n}$ wird für die konkrete Belegung $\psi_n^{-1}(b_1, \ldots, b_n)$ falsch (0) und wahr für alle anderen Belegungen. Durch die Disjunktion über alle $(b_1, \ldots, b_n) \in f^{-1}(0)$ erzeugt man eine Formel, welche die vorgegebene Funktion f repräsentiert.

Die Korrektheit der Formeln in den beiden Spezialfällen ist offensichtlich. An dieser Stelle wird klar, warum wir auf die Verwendung von 0 und 1 als Primformeln Wert gelegt haben. Es ist aber genauso offensichtlich, dass man zur Repräsentation der konstanten Booleschen Funktionen auch auf diese Primterme verzichten kann, denn $1 \equiv x_1 \vee \neg x_1$ und $0 \equiv x_1 \wedge \neg x_1$.

Beispiel: In der folgenden Tabelle ist eine 3–stellige Boolesche Funktion f dargestellt, zusammen mit den vollständigen Mintermen von Tupeln aus $f^{-1}(1)$ und den vollständigen Maxtermen von Tupeln 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	$x_1 \wedge x_2 \wedge x_3$	—

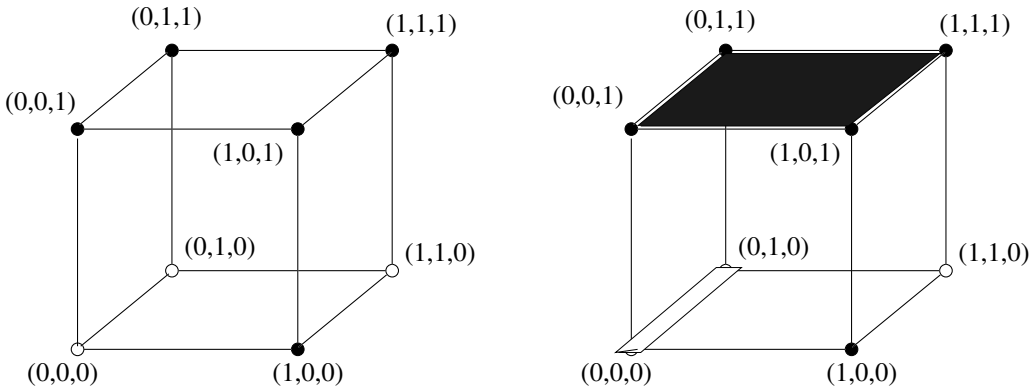
Daraus resultieren die folgende kanonische DNF und KNF:

$$\begin{aligned} d_f &= (\neg x_1 \wedge \neg x_2 \wedge x_3) \vee (\neg x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge \neg x_2 \wedge \neg x_3) \vee (x_1 \wedge \neg x_2 \wedge x_3) \\ &\quad \vee (x_1 \wedge x_2 \wedge x_3) \\ k_f &= (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3) \end{aligned}$$

Zum Verständnis von DNF und KNF kann man auch geometrische bzw. graphentheoretische Hilfsmittel einsetzen: Die Menge der n-Tupel $\{0,1\}^n$ bildet auch die Knotenmenge des n-dimensionalen Würfelgraphen Q_n. Jede Konjunktion von k Literalen

repräsentiert einen $(n-k)$-dimensionalen Unterwürfel. Eine DNF muss durch die Unterwürfel der in ihr auftretenden Minterme die Knotenmenge $f^{-1}(1)$ überdecken. In der kanonischen DNF wird jeder Knoten einzeln (als 0-dimensionaler Unterwürfel) überdeckt. Man sieht, dass man diese kanonische DNF vereinfachen kann, wenn mehrere Knoten aus $f^{-1}(1)$ einen höherdimensionalen Unterwürfel bilden.

Die folgende Abbildung veranschaulicht diese Idee für das oben besprochene Beispiel. Auf der linken Seite sind alle Tupel, die auf 1 abgebildet werden, mit vollen (schwarzen) Kreisen dargestellt, Tupel, die auf 0 abgebildet werden, durch einen Kreisring mit weißer Fläche.
Auf der rechten Seite wird hervorgehoben, dass die vier Tupel auf der oberen Seite alle gleichgefärbt sind. Deshalb kann man die vier Minterme dieser Tupel durch den einfachen Minterm x_3 ersetzen. Analog können die zwei Maxterme der weiß markierten Kante durch den Maxterm $x_1 \vee x_3$ ersetzt werden.



Verfolgt man diese anschauliche Idee der Vereinfachung zurück zur formalen Darstellung, dann kann man jeweils zwei Minterme (bzw. Maxterme) t und t' zu einem Minterm (bzw. Maxterm) s zusammenfassen, wenn t und t' identisch sind bis auf eine Variable x, die unnegiert in t und negiert in t' vorkommt. Der gemeinsame, identische Teil von t und t' bildet dann den Ersatzterm s. Diese Regel ist ein Spezialfall des folgenden leicht zu beweisenden Fakts.

Satz: Sei s ein beliebiger Term und x eine Variable, dann ist s logisch äquivalent zu den Termen $(s \vee x) \wedge (s \vee \neg x)$ und $(s \wedge x) \vee (s \wedge \neg x)$.

Im Allgemeinen wird man durch die wiederholte Anwendung dieser Regel nur einen gewissen Grad an Vereinfachung der kanonischen Normalformen erreichen, aber nicht zur einfachsten DNF bzw. KNF kommen.

1.6 Vollständige Signaturen

Definition:

- Symbole, die man zur Bildung von Formeln aus Variablen einsetzt, bezeichnet man als *logische Junktoren*. Insbesondere muss man auch die elementaren

Aussagen *true* und $false$ bzw. die Werte 1 und 0 als 0–stellige Junktoren betrachten, wenn man sie zur Formelbildung verwenden will.

- Eine *logische Signatur* ist eine Menge von Junktoren.
- Eine logische Signatur ist *funktional vollständig*, wenn jede Boolesche Funktion durch eine mit dieser Signatur gebildeten Formel repräsentierbar ist. Man spricht in diesem Fall auch von einer vollständigen Basis.
- Die Signatur $\{\neg, \vee, \wedge\}$ heißt *Boolesche Signatur*. Bei dieser Festlegung werden keine elementaren Aussagen zum Formelaufbau verwendet, insbesondere auch nicht 0 und 1.

Satz: Die Boolesche Signatur sowie die Signaturen $\{\neg, \wedge\}$ und $\{\neg, \vee\}$ sind funktional vollständig.

Beweis: Die Vollständigkeit der Booleschen Signatur folgt aus der Existenz von DNF's bzw. KNF's. Man beachte, daß die Formeln $x \wedge \neg x$ und $x \vee \neg x$ die konstanten Funktionen 0 und 1 repräsentieren.

Die Vollständigkeit einer anderen Signatur Σ wird in der Regel damit nachgewiesen, dass man die Operationen der Booleschen Signatur mit Operationen aus Σ logisch äquivalent darstellen kann. So kann man für die Signatur $\Sigma_1 = \{\neg, \wedge\}$ die fehlende Disjunktion mit der deMorganschen Regel durch Negation und Konjunktion beschreiben und für die Signatur $\Sigma_2 = \{\neg, \vee\}$ die fehlende Konjunktion mit der deMorganschen Regel durch Negation und Disjunktion. Somit wird die Vollständigkeit der Signaturen Σ_1 und Σ_2 auf die Vollständigkeit der Booleschen Signatur zurückgeführt.

Es gibt sogar einelementige Signaturen, die funktional vollständig sind, z.B. die Signatur $\{\,|\,\}$, wobei $|$ den sogenannten NAND–Junktor ($x|y \equiv \neg(x \wedge y)$) bezeichnet. Hier beginnt die Begründung mit der Negation durch die Äquivalenz $\neg x \equiv x|x$. Im zweiten Schritt gibt man eine Konstruktion für die Konjunktion an: $x \wedge y \equiv \neg(x|y) \equiv (x|y)|(x|y)$. Damit hat man das Problem auf die Vollständigkeit der Signatur Σ_1 zurückgeführt.

Es ist leicht zu begründen, daß die Signaturen $\{\vee\}$, $\{\wedge\}$ und $\{\vee, \wedge\}$ nicht funktional vollständig sind, denn die konstanten Booleschen Funktionen sind nicht darstellbar. Man kann auch zeigen, dass die Signatur $\{0, 1, \vee, \wedge\}$ nicht vollständig ist, aber das erfordert einen wesentlich höheren Aufwand.

Weitere Beispiele für unvollständige Signaturen sind $\{\rightarrow\}$, $\{\leftrightarrow\}$, $\{\neg, \leftrightarrow\}$, $\{\oplus\}$, $\{\neg, \oplus\}$ und $\{1, \oplus\}$. Bis auf die erste sind alle hier genannten Signaturen Teilmengen der unvollständigen Signatur $\{0, 1, \neg, \leftrightarrow, \oplus\}$.

2 Grundbegriffe der Mengenlehre

2.1 Mengen und Operationen auf Mengen

Moderne Mengentheorie wird in Form eines axiomatischen Kalküls betrieben. Dieser Ansatz hat aber den Nachteil, daß einfache inhaltliche Fragen oft durch einen technisch komplizierten Apparat verdeckt werden. Wir werden uns deshalb auf die Entwicklung einer "naiven" Mengenlehre beschränken, die als sprachliches Werkzeug für die nachfolgenden Teile der Vorlesung völlig ausreichend ist.
Nach Georg Cantor ist eine *Menge* "eine Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens (welche die Elemente der Menge genannt werden) zu einem Ganzen".
Der Sachverhalt, dass ein Objekt a Element einer Menge A ist, wird durch $a \in A$ dargestellt, anderenfalls schreibt man $a \notin A$. Zwei Mengen A und B sind *gleich*, wenn sie die gleichen Elemente besitzen, d.h. wenn für alle a gilt: $a \in A$ dann und nur dann, wenn $a \in B$.

Darstellungen von Mengen

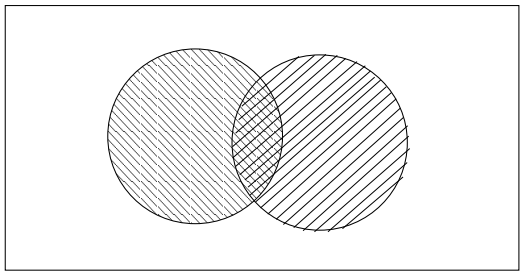
a) Mengen können durch *Auflistung ihrer Elemente* in geschweiften Klammern dargestellt werden. Das betrifft insbesondere endliche Mengen, wie z.B. $A = \{2, 3, 5, 7\}$ oder $B = \{rot, gelb, blau\}$. Dabei ist die Reihenfolge der Elemente in der Auflistung ohne Bedeutung. Auch die Mehrfachnennung von Elementen ist erlaubt (sollte aber zur Vermeidung von Missverständnissen möglichst vermieden werden), sie hat aber nur Einfluss auf die Darstellung der Menge und nicht auf die Menge selbst, z.B. $\{2, 3, 5, 7\} = \{5, 7, 3, 2, 2, 5, 2\}$.
Wir vereinbaren, dass auch unendliche Mengen durch Auflistung dargestellt werden können, sofern dies unmissverständlich ist, wie z.B. $\{0, 1, 2, 3, \ldots\}$ für die natürlichen Zahlen oder $\{2, 4, 6, 8, \ldots\}$ für die positiven, geraden Zahlen.

b) Die in der Mathematik gebräuchlichste Darstellungsform von Mengen beruht auf dem sogenannten *Abstraktionsprinzip*, nach dem man Mengen – im Sinne der Cantorschen Definition – durch wohlbestimmte Eigenschaften definieren kann. Dazu werden Prädikate $P(x)$ über einem festgelegten Individuenbereich für x benutzt. Dann wird mit $\{x \mid P(x)\}$ oder (wenn der Bereich B explizit genannt werden soll) mit $\{x \in B \mid P(x)\}$ die Menge bezeichnet, die sich aus allen Individuen aus dem Bereich zusammensetzt, für die $P(x)$ wahr ist. Man bezeichnet diese Darstellungsart von Mengen nach den Mathematikern Ernst Zermelo und Abraham Fraenkel auch als ZF-Notation.

c) Zur Veranschaulichung können Mengen durch sogenannte *Venn–Diagramme* als Kreisscheiben oder andere Flächen in der Ebene dargestellt werden.

Oft ist es sinnvoll, den zu betrachtenden Individuenbereich generell festzulegen, z.B. wenn man nur Mengen von natürlichen Zahlen betrachten will. Ein solcher Bereich wird *Universum* genannt und allgemein mit U bezeichnet. Es ist klar, dass Aussageformen über U immer Teilmengen von U definieren. Im folgenden Venn-Diagramm sind zwei Mengen A und B als Kreisflächen in dem durch das Rechteck symbolisierten

Universum U dargestellt:



Bemerkung: In manchen Anwendungen (z.B. für die Buchstaben in einem Wort) benötigt man auch ein Konstrukt, in dem Elemente auch mehrfach auftreten können. In diesem Fall sprechen wir von einer *Multimenge*. Obwohl sich diese konzeptionell wesentlich von einer Menge unterscheidet, wird in der Regel die gleiche Notation verwendet, was natürlich zu Missverständnissen führen kann. Deshalb werden wir bei Verwendung von Multimengen diesen Begriff explizit nennen, anderenfalls ist immer eine Menge gemeint. Demnach sind $\{b, a, b, b\}$ und $\{a, b\}$ (ohne Zusatz) zwei identische Mengen, die Multimengen $\{b, a, b, b\}$ und $\{a, b\}$ unterscheiden sich, aber die Multimengen $\{b, a, b, b\}$ und $\{a, b, b, b\}$ sind wiederum identisch.

Definition: Eine Menge A ist *Teilmenge* (oder *Untermenge*) einer Menge B (Schreibweise $A \subseteq B$), wenn aus $a \in A$ auch $a \in B$ folgt.

Die Teilmengenrelation entspricht also einer Implikation der definierenden Prädikate. Deshalb kann man aus den Eigenschaften der logischen Implikation zwei elementare Eigenschaften der Teilmengenrelation ableiten:

- Die Mengen A und B sind gleich genau dann, wenn $A \subseteq B$ und $B \subseteq A$.
- Ist A eine Teilmenge von B und B eine Teilmenge von C, dann ist auch A eine Teilmenge von C.

Die folgende Definition enthält eine Zusammenfassung der wichtigsten Mengenoperationen. Man beachte insbesondere den Zusammenhang zu den entsprechenden logischen Operationen.

Definition:

- Zwei Mengen A und B sind *disjunkt*, wenn sie keine gemeinsamen Elemente besitzen, d.h wenn aus $a \in A$ folgt $a \notin B$.
- Die *Vereinigung* $A \cup B$ der Mengen A und B besteht aus allen Objekten, die Elemente von A oder von B sind, dh. $A \cup B = \{x \mid x \in A \ \vee \ x \in B\}$.
- Der *Durchschnitt* $A \cap B$ der Mengen A und B besteht aus allen Objekten, die Elemente von A und von B sind, dh. $A \cap B = \{x \mid x \in A \ \wedge \ x \in B\}$.

- Die *Differenz* $A \setminus B$ der Mengen A und B besteht aus allen Objekten, die Elemente von A, aber nicht von B sind, dh. $A \setminus B = \{x \mid x \in A \ \wedge \ x \notin B\}$.
- Die Vereingung der Mengendifferenzen $A \setminus B$ und $B \setminus A$ wird die *symmetrische Differenz* aus A und B genannt und mit $A \oplus B$ oder auch mit $A \div B$ bezeichnet.
- Die Menge, die kein Element enthält, wird *leere Menge* genannt und mit $\emptyset$ bezeichnet.
- Ist A Teilmenge eines festgelegten Universums U, dann ist das *Komplement* von A definiert als $U \setminus A$. Es wird mit $\overline{A}$ bezeichnet.

Der Zusammenhang zwischen Teilmengenbeziehungen sowie Operationen auf Mengen und den entsprechenden logischen Operationen wird noch einmal in der folgenden Tabelle zusammengefasst:

Mengenlehre		Logik	
Gleichheit	$A = B$	$x \in A \leftrightarrow x \in B$	Äquivalenz
Inklusion	$A \subseteq B$	$x \in A \rightarrow x \in B$	Implikation
Vereinigung	$A \cup B$	$x \in A \vee x \in B$	Disjunktion
Durchschnitt	$A \cap B$	$x \in A \rightarrow x \in B$	Konjunktion
Komplement	$A = \overline{B}$	$x \in A \leftrightarrow \neg(x \in B)$	Negation
symmetr. Differenz	$A \oplus B = A \div B$	$x \in A \oplus x \in B$	Antivalenz

Damit können auch - wie im nachfolgenden Satz formuliert - die bekannten Gesetze der Aussagenlogik in die Mengenlehre übertragen werden.

Satz: Folgende Identitäten gelten für alle Untermengen A, B, C eines Universums U:

Kommutativität:	$A \cup B = B \cup A$	und	$A \cap B = B \cap A$
Assoziativität:	$A \cup (B \cup C) = (A \cup B) \cup C$		
	und $A \cap (B \cap C) = (A \cap B) \cap C$		
Distributivität:	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$		
	und $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$		
Idempotenz:	$A \cup A = A$	und	$A \cap A = A$
Dominanz:	$A \cup U = U$	und	$A \cap \emptyset = \emptyset$
Identität:	$A \cup \emptyset = A$	und	$A \cap U = A$
Absorbtion:	$A \cup (A \cap B) = A$	und	$A \cap (A \cup B) = A$
De Morgan'sche Regel:	$\overline{A \cup B} = \overline{A} \cap \overline{B}$	und	$\overline{A \cap B} = \bar{A} \cup \bar{B}$
Komplementierung:	$A \cup \overline{A} = U$	und	$A \cap \overline{A} = \emptyset$
	$\overline{(\overline{A})} = A$	und	$A \setminus B = A \cap \overline{B}$

Auf Grund der Assoziativität kann man bei der Vereinigung (bzw. beim Durchschnitt) von n Mengen $A_1, A_2, \ldots, A_n$ auf Klammerungen verzichten und die folgende

Schreibweise nutzen:

$$A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$$
$$A_1 \cap A_2 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i$$

Die Bildung einer solchen Vereinigung bzw. eines solchen Duchschnitts erfolgt somit (in Analogie zu endlichen Summen oder Produkten von n Zahlen) durch die $(n-1)$-fache Anwendung der einfachen Vereinigungs- bzw. Durchschnittsoperation. Im Gegensatz zu den Zahlen kann man aber für Mengen auch uneingeschränkt unendliche Vereinigungen und Durchschnitte bilden.

Definition: Ist I eine beliebige Menge und ist für jedes $i \in I$ eine Menge A_i gegeben, dann nennen wir die Multimenge dieser Mengen eine *Mengenfamilie* und bezeichnen sie durch $\{A_i \mid i \in I\}$. Die Vereinigung (bzw. der Durchschnitt) dieser Mengenfamilie ist definiert durch

$$\bigcup_{i \in I} A_i = \{x \mid \text{es gibt ein } i \in I, \text{ so dass } x \in A_i\}$$

$$\bigcap_{i \in I} A_i = \{x \mid \text{für alle } i \in I \text{ gilt } x \in A_i\}$$

Beispiel: Sei $I = \{2, 3, 4, \ldots\}$ die Menge der natürlichen Zahlen, die größer oder gleich 2 sind und sei $A_i = \{n \in \mathbb{N} \mid i | n\}$ die Menge der durch i teilbaren Zahlen für jedes $i \in I$. Dann ergibt sich

$$\bigcup_{i \in I} A_i = \{0, 2, 3, \ldots\} = \mathbb{N} \setminus \{1\} \quad \text{und} \quad \bigcap_{i \in I} A_i = \{0\}.$$

Zur Begründung betrachten wir jeweils drei Fälle. Da die 0 durch jede andere Zahl teilbar ist, liegt sie in jedem A_i und somit auch in der Vereinigung und im Durchschnitt. Die 1 ist durch keine andere Zahl teilbar, gehört also zu keinem A_i und liegt somit weder im Durchschnitt noch in der Vereingung. Da für jedes $i \in I$ gilt $i \in A_i$ und $i \notin A_{2i}$ gilt, gehört i zur Vereinigung, aber nicht zum Durchschnitt.

Bemerkung: Für die Vereinigung und den Durchschnitt ist es unerheblich, ob wir die Mengenfamilie als Multimenge oder als Menge auffassen, für andere Betrachtungen könnte dieser Unterschied aber durchaus bedeutsam sein.

Definition: Eine Familie $\{A_i \mid i \in I\}$ von nichtleeren Mengen wird *Partition* oder *Zerlegung* einer Menge A genannt, falls

1. $A = \bigcup_{i \in I} A_i$
2. Für beliebige, voneinander verschiedene $i, j \in I$ gilt $A_i \cap A_j = \emptyset$.

Beispiele: $\{\{a, c\}, \{b, e, f\}, \{d\}\}$ ist eine Partition der Menge $\{a, b, c, d, e, f\}$ in drei Teilmengen. Die Menge der geraden und die Menge der ungeraden Zahlen bilden eine Partition der Menge $\mathbb{N}$ in zwei Teilmengen.

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

Satz: Ist A eine endliche, n-elementige Menge, dann hat die Potenzmenge $\mathcal{P}(A)$ genau 2^n Elemente.

2.2 Das Kartesische Produkt und Relationen

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

Definition: 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 $b \in B$, als Formel:

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

Beispiel: $\{a, b, c\} \times \{1, 2\} = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$.

Definition: Eine Untermenge R eines kartesischen Produkts $A \times B$ wird *binäre Relation* oder kurz *Relation* zwischen A und B genannt. Im Fall $(a, b) \in R$ sagt man, dass a in Relation zu b steht. Für $(a, b) \in R$ kann auch $a\, R\, b$ geschrieben werden. Diese alternative Notation wird vor allem dann verwendet, wenn der Relationsbezeichner kein Buchstabe sondern ein Relationssymbol wie $<, \leq, \equiv$ oder $\sim$ ist.
Eine Untermenge R eines kartesischen Produkts der Form $A \times A$ wird (binäre) Relation auf A (oder über A) genannt.

Die ersten drei Relationen in den folgenden Beispielen sind generisch, d.h. man kann sie über beliebigen Grundmengen betrachten:

- $\emptyset \subseteq A \times B$ wird *leere Relation* genannt.
- $A \times B$ wird *Allrelation* zwischen A und B genannt.
- Die Menge $\{(a, a) \mid a \in A\}$ wird die *identische Relation* über A genannt und kurz mit Id_A bezeichnet.
- Die Teilbarkeitsrelation $|$ kann man als Relation über den natürlichen Zahlen (aber auch über den ganzen Zahlen) betrachten. Wie bereits besprochen, ist diese Relation wie folgt definiert:

$$\forall\, a, b \in \mathbb{N} \quad a \,|\, b \quad \Longleftrightarrow \quad \exists c \in \mathbb{N} \quad b = a \cdot c$$

- Über den natürlichen Zahlen $\mathbb{N}$, den ganzen Zahlen $\mathbb{Z}$, den rationalen Zahlen $\mathbb{Q}$ und den reellen Zahlen $\mathbb{R}$ kennen wir eine Reihe von Vergleichsrelationen, nämlich $<, \leq, \geq, >$.
- Ist A die Menge aller Informatikstudenten an der FU Berlin und B die Menge aller Pflichtmodule im Informatikstudium, dann ist
  $R = \{(a, b) \in A \times B \mid$ Student a hat das Modul b abgeschlossen $\}$ eine binäre Relation.
- Jede Abbildung $f : A \longrightarrow B$ kann auch als binäre Relation
  $f = \{(a, b) \in A \times B \mid a \in A \wedge b = f(a)\}$ gesehen werden.

Zur Darstellung von Relationen sind verschiedene Methoden gebräuchlich:

- Darstellungen in Tabellenform, bei denen für jedes $a \in A$ eine Spalte und für jedes $b \in B$ eine Zeile angelegt wird. Die Zelle in der Spalte von a und der Zeile von b wird mit einer 1 gefüllt, wenn $a\,R\,b$ gilt, und sonst mit einer 0. (Verwendung in relationalen Datenbanken);
- Anschauliche Darstellungen durch Diagramme in einem Rechteck;
- Bipartite Graphen, bei denen die Elemente aus A und B als Knoten getrennt auf zwei Seiten gezeichnet werden, wobei zwei Elemente, die zueinander in Relation stehen, durch eine Kante (Verbindungsstrecke) verbunden werden.
- Für Relationen über einer Menge A kann man gerichtete Graphen verwenden. Dabei wird jedes Element von A als ein Knoten und jedes Paar $(a, b) \in R$ durch eine von a nach b gerichtete Kante gezeichnet.

Relationsoperationen

1. Sind R und R' Relationen zwischen A und B, dann sind auch die Vereinigung $R \cup R'$, der Durchschnitt $R \cap R'$ sowie das Komplement $\overline{R} = (A \times B) \setminus R$ Relationen zwischen A und B.
2. Die zu einer Relation $R \subseteq A \times B$ *inverse Relation* $R^{-1} \subseteq B \times A$ ist definiert durch $R^{-1} = \{(b, a) \in B \times A \mid (a, b) \in R\}$.
3. Die *Verkettung* oder *Komposition* $R \circ S$ von zwei Relationen $R \subseteq A \times B$ und $S \subseteq B \times C$ ist definiert durch

$$\{(a, c) \in A \times C \mid \text{es gibt ein } b \in B \text{ mit } (a, b) \in R \text{ und } (b, c) \in S\}.$$

Beispiele:

1) Wir betrachten die Vergleichsrelationen $<$, $\leq$, $\geq$ und $=$ über den natürlichen Zahlen $\mathbb{N}$. Offensichtlich ist die Vereinigung der Relationen $<$ und $=$ die Relation $\leq$. Das Komplement der Relation $<$ ist die Relation $\geq$. Der Durchschnitt der Relationen $\leq$ und $\geq$ ist die identische Relation $Id_{\mathbb{N}}$. Die zu $\leq$ inverse Relation ist $\geq$, die identische Relation ist zu sich selbst invers.

2) Sei M die Menge aller Menschen und $R \subseteq M \times M$ "Elternrelation", also $a\,R\,b$, falls a Vater oder Mutter von b ist. Dann kann man die inverse Relation R^{-1} sowie die Verkettungen $R \circ R$, $R \circ R^{-1}$ und $R^{-1} \circ R$ wie folgt charakterisieren:

- $a\,R^{-1}\,b$, falls a Kind von b ist,
- $a\,R \circ R\,b$, falls a Großvater oder Großmutter von b ist,
- $a\,R \circ R^{-1}\,b$, falls a und b ein gemeinsames Kind haben oder falls $a = b$ und a hat ein Kind,

- $a\, R^{-1} \circ R\, b$, falls $a = b$ oder falls a und b Geschwister oder Halbgeschwister sind.

Eigenschaften von Relationen über Mengen

Definition: Sei R eine Relation über A.

- R ist *reflexiv*, falls für jedes $a \in A$ gilt, dass $a\, R\, a$, d.h. $Id_A \subseteq R$.
- R ist *symmetrisch*, falls aus $a\, R\, b$ folgt, dass $b\, R\, a$, d.h. $R^{-1} \subseteq R$.
- R ist *transitiv*, falls aus $a\, R\, b$ und $b\, R\, c$ folgt, dass $a\, R\, c$, d.h. $R \circ R \subseteq R$.
- R ist *antisymmetrisch*, falls aus $a\, R\, b$ und $b\, R\, a$ die Gleichheit von a und b folgt, d.h. $R \cap R^{-1} \subseteq Id_A$.
- R ist *asymmetrisch*, falls aus $a\, R\, b$ folgt, dass $(b, a) \notin R$, d.h. $R \cap R^{-1} = \emptyset$.

Beispiele:
1) Die Vergleichsrelationen $\leq$ und $\geq$ sind über $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ und $\mathbb{R}$ reflexiv, transitiv und antisymmetrisch. Die Relationen $<$ und $>$ sind nicht reflexiv, aber transitiv, antisymmetrisch und asymmetrisch.
2) Die oben definierte Teilbarkeitsrelation ist reflexiv und transitiv über $\mathbb{N}$ und über $\mathbb{Z}$. Sie ist antisymmetrisch über $\mathbb{N}$, aber als Relation über $\mathbb{Z}$ ist sie nicht antisymmetrisch, denn $1|-1$ und $-1|1$, aber $1 \neq -1$.

2.3 Äquivalenzrelationen

Definition: Eine Relation über einer Menge A wird *Äquivalenzrelation* genannt, wenn sie reflexiv, symmetrisch und transitiv ist.

Beispiele:

- Sei $\mathbb{N}$ die Menge der natürlichen Zahlen und R definiert durch $a\, R\, b$, genau dann wenn a und b beim Teilen durch 5 den gleichen Rest haben. Dann ist R eine Äquivalenzrelationüber $\mathbb{N}$.
- Die Relation $\equiv$ der logischen Äquivalenz über der Menge der Booleschen Formeln ist eine Äquivalenzrelation.
- Sei M die Menge aller Menschen und R die Relation, die zwei Menschen in Beziehung setzt, wenn sie denselben Vater und dieselbe Mutter haben. Diese Geschwisterrelation ist offensichtlich eine Äquivalenzrelation. Die Halbgeschwister-Relation ist im Allgemeinen nicht transitiv und somit keine Äquivalenzrelation.

Definition: Ist $R \subseteq A \times A$ eine Äquivalenzrelation und ist $a \in A$, dann nennt man die Menge $\{x \in A \mid xRa\}$ die *Äquivalenzklasse* von a (bezüglich R). Sie wird mit $[a]_R$ oder auch mit $a/_R$ bezeichnet. Ein Element einer Äquivalenzklasse wird *Repräsentant* dieser Klasse genannt.

Lemma: Sei R eine Äquivalenzrelation, dann sind zwei Äquivalenzklassen $[a]_R$ und $[b]_R$ entweder gleich oder disjunkt. Sie sind genau dann gleich, wenn $a\,R\,b$ gilt.

Beweis: Wir verifizieren dazu die folgende Schlusskette:

$$[a]_R \cap [b]_R \neq \emptyset \quad \overset{(1)}{\Longrightarrow} \quad a\,R\,b \quad \overset{(2)}{\Longrightarrow} \quad [a]_R = [b]_R$$

(1) Sei $c \in [a]_R \cap [b]_R$, dann gilt cRa und cRb, wegen der Symmetrie auch aRc und wegen der Transitivität auch aRb.
(2) Sei d ein beliebiges Element aus $[a]_R$ und gelte aRb. Dann gilt dRa und wegen der Transitivität auch dRb. Damit liegt d in der Äquivalenzklasse$[b]_R$ und folglich ist $[a]_R \subseteq [b]_R$. Wegen der Symmetrie kann man aber auch die Rollen von a und b vertauschen und somit $[b]_R \subseteq [a]_R$ ableiten, woraus letztlich die Gleichheit der beiden Äquivalenzklassen folgt. □

Die erste Aussage des folgenden Satzes kann als einfache Schlussfolgerung aus dem Lemma abgeleitet werden.

Satz: Ist $R \subseteq A \times A$ eine Äquivalenzrelation, dann bildet die Menge aller Äquivalenzklassen eine Partition von A. Umgekehrt, ist eine Partition $\{A_i \mid i \in I\}$ von A gegeben, dann ist die durch

$$a\,R\,b \iff \exists i \in I \qquad a \in A_i \ \wedge \ b \in A_i$$

definierte Relation R eine Äquivalenzrelation.

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

Beispiel: Wir verallgemeinern das Beispiel der Relation, die durch gleiche Reste beim Teilen durch 5 definiert ist, indem wir an Stelle der 5 einen beliebigen ganzzahligen Teiler $d \geq 2$ betrachten. Man nennt zwei Zahlen $a, b \in \mathbb{N}$ *kongruent modulo d*, wenn sie beim Teilen durch d den gleichen Rest haben (Kongruenzrelationen sind spezielle Äquivalenzrelationen, die verträglich mit bestimmten Operationen sind, in diesem Fall mit Addition und Multiplikation). Zur Notation verwendet man $a \bmod d$ für den Rest beim Teilen von a durch d und den Ausdruck $a \equiv b \pmod d$ für die Tatsache, dass a und b beim Teilen durch d den gleichen Rest haben, also dass $a \bmod d$ und $b \bmod d$ gleich sind.
Bei der Kongruenzrelation modulo d werden die einzelnen Äquivalenzklassenjeweils durch die Zahlen mit gleichem Rest gebildet, was im Beispiel $d = 5$ zu der folgenden Partition von $\mathbb{N}$ führt:

$$\{\{0, 5, 10 \ldots\}, \{1, 6, 11, \ldots\}, \{2, 7, 12, \ldots\}, \{3, 8, 13, \ldots\}, \{4, 9, 14, \ldots\}\}$$

Offensichtlich bilden die Reste {0,1,2,3,4} ein Repräsentantensystem (das sogenannte Standard–Repräsentantensystem), aber auch die Menge $\{3, 10, 7, 21, 9\}$ ist ein Repräsentantensystem.

Satz: Die identische Relation Id_A und die Allrelation $A \times A$ sind Äquivalenzrelationen. Sind R und R' Äquivalenzrelationen in A, dann ist auch $R \cap R'$ eine Äquivalenzrelation in A.

Achtung: Die letzte Aussage gilt im Allgemeinen nicht für Vereinigungen. Als Gegenbeispiel kann man die Kongruenzrelationen modulo 2 und modulo 3 betrachten. Offensichtlich ist das Paar $(1, 6)$ nicht in der Vereinigung, denn 1 und 6 haben sowohl beim Teilen durch 2 als auch beim Teilen durch 3 verschiedene Reste. Andererseits sind die Paare $(1, 4)$ - gleicher Rest beim Teilen durch 3 - und $(4, 6)$ - gleicher Rest beim Teilen durch 2 - in der Relationsvereinigung. Folglich ist diese Vereinigung nicht transitiv.

Allgemein kann jede Relation $R \subseteq A \times A$ durch die folgenden 3 Schritte zu einer Äquivalenzrelation erweitert werden:

1) reflexiver Abschluss: $R_r = R \cup Id_A$
2) symmetr. Abschluss: $R_{rs} = R_r \cup R_r^{-1} = R \cup R^{-1} \cup Id_A$
3) transitiver Abschluss: $R_{rst} = R_{rs} \cup R_{rs} \circ R_{rs} \cup R_{rs} \circ R_{rs} \circ R_{rs} \cup \ldots = \bigcup_{i=1}^{\infty} R_{rs}^i$

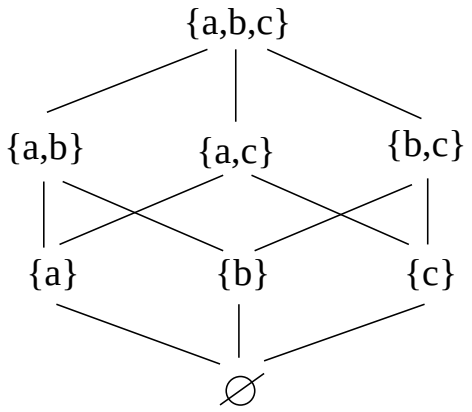
wobei R_{rs}^i die i-fache Verkettung von R_{rs} ist.

Beispiel: Der reflexiv–symmetrisch–transitive Abschluss der Nachfolgerrelation auf den natürlichen Zahlen ist die Allrelation. Setzt die Relation R jede natürliche Zahl n zum Nachfolger ihres Nachfolgers, also zu $n+2$ in Beziehung, dann ist der reflexiv–symmetrisch–transitive Abschluss dieser Relation die Kongruenz modulo 2.

2.4 Ordnungsrelationen

Definition: Eine Relation R über einer Menge A, die reflexiv, transitiv und antisymmetrisch ist, wird *Halbordnungsrelation* oder auch *partielle Ordnungsrelation* genannt. Das Paar (A, R) nennt man eine *halb– (partiell) geordnete Menge* oder kurz *poset* als Abkürzung für partially ordered set.

Endliche, halbgeordnete Mengen werden oft durch sogenannte *Hasse–Diagramme* dargestellt. Dabei werden die Elemente der Menge als Punkte in der Ebene gezeichnet, wobei direkte Nachfolger jeweils höher als ihre Vorgänger liegen und mit ihnen durch ein Liniensegment verbunden sind. Formal betrachtet beschreibt das Hasse–Diagramm eines Posets (A, R) die kleinste Unterrelation von R, deren reflexiver und transitiver Abschluss R ergibt. Die folgende Abbildung zeigt das Hasse–Diagramm der Potenzmenge einer 3–elementigen Menge $M = \{a, b, c\}$:



Beispiele:
1) Für jede beliebige Menge M ist $(\mathcal{P}(M), \subseteq)$ eine halbgeordnete Menge.
2) Die Teilbarkeitsrelation $|$ ist eine Halbordnungsrelation in der Menge der positiven ganzen Zahlen $\mathbb{Z}^+$.
3) In der Menge der reellen Zahlen $\mathbb{R}$ ist die Relation $\leq$ eine Halbordnungsrelation.
4) Die Menge der Wörter einer Sprache wird durch die "lexikographische Ordnung" geordnet.

Zwei Begriffe sind eng verwandt mit partiellen Ordnungsrelationen: totale Ordnungsrelationen und strikte (oder strenge) Ordnungsrelationen. Diese Begriffe werden durch die folgenden Definitionen genauer erläutert.

Definition: Zwei Elemente a und b einer halbgeordneten Menge (A, R) nennt man *vergleichbar*, falls $a\,R\,b$ oder $b\,R\,a$ gilt. Anderenfalls nennt man sie unvergleichbar.
Eine Halbordnungsrelation R in einer Menge A wird *totale* (oder auch *lineare*) *Ordnungsrelation* genannt, wenn jedes Paar von Elementen vergleichbar ist.

Beispiele: In den obigen Beispielen sind die Relationen aus 1) und 2) keine totalen Ordnungsrelationen. So sind für $M = \{a, b, c\}$ die Untermengen $\{a\}$ und $\{c\}$ unvergleichbar bezüglich der Teilmengenrelation. Die Zahlen 6 und 20 sind unvergleichbar bezüglich der Teilbarkeitsrelation. Dagegen ist $\leq$ eine totale Ordnungsrelation für die reellen Zahlen. Die lexikographische Ordnung ist eine totale Ordnungsrelation.

Bemerkung: Taucht in der Literatur der Begriff "Ordnungsrelation" auf, so ist darunter in der Regel eine "Halbordnungsrelation" zu verstehen.

Definition: Eine Relation R über einer Menge A wird *strikte* oder *strenge Ordnungsrelation* genannt, wenn sie transitiv und asymmetrisch ist.

Typische Beispiele für strikte Ordnungsrelationen sind die "echt–kleiner"–Relation $<$ oder die Relation, ein echter Teiler zu sein. Generell kann man aus jeder Halbordnungsrelation R über einer Menge A eine strikte Ordnungsrelation $R' = R \setminus Id_A$ ableiten und umgekehrt kann aus jeder strikten Ordnungsrelation durch Vereinigung mit Id_A eine Halbordnungsrelation gemacht werden.

Weitere Begriffe, die wie das Maximum und Minimum aus der Schulmathematik

bekannt sind, müssen im Kontext mit Halbordnungsrelationen noch einmal neu definiert werden.

Definition: Sei $(A, \prec)$ ein Poset, $B \neq \emptyset$ eine Teilmenge von A und $a \in A$.

- a ist eine *obere* (bzw. *untere*) *Schranke* von B, wenn $b \prec a$ (bzw. $a \prec b$) für alle $b \in B$ gilt.
- a ist ein *maximales* (bzw. *minimales*) *Element* von B, wenn $a \in B$ und es kein von a verschiedenes $b \in B$ gibt, so dass $a \prec b$ (bzw. $b \prec a$).
- a wird *größtes Element* von B (bzw. *kleinstes Element* von B) genannt, wenn $a \in B$ und a eine obere (bzw. untere) Schranke von B ist.
- a wird *Supremum* oder *obere Grenze* von B (bzw. *Infimum* oder *untere Grenze* von B) genannt, wenn die Menge $O(B)$ der oberen Schranken von B (bzw. die Menge $U(B)$ der unteren Schranken von B) nicht leer ist und a das kleinste Element von $O(B)$ (bzw. das größte Element von $U(B)$) ist.

Bemerkungen zur Definition:

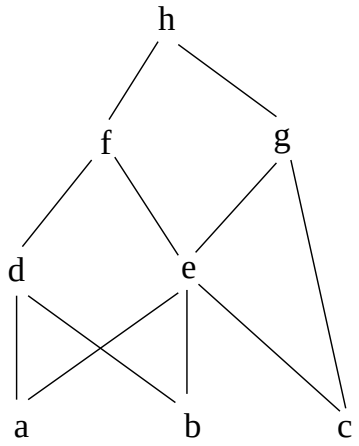
- Eine Teilmenge B kann mehrere maximale bzw. mehrere minimale Elemente haben. Ist B endlich, dann gibt es mindestens ein maximales und ein minimales Element von B.Dagegen gibt es einfache Beispiele von endlichen Mengen, für die es keine obere und untere Schranken und somit auch weder größte oder kleinste Elemente noch obere oder untere Grenzen gibt.
- Existiert ein größtes (bzw. kleinstes) Element von B, dann ist es eindeutig. Somit ist auch die obere bzw. untere Grenze von B eindeutig sofern sie überhaupt existiert.
- Etwas verwirrend ist die Konvention, das größte (bzw. kleinste) Element von B, sofern es existiert, als *Maximum* (bzw. *Minimum* von B zu bezeichnen. Es gilt der Merksatz, dass das Maximum (wenn existent) immer ein maximales Element ist, aber nicht die Umkehrung.
- Da Maximum, Minimum, Supremum oder Infimum einer Teilmenge B bei Existenz immer eindeutig sind, kann man für sie die Bezeichnungen $max(B)$, $min(B)$, $sup(B)$,$inf(B)$ einführen. Diese stehen dann entweder für ein eindeutig existierendes Element oder für nicht existierend.

Beobachtung: Wenn für eine Teilmenge B einer Halbordnung $(A, \prec)$ das Maximum (bzw. Minumum) existiert, dann existiert auch das Supremum (bzw. das Infimum) und es gilt $sup(B) = max(B)$ (bzw. $inf(B) = min(B)$).

Beweis: Sei a das Maximum von B, dann gilt (i) $a \in B$ und (ii) ist a obere Schranke von B, d.h. $b \prec a$ für alle $b \in B$. Man muss zeigen, dass a in der Menge der oberen Schranken $O(B)$, das kleinste Element ist. Nach Definition muss also a Element und untere Schranke von $O(B)$ sein. Die erste Eigenschaft ist bereits durch

die Voraussetzung (ii) gegeben. Für die zweite Eigenschaft müssen wir $a \prec c$ für alle $c \in O(B)$ nachweisen. Aus $c \in O(B)$ folgt aber $b \prec c$ für alle $b \in B$, also auch insbesondere für $b = a$. Somit ist a auch das Supremum von B.

Beispiel: Wir betrachten die durch nachfolgendes Hasse-Diagramm dargestellte Halbordnung $(A, \prec)$ mit den Teilmengen $B = \{d, e\}$ und $C = \{f, g, h\}$:



- Die Gesamtmenge A hat ein maximales Element, nämlich h und drei minimale Elemente, nämlich a, b und c. Damit is h auch das Maximum und Supremum von A, Minimum und Infimum existieren nicht.
- B hat zwei obere Schranken, nämlich f und h und zwei untere Schranken, nämlich a und b.
- Sowohl d als auch e sind maximale und minimale Elemente von B.
- B hat weder ein Maximum noch ein Minimum noch ein Infimum, aber ein Supremum $sup(B) = f$.
- C hat eine obere Schranke, nämlich h und vier untere Schranken, nämlich a, b, c und e.
- C besitzt ein maximales Element, nämlich h, und die minimalen Elemente f und g.
- C hat kein Minimum, aber $inf(C) = e$ und $sup(C) = max(C) = h$.

2.5 Funktionen

Definition: Unter einer *Funktion* (oder *Abbildung*) f von einer Menge A in eine Menge B versteht man eine Zuordnung, bei der jedem Element aus A ein eindeutig bestimmtes Element aus B entspricht. Formal kann f als eine Relation zwischen A und B charakterisiert werden, so dass für jedes $a \in A$ genau ein $b \in B$ existiert mit $a\, f\, b$. Als übliche Schreibweise verwenden wir $f : A \longrightarrow B$ um auszudrücken,

dass f eine Funktion von A nach B ist, und $f(a) = b$, um auszudrücken, dass dem Element a durch die Funktion f der Wert b zugeordnet wird. Die Menge A wird *Definitionsbereich* von f und die Menge B wird *Wertebereich* oder *Wertevorrat* von f genannt.

Anmerkung: Die hier verwendete Definition von Funktionen ist die in der Mathematik übliche Form. In der Informatik spielen oft auch sogenannte partielle Funktionen eine Rolle. Das sind Relationen zwischen A und B, so dass jedes $a \in A$ zu höchstens einem $b \in B$ in Beziehung steht. Im Sinne der oben gegebenen Definition ist also eine partielle Funktion nicht ein Spezialfall einer Funktion, sondern im Allgemeinen gar keine Funktion, aber eine Funktion ist auch immer eine partielle Funktion.

Definition: Ist $f : A \longrightarrow B$ eine Funktion, $M \subseteq A$ und $N \subseteq B$, dann nennt man die Menge
$f(M) = \{y \in B \mid \text{es gibt ein } x \in M \text{ mit } f(x) = y\}$ das *Bild* von M unter f
und die Menge
$f^{-1}(N) = \{x \in A \mid f(x) \in N\}$ das *vollständige Urbild* von N unter f.

Definition:

- Eine Funktion $f : A \longrightarrow B$ heißt *surjektiv*, falls jedes Element von B im Bild von A auftritt, d.h. $f(A) = B$. Man kann sagt dann auch, dass f die Menge A *auf* die Menge B abbildet.
- Eine Funktion $f : A \longrightarrow B$ heißt *injektiv* oder *eineindeutig*, falls je zwei verschiedene Elemente aus A auch verschiedene Bilder haben, d.h. wenn aus $f(a) = f(a')$ die Gleichheit von a und a' folgt.
- Eine Funktion wird *bijektiv* genannt, wenn sie injektiv und surjektiv ist.

Beispiel: Wir betrachten die bekannte Sinusfunktion. Als Funktion von den reellen Zahlen in die reellen Zahlen ist $sin : \mathbb{R} \longrightarrow \mathbb{R}$ weder injektiv noch surjektiv.
Durch Einschränkungen von Definitions– und/oder Wertebereich kann man diese Eigenschaften erzwingen:

- $sin : \mathbb{R} \longrightarrow [-1, 1]$ ist surjektiv, aber nicht injektiv
- $sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \longrightarrow \mathbb{R}$ ist injektiv, aber nicht surjektiv
- $sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \longrightarrow [-1, 1]$ ist bijektiv.

Betrachtet man eine Funktion $f : A \longrightarrow B$ als Relation, dann ist die zu f inverse Relation f^{-1} genau dann eine Funktion, wenn f bijektiv ist. In diesem Fall wird f^{-1} die zu f *inverse Funktion* genannt.

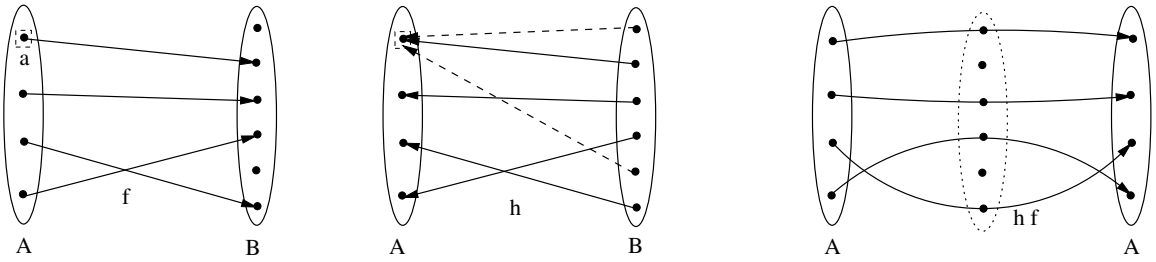
Definition: Sind $f : A \longrightarrow B$ und $g : B \longrightarrow C$ Funktionen, dann ist die Relationsverkettung $f \circ g$ eine Funktion von A in C. Sie wird *Verknüpfung* oder *Komposition* von f und g genannt und durch $gf : A \longrightarrow C$ bezeichnet, wobei $gf(a) = g(f(a))$

gilt. Man beachte, dass Relationsverkettungen von links nach rechts und Funktionsverknüpfungen von rechts nach links geschrieben werden.

Satz: Die folgenden Fakten ergeben sich als einfache Schlussfolgerungen aus den Definitionen. Seien $f : A \longrightarrow B$ und $g : B \longrightarrow C$ zwei Funktionen, dann gilt:

- Ist f bijektiv, dann ist $f^{-1}f = Id_A$ und $ff^{-1} = Id_B$
- f ist genau dann injektiv, wenn eine Funktion $h : B \longrightarrow A$ existiert mit $hf = Id_A$.
- f ist genau dann surjektiv, wenn eine Funktion $h : B \longrightarrow A$ existiert mit $fh = Id_B$.
- Sind f und g injektiv, dann ist auch gf injektiv.
- Sind f und g surjektiv, dann ist auch gf surjektiv.
- Sind f und g bijektiv, dann ist auch gf bijektiv und es gilt $(gf)^{-1} = f^{-1}g^{-1}$.

Beispiel 1: Die folgende Abbildung illustriert den zweiten Fakt. Die links dargestellte Funktion $f : A \longrightarrow B$ ist injektiv. Zur Konstruktion der Funktion $h : B \longrightarrow A$ bilden wir alle Elemente b aus dem Bild $f(A)$ auf das eindeutige Element aus dem Urbild $f^{-1}(\{b\})$ ab (die Eindeutigkeit folgt aus der Injektivität von f, in der graphischen Darstellung bedeutet das einfach die Umkehrung der Pfeile). Zusätzlich fixieren wir ein beliebiges Element $a \in A$ und bilden alle Elemente aus $B \setminus f(A)$ auf a ab (gestrichelte Pfeile in der Mitte). Die Komposition hf ist dann die Identische Abbildung auf A (rechts).



Beispiel 2: Bezeichne $\mathbb{R}$ die Menge der reellen Zahlen und $\mathbb{R}^{\geq 0}$ die Menge der nicht negativen reellen Zahlen. Dann ist die Funktion $f(x) = x^2$ eine surjektive Funktion von $\mathbb{R}$ auf $\mathbb{R}^{\geq 0}$. Wir suchen eine passende Funktion $h : \mathbb{R}^{\geq 0} \longrightarrow \mathbb{R}$, so dass $fh = id_{\mathbb{R}^{\geq 0}}$, d.h. h muss jedes $y \in \mathbb{R}^{\geq 0}$ auf eine Zahl x abbilden, deren Quadrat wieder y ist. Diese Eigenschaft wird offensichtlich durch die Quadratwurzelfunktion erfüllt, d.h. $h(y) = \sqrt{y}$ wäre eine geeignete Wahl, aber auch die Funktion $h(y) = -\sqrt{y}$ hat die geforderte Eigenschaft.

Satz: Jede Funktion $f : A \longrightarrow B$ induziert eine Äquivalenzrelation $\sim_f$ über A durch $a \sim_f b$ genau dann, wenn $f(a) = f(b)$.

Diese Äquivalenzrelation wird auch die *Faserung* von A durch f genannt.

2.6 Natürliche Zahlen und vollständige Induktion

Alle aus der Schulmathematik bekannten Aussagen über natürliche Zahlen können aus einigen wenigen Grundannahmen abgeleitet werden. Diese grundlegenden Voraussetzungen für die Existenz der natürlichen Zahlen (und letzlich Grundlagen der gesamten Zahlentheorie) gehen auf Richard Dedekind und Giuseppe Peano zurück und sind als die Peano'schen Axiome bekannt:

1. 0 ist eine natürliche Zahl.
2. Jede natürliche Zahl n hat einen eindeutigen Nachfolger $S(n)$, der auch eine natürliche Zahl ist.
3. Aus $S(n) = S(m)$ folgt $n = m$.
4. 0 ist kein Nachfolger einer natürlichen Zahl.
5. Jede Menge X, die 0 enthält und für die gilt, dass aus $n \in X$ auch $S(n) \in X$ folgt, enthält alle natürlichen Zahlen.

Achtung: Wir schreiben für den Nachfolger $S(n)$ auch $n + 1$, aber das ist als symbolische Schreibweise und nicht als Anwendung der Operation Addition zu verstehen. Im Gegenteil, wie die folgenden Betrachtungen zeigen, kann die Addition durch Anwendung der Nachfolgerfunktion rekursiv definiert werden.

Konsequenz 1: Man kann Funktionen $f : \mathbb{N} \longrightarrow A$ definieren, indem man $f(0)$ festlegt und $f(S(n))$ auf $f(n)$ zurückführt. Dieses Prinzip der Definition von Funktionen nennt man *Rekursion*.

Beispiel: Um die Addition von natürlichen Zahlen einzuführen, definieren wir für jede fest gewählte Zahl m die Funktion $m+ : \mathbb{N} \longrightarrow \mathbb{N}$, die jedem n aus dem Definitionsbereich die Summe $m + n$ zuordnen soll. Diese Funktion hat die folgende rekursive Definition: $m + (0) := m$ und $m + (S(n)) := S(m + n)$. Das entspricht den Regeln $m + 0 := m$ und $m + (n + 1) := (m + n) + 1$.

Analog kann man die Multiplikation durch $m\cdot : \mathbb{N} \longrightarrow \mathbb{N}$ mit $m \cdot (0) := 0$ und $m \cdot (S(n)) := (m \cdot (n)) + m$ definieren, was den Regeln $m \cdot 0 := 0$ und $m \cdot (n + 1) := (m \cdot n) + m$ entspricht.

Konsequenz 2: Man kann allgemeine Aussagen über natürliche Zahlen nach dem folgenden Schema beweisen. Eine Aussageform $P(x)$ über dem Bereich der natürlichen Zahlen ist wahr für alle natürlichen Zahlen, wenn sie die folgenden zwei Bedingungen erfüllt:

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

Dieses Beweisprinzip nennt man *vollständige Induktion*. Die erste Bedingung wird *Induktionsanfang* oder *Induktionsbasis*, die zweite Bedingung *Induktionsschluss* genannt. Dabei heißt $P(n)$ *Induktionsvoraussetzung* oder *Induktionsannahme* und $P(n + 1)$

Induktionsbehauptung.

Beispiele für Aussagen, die man mit Induktion beweisen kann:

- Für jede natürliche Zahl n ist die Zahl $a_n = n^3 + 2n$ durch 3 teilbar.
- Für beliebige reelle Zahlen a und $r \neq 1$ und für jede natürliche Zahl n gilt

$$\sum_{i=0}^{n} ar^i = \frac{ar^{n+1} - a}{r - 1}.$$

Exemplarisch für das zu verwendende Schema stellen wir hier den Beweis der ersten Aussage in einer sehr ausführlichen Version vor.

Induktionsanfang: Für $n = 0$ ist $a_n = 0^3 + 2 \cdot 0 = 0$ durch 3 teilbar (nach Teilbarkeitsdefinition: $a_n = 0 = 3 \cdot 0$).

Induktionsvoraussetzung: $a_n = n^3 + 2n$ ist durch 3 teilbar (für ein bestimmtes $n \in \mathbb{N}$), d.h. $a_n = 3k$ für ein $k \in \mathbb{N}$

Induktionsbehauptung: $a_{n+1} = (n+1)^3 + 2(n+1)$ ist durch 3 teilbar.

Induktionsschritt:

a_{n+1}	$= (n+1)^3 + 2(n+1)$	*Binomische Formel anwenden*
	$= (n^3 + 3n^2 + 3n + 1) + (2n + 2)$	*geeignet zusammenfassen*
	$= (n^3 + 2n) + 3n^2 + 3n + 3$	
	$= a_n + 3n^2 + 3n + 3$	*Induktionsvoraussetzung anwenden*
	$= 3k + 3n^2 + 3n + 3$	3 *ausklammern* (*Distributivgesetz*)
	$= 3(k + n^2 + n + 1)$	$k' = k + n^2 + n + 1$
	$= 3k'$	$k' \in \mathbb{N}$

Folglich ist auch a_{n+1} durch 3 teilbar und somit die Induktionsbehauptung bewiesen. □

Für mit dem Beweisschema vertraute Leser kann man diesen Induktionbeweis auch in einer verkürzten Form aufschreiben. Wir verwenden die Kürzel IA, IV, IB und IS für Induktionsanfang, Induktionsvoraussetzung, Induktionsbehauptung und Induktionsschritt. Da Induktionsvoraussetzung und Induktionsbehauptung sich im Allgemeinen schon aus der Formulierung der Aussage ablesen lassen, kann man darauf verzichten, sie noch einmal explizit aufzuschreiben. An Stelle dessen vermerkt man beim Induktionsschritt, ob sich die Voraussetzung auf n und die Behauptung auf $n + 1$ bezieht oder ob man von $n - 1$ auf n schließen will (was manchmal der bequemere Weg sein kann). Hier ist eine Kurzversion des letzten Beweises:

IA: Für $n = 0$ ist $a_0 = 0$ durch 3 teilbar.

IS: $n \longrightarrow n + 1$

$a_{n+1} = (n+1)^3 + 2(n+1) = n^3 + 3n^2 + 3n + 1 + 2n + 2 = a_n + 3(n^2 + n + 1)$

a_{n+1} ist durch 3 teilbar, weil a_n nach IV und der zweite Summand nach Definition durch 3 teilbar ist. □

Zwei Varianten des Induktionsprinzips werden häufig verwendet:

Variante 1: Wird die Induktionsbasis nicht für $n = 0$ sondern für einen anderen festen Anfangswert $k > 0$ bewiesen, so gilt die Aussage für alle natürlichen Zahlen $n \geq k$.

Beispiele:

- Für jede natürliche Zahl $n > 0$ ist die Summe der ungeraden Zahlen von 1 bis $2n - 1$ gleich n^2.
- Jeden ganzzahligen Wert $n \geq 8$ kann man durch Briefmarken mit den Werten 3 und 5 zusammenstellen.

Variante 2: Beim Induktionsschritt ist es erlaubt, nicht nur auf $P(n)$, sondern auf beliebige kleinere Zahlen zurückzugreifen, d.h. an Stelle von $P(n) \longrightarrow P(n+1)$ zeigt man $P(k) \wedge P(k+1) \wedge \ldots \wedge P(n) \longrightarrow P(n+1)$, wobei k der Anfangswert aus der Induktionsbasis ist. Dieses Prinzip wird *verallgemeinerte vollständige Induktion* genannt.

Der folgende Satz gibt ein typisches Beispiel für eine Aussage, die man mit verallgemeinerter Induktion beweisen kann.

Satz: Jede natürliche Zahl $n \geq 2$ kann man als Produkt von Primzahlen darstellen.

2.7 Endliche Mengen und Kardinalzahlen

Definition: Zwei Mengen A und B werden *gleichmächtig* genannt, wenn eine bijektive Funktion $f : A \longrightarrow B$ existiert.
In diesem Sinne bedeutet eine Abzählung einer endlichen (n-elementigen) Menge M, dass eine Bijektion zwischen der Menge $\{1, 2, \ldots, n\}$ und M (oder zwischen $\{0, 1, \ldots, n-1\}$ und M) konstruiert wird. Problematisch ist dabei, dass die Beschreibung von $\{1, 2, \ldots, n\}$ (oder $\{0, 1, \ldots, n-1\}$) genau gesehen eine rekursive Definition ist. John von Neumann löste dieses Problem durch eine spezielle Konstruktion der natürlichen Zahlen.
Definition: Die natürlichen Zahlen in Sinne von *von Neumann* bilden die kleinste Menge mit den folgenden zwei Eigenschaften:

1) $0 = \emptyset \in \mathbb{N}$
2) Ist $n \in \mathbb{N}$, dann ist auch der Nachfolger $S(n) := n \cup \{n\}$ in $\mathbb{N}$.

Damit erhalten wir die folgende Darstellung der natürlichen Zahlen:
$0 = \emptyset$,
$1 = \{\emptyset\} = \{0\}$,
$2 = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$,
$3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\}$
. . . . . .

Jede natürliche Zahl ist nach dieser Konstruktion durch die Anzahl ihrer Elemente bestimmt.

Definition: Eine Menge A wird *endlich* genannt, wenn es eine von Neumannsche natürliche Zahl n gibt, so dass A und n gleichmächtig sind. In diesem Fall wird n die *Kardinalzahl* von A genannt (Schreibweise $|A| = n$). Zwei endliche Mengen sind also gleichmächtig, wenn die Anzahlen ihrer Elemente gleich sind.

Satz: Sind A, B und M endliche Mengen, dann gelten die folgenden Aussagen über ihre Mächtigkeiten:

1. $|A \setminus B| = |A| - |A \cap B|$
2. $|A \cup B| = |A| + |B| - |A \cap B|$
3. $|A \times B| = |A| \cdot |B|$
4. $|\mathcal{P}(M)| = 2^{|M|}$

Anmerkungen zu diesen Punkten 1) Die Summenregel ist eine Grundregel der Kombinatorik, die besagt, dass die Größe der Vereinigung von zwei disjunkten endlichen Mengen A und B sich als Summe aus $|A|$ und $|B|$ ergibt. Diese Regel ist als Spezialfall in der ersten und zweiten Aussage enthalten.

2) Die Verallgemeinerung der zweiten Aussage für drei und mehr Mengen führt zum sogenannten *Prinzip der Inklusion–Exklusion*:

$$|A \cup B \cup C| = \quad |A| + |B| + |C| - (|A \cap B| + |A \cap C| + |B \cap C|) + |A \cap B \cap C|$$

$$|\bigcup_{i=1}^{k} A_i| = \quad \sum_{1 \le i \le k} |A_i| - (\sum_{1 \le i_1 < i_2 \le k} |A_{i_1} \cap A_{i_2}|) + (\sum_{1 \le i_1 < i_2 < i_3 \le k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}|) - \ldots + \ldots \quad + (-1)^{k+1} |\bigcap_{i=1}^{k} A_i|$$

3) Die dritte Aussage ist auch eine Grundregel der Kombinatorik und wird Produktregel genannt.

4) Wie man leicht sieht, gibt es eine bijektive Abbildung von $\mathcal{P}(M)$ auf die Menge der Funktionen von M nach $\{0, 1\}$. Dabei wird jedem $A \in \mathcal{P}(M)$ die charakteristische Funktion $\chi_A : M \longrightarrow \{0, 1\}$ zugeordnet, welche jedes $x \in A$ auf 1 abbildet und für alle $x \notin A$ den Wert 0 hat. Bezeichnet man für zwei Mengen N und M mit N^M die Menge aller Funktionen von M nach N, dann kann die oben beschriebene Bijektion auch wie folgt gelesen werden:
$\mathcal{P}(M)$ ist gleichmächtig mit $\{0, 1\}^M$ (also mit 2^M, da $2 = \{0, 1\}$ im Sinne der von Neumannschen natürlichen Zahlen).

Aus dieser Sicht kann man die vierte Aussage wie folgt verallgemeinern: Sind N und M zwei beliebige endliche Mengen, dann ist die Anzahl der Funktionen von M in N gleich $|N|^{|M|}$, also $|N^M| = |N|^{|M|}$.

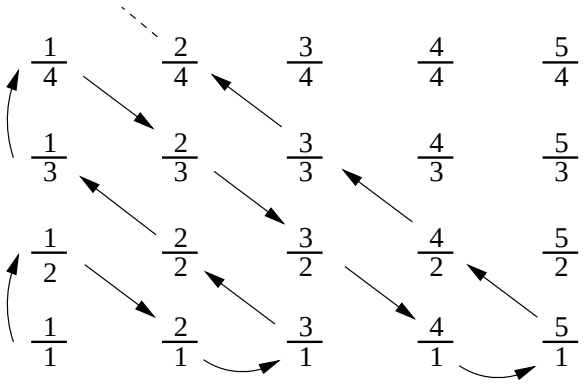
Definition: Eine Menge A, die nicht endlich ist, wird *unendlich* genannt (Schreibweise $|A| = \infty$).
Eine Menge A wird *abzählbar* genannt, wenn sie endlich ist oder wenn sie mit der Menge der natürlichen Zahlen $\mathbb{N}$ gleichmächtig ist.

Satz: Die Menge $\mathbb{Z}$ der ganzen Zahlen und die Menge $\mathbb{Q}$ der rationalen Zahlen sind abzählbar.

Beweisidee: Für $\mathbb{Z}$ reicht es aus, eine unendliche Folge $z_0, z_1, z_2, \ldots$ zu bilden in der jede ganze Zahl genau einmal auftritt, denn dann wird durch $f(n) := z_n$ eine Bijektion zwischen $\mathbb{N}$ und $\mathbb{Z}$ definiert. Eine solche Folge ist z.B. $0, 1, -1, 2, -2, 3, -3, \ldots$. In diesem Fall ist es auch leicht, eine geschlossene Formel für die Bijektion $f : \mathbb{N} \longrightarrow \mathbb{Z}$ anzugeben:

$$f(n) = (-1)^{n+1} \cdot \left\lceil \frac{n}{2} \right\rceil = \begin{cases} -\frac{n}{2} & \text{falls } n \text{ gerade} \\ \frac{n+1}{2} & \text{falls } n \text{ ungerade} \end{cases}$$

Für $\mathbb{Q}$ kombiniert man diesen Trick der alternierenden Aufzählung mit einem Schema zur Aufzählung aller positiven rationalen Zahlen. Dazu schreibt man alle Brüche mit positiven Zählern und Nennern in eine Tabelle und läuft die Zellen systematisch nach folgendem Schema ab. Daraus ergibt sich eine surjektive Abbildung $g : \mathbb{N} \longrightarrow \mathbb{Q}^+$.



Um nicht nur eine surjektive Abbildung, sondern eine Bijektion zu erhalten, muss man letzlich noch dafür sorgen, dass alle Brüche übersprungen werden, deren rationaler Wert schon vorher erreicht wurde (das sind alle Brüche, die sich kürzen lassen). Im abgebildeten Schema ist der Bruch $\frac{2}{2}$ der erste, den man überspringen muss, danach die Brüche $\frac{4}{2}$, $\frac{3}{3}$ und $\frac{2}{4}$. Da es in diesem Fall wesentlich schwerer ist, eine geschlossene Formel für g anzugeben, begnügen wir uns hier mit der verbalen Konstruktionsbeschreibung einer solchen Bijektion.

Satz: Die Menge der reellen Zahlen ist nicht abzählbar.

Beweis: Wir führen einen Widerspruchsbeweis, um zu zeigen, dass schon die reellen Zahlen aus dem Intervall $[0, 1]$ nicht abzählbar sind. Dazu nehmen wir an, dass es eine bijektive Abbildung $f : \mathbb{N} \longrightarrow [0, 1]$ gibt und schreiben die Bilder $f(0), f(1), f(2), \ldots$

als Dezimalbrüche

$$\begin{array}{lcll} f(0) & = & 0, a_{0,0}\, a_{0,1}\, a_{0,2}\, a_{0,3} \ldots & \\ f(1) & = & 0, a_{1,0}\, a_{1,1}\, a_{1,2}\, a_{1,3} \ldots & a_{i,j} \in \{0, 1, 2, \ldots, 9\} \\ f(2) & = & 0, a_{2,0}\, a_{2,1}\, a_{2,2}\, a_{2,3} \ldots & \\ \vdots & \vdots & \vdots & \end{array}$$

Es ist bekannt, dass Dezimalbruch–Darstellungen von reellen Zahlen immer dann eindeutig sind, wenn sie nicht mit $\overline{9}$ (d.h. 9–Periode) enden und auch nicht abbrechen (d.h. 0–Periode).
Jetzt ist es leicht, eine Zahl $r \in [0, 1]$ mit der Darstellung $0, b_0 b_1 b_2 \ldots$ zu konstruieren, die nicht im Bild von f liegt, indem man die einzelnen Dezimalstellen wie folgt festlegt

$$b_i = \begin{cases} 1 & \text{falls } a_{i,i} \neq 1 \\ 2 & \text{falls } a_{i,i} = 1 \end{cases}$$

Offensichtlich hat die Zahl r eine eindeutige Darstellung und liegt nicht im Bild von f, weil sie sich von jedem $f(n)$ aus dem Bild an der n–ten Stelle unterscheidet. Folglich ist die Abbildung f nicht surjektiv, ein Widerspruch zur Annahme, dass f eine Bijektion ist.

Die hier verwendete Beweistechnik bezeichnet man als Diagonalisierung. In der theoretischen Informatik wird sie in abgewandelter Form für Unmöglichkeitsbeweise eingesetzt (z.B. um zu zeigen, dass bestimmte Funktionen nicht berechenbar sind). Der folgende Satz beinhaltet eine weitere Anwendung dieser Technik.

Satz: Es gibt keine surjektive Abbildung von einer Menge M auf ihre Potenzmenge.

Beweis durch Widerspruch: Angenommen M wäre eine Menge mit einer surjektiven Abbildung $f : M \longrightarrow \mathcal{P}(M)$. Wir bilden die Menge $A = \{x \in M \mid x \notin f(x)\}$. Da A Untermenge von M, also Element von $\mathcal{P}(M)$ ist, muss es ein $a \in M$ mit $f(a) = A$ geben, anderenfalls wäre f nicht surjektiv. Der Widerspruch besteht darin, dass aus der Definition von A die logische Äquivalenz der Aussagen $a \in A$ und $a \notin A$ folgt.

3 Kombinatorik

3.1 Elementare Zählprinzipien

Die Abzählung von endlichen Mengen ist das klassische Thema der Kombinatorik. Dabei wird eine unendliche Familie $\{A_n \mid n \in I\}$ von endlichen Mengen A_n betrachtet, wobei n eine Indexmenge I durchläuft (in der Regel die natürlichen Zahlen). Zu bestimmen ist die *Zählfunktion* $f : I \longrightarrow \mathbb{N}$ mit $f(n) = |A_n|$ für alle $n \in I$. Hier und im Folgenden bezeichnet $|M|$ die Anzahl der Elemente (Mächtigkeit) einer Menge M, falls diese endlich ist und ∞ anderenfalls.

Oft werden die Mengen A_n aus einer gegebenen n–elementigen Menge M (auch kurz *n–Menge* genannt) durch einfache strukturelle oder kombinatorische Bedingungen abgeleitet, z.B. die Menge $S(M)$ aller Permutationen von M (das sind bijektive Abbildungen von M nach M), die Menge $\mathcal{P}(M)$ aller Untermengen oder die Menge $\binom{M}{k}$ aller k–Untermengen von M. Im letzten Fall ist die Zählfunktion von den zwei Größen n und k abhängig. Ziel ist es, eine möglichst kurze, geschlossene Formel für $f(n)$ bzw. $f(n,k)$ zu finden.
Die meisten Aufgabenstellungen dieser Art lassen sich (auch wenn die Lösung teilweise enormen technischen Aufwand erfordert) auf vier Grundregeln zurückführen.

- **Summenregel:** Ist S die Vereinigung von paarweise disjunkten, endlichen Mengen $S_1, S_2, \ldots, S_t$, dann gilt $|S| = \sum_{i=1}^{t} |S_i|$.
- **Produktregel:** Ist S das Kartesische Produkt von t endlichen Mengen $S_1, S_2, \ldots, S_t$, dann gilt $|S| = \prod_{i=1}^{t} |S_i|$.
- **Gleichheitsregel:** Existiert eine bijektive Abbildung zwischen zwei Mengen S und T, so gilt $|S| = |T|$.
- **Die Regel vom doppelten Abzählen** wird etwas später erklärt.

Die Gleichheitsregel ist letzlich nur eine Umformulierung der Definition für die Gleichmächtigkeit von zwei Mengen. Die Summen- und Produktregel kann man mit vollständiger Induktion beweisen.

Eine typische Anwendung der Summenregel besteht darin, die Elemente der Menge S nach gewissen, sich gegenseitig ausschließenden Eigenschaften E_i $(i = 1, \ldots, t)$ zu klassifizieren und die Mengen S_i wie folgt zu setzen $\{x \in S \mid x\ hat\ Eigenschaft\ E_i\}$.

Beispiel: Wir haben bereits eine Plausibilitätserklärung gefunden, warum die Potenzmenge einer n–Menge M genau 2^n Elemente hat. Hier ist ein kompletter Beweis mit vollständiger Induktion nach n.

Induktionsanfang: Für $n = 0$ ist $M = \emptyset$, $\mathcal{P}(M) = \{\emptyset\}$ und folglich $|\mathcal{P}(M)| = 1 = 2^0$.
Induktionsschritt von n nach $n+1$: Sei M eine $(n+1)$–Menge und a ein fest gewähltes Element aus M. Wir zerlegen $\mathcal{P}(M)$ in zwei disjunkte Teilmengen $S_1 = \{A \subseteq M \mid a \in A\}$ und $S_2 = \{A \subseteq M \mid a \notin A\}$. Offensichtlich ist $S_2 = \mathcal{P}(M \setminus \{a\})$ die Potenzmenge einer n–Menge und nach Induktionsvoraussetzung gilt $|S_2| = 2^n$.

Wir konstruieren eine Bijektion zwischen S_1 und S_2, indem wir jeder Menge $A \in S_1$ die Menge $B = A \setminus \{a\} \in S_2$ zuordnen. Man erkennt die Bijektivität dieser Abbildung daran, dass eine Umkehrabbildung existiert, die jedem $B \in S_2$ die Menge $A = B \cup \{a\} \in S_1$ zuordnet. Aus der Gleichheitsregel folgt $|S_1| = |S_2|$ und aus der Summenregel $|\mathcal{P}(M)| = |S_1| + |S_2| = 2|S_2| = 2 \cdot 2^n = 2^{n+1}$. □

Man kann die gleichen Argumente verwenden, um eine Rekursion für die Anzahl der k-Teilmengen einer n-Menge zu entwickeln. Sei $S = \binom{M}{k}$ die Menge aller k–Untermengen einer n–Menge M und bezeichne $\binom{n}{k}$ die Kardinalzahl von S (zunächst nur als ein formaler Ausdruck). Wir fixieren ein $a \in M$ und klassifizieren die k–Untermengen von M nach der Eigenschaft, ob sie a enthalten (E_1) oder nicht (E_2). Also ist $S_1 = \{A \in S \mid a \in A\}$ und $S_2 = \{A \in S \mid a \notin A\}$. Wie man leicht sieht, ist $S_2 = \binom{M\setminus\{a\}}{k}$ und auf der anderen Seite gibt es eine Bijektion zwischen S_1 und der Menge aller $(k-1)$–Untermengen von $M \setminus \{a\}$. Durch Anwendung von Gleichheits– und Summenregel erhalten wir die folgende rekursive Formel

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Definitition: Ein *Inzidenzsystem* (S, T, I) besteht aus zwei (endlichen) Mengen S und T und einer *Inzidenzrelation* $I \subseteq S \times T$.

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 $a \in S$, $b \in T$, dann gilt:

$$\sum_{a \in S} r(a) = \sum_{b \in T} r(b).$$

In der Tat sind beide Summen gleich $|I|$, denn die Paare der Relation I wurden nur auf zwei verschiedene Weisen aufgezählt. Oft wird diese Regel auch in etwas modifizierter Form unter Verwendung von Inzidenzmatrizen oder bipartiten Graphen verwendet:

Sei $M = (m_{ij})$ eine $m \times n$ Matrix, d.h. ein Rechteck–Schema mit m Zeilen und n Spalten in dem Zahlen m_{ij} (i-te Zeile, j-te Spalte) eingetragen sind. Sind $z_i = \sum_{j=1}^{n} m_{ij}$ $(1 \le i \le m)$ die Zeilensummen, und $s_j = \sum_{i=1}^{m} m_{ij}$ $(1 \le j \le m)$ die Spaltensummen, dann gilt $\sum_{i=1}^{m} z_i = \sum_{j=1}^{n} s_j$, denn beide Summen sind gleich der Summe aller Matrixelemente. Die Regel vom zweifachen Abzählen ergibt sich nun als Spezialfall für die sogenannte *Inzidenzmatrix*. Dazu werden die Elemente aus S und T nummeriert, $S = \{a_1, \ldots, a_m\}$, $T = \{b_1, \ldots, b_n\}$ und $m_{ij} = \begin{cases} 1 & falls\ (a_i, b_j) \in I \\ 0 & sonst \end{cases}$ gesetzt.

3.2 Die fundamentalen Zählkoeffizienten

k-Kombinationen und k-Variationen

Sei eine n–Menge N gegeben (ohne Einschränkung kann man $N = \{1, 2, \ldots, n\}$ annehmen) und $0 \le k \le n$ eine natürliche Zahl. In diesem Abschnitt werden wir

die Werte einiger Zählfunktionen bestimmen, die sehr häufig auftreten und deshalb mit eigenen Symbolen bezeichnet werden. Eine k–elementige Untermenge von N wird auch *k–Kombination* genannt. Die Anzahl der k–Kombinationen einer n–Menge wird mit $\binom{n}{k}$ bezeichnet, und diese Anzahlen werden *Binomialkoeffizienten* genannt. Dabei werden k–Kombinationen (wie für Mengen üblich) als nicht geordnet betrachtet, d.h. $\{1,3,5\}$ und $\{3,5,1\}$ sind ein und dieselbe 3–Kombination. "Geordnete" k–Untermengen werden auch *k–Variationen (ohne Wiederholung)* genannt, alternativ kann dafür auch der Begriff *k–Permutation einer n–Menge* verwendet werden. Achtung: Geordnet heißt hier gerade nicht nach Größe geordnet, sondern in einer bestimmten Reihenfolge angeordnet! Man kann sie als Wörter der Länge k aus voneinander verschiedenen Elementen aus N kodieren (also $135 \neq 351$).

k-Partitionen

Der Begriff der Partition wurde bereits eingeführt und im Zusammenhang mit Äquivalenzrelationen verwendet. Wir verstehen darunter die Zerlegung einer Menge in disjunkte, nichtleere Untermengen (Blöcke). Bezeichnet k die Anzahl der Blöcke, so sprechen wir von *k–Partitionen*. An dieser Stelle ist es wichtig, auf einen bisher eher nebensächlichen Aspekt einzugehen: Betrachtet man die Partion als **Menge** von k Blöcken, dann spielt die Reihenfolge der Blöcke keine Rolle und wir sprechen einfach von einer *k–Partition*, oder (wenn wir diesen Aspekt besonders betonen wollen) von einer *ungeordneten k–Partition*. Die Anzahl der k–Partitionen einer n–Menge wird mit $S_{n,k}$ bezeichnet. Die Zahlen $S_{n,k}$ werden *Stirling–Zahlen zweiter Art* genannt. Will man dagegen auch die Reihenfolge des Auftretens der Blöcke berücksichtigen, so spricht man von *geordneten k–Partitionen*. Zur Unterscheidung umfasst man die Blöcke einer ungeordneten Partition in Mengenklammern, während geordnete Partitionen als Tupel geschrieben werden (runde Klammern).

Beispiel: Die Mengen $\{\{2\},\{1,3,5\},\{4,6\}\}$ und $\{\{3,5,1\},\{2\},\{4,6\}\}$ sind gleich, d.h. sie stellen ein und dieselbe (ungeordnete) 3–Partition dar.
Dagegen sind $(\{2\},\{1,3,5\},\{4,6\})$ und $(\{1,3,5\},\{4,6\},\{2\})$ zwei verschiedene geordnete 3–Partitionen.

Binomialkoeffizienten und fallende Faktorielle

Die Anzahl der k–Variationen von N kann mit der Produktregel bestimmt werden (n Möglichkeiten für die erste Stelle und ist diese festgelegt, dann $n-1$ Möglichkeiten für die zweite Stelle,...):

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

Diese Anzahlen werden auch kurz mit $n^{\underline{k}}$ bezeichnet und *fallende Faktorielle* genannt. Streng genommen muss ein solcher Beweis auch wieder mit vollständiger Induktion geführt werden, aber inzwischen sollte jeder genug Übung darin haben, dieses Detail auszufüllen. Analog definiert man die *steigenden Faktoriellen* mit

$$n^{\overline{k}} = n(n+1)\dots(n+k-1) = \frac{(n+k)!}{n!},$$

wobei wir $n^{\overline{0}} = n^{\underline{0}} = 1$ setzen.

Da jede k–Untermenge sich auf $k!$ Arten als k–Permutation darstellen lässt, gilt nach Produktregel $n^{\underline{k}} = k!\binom{n}{k}$. Daraus kann eine explizite Formel für die Binomialkoeffizienten abgeleitet werden:

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n!}{(n-k)!k!}$$

Analog kann auch die Anzahl der geordneten k–Partitionen durch das Produkt $k!\, S_{n,k}$ beschrieben werden.

Beispiel: Der Ziehungsvorgang beim Lotto "6 aus 49" liefert zunächst eine 6–Variation der Menge $\{1, 2, \ldots, 49\}$ (wir vernachlässigen hier die Zusatzzahl). Die Anzahl dieser 6–Variationen ist $49^{\underline{6}}$. Für das Ergebnis der Ziehung ist nur die Menge der gezogenen Zahlen entscheidend und nicht die Reihenfolge, in der sie gezogen wurden. Da jedes Ziehungsergebnis in 6! verschiedenen Reihenfolgen gezogen werden kann, ist die Anzahl der möglichen Ziehungsergebnisse (und damit auch die Anzahl der möglichen Tipps) gleich $\frac{49^{\underline{6}}}{6!} = \binom{49}{6}$.

Geordnete und ungeordnete Zahlpartitionen

Eine Abwandlung des Begriffs der Mengenpartition führt zu den sogenannten Zahlpartitionen. Ist $n \in \mathbb{N}$ dann wird eine Summe $n = n_1 + \ldots + n_k$ mit $n_i \in \mathbb{N}, n_i > 0$ eine k–*Zahlpartition* von n genannt. Die Anzahl solcher Partitionen wird mit $P_{n,k}$ bezeichnet. Auch hier spielt die Reihenfolge der n_i **keine** Rolle, so dass man ohne Einschränkung $n_1 \geq n_2 \geq \ldots \geq n_k$ annehmen kann. Soll die Reihenfolge Berücksichtigung finden, so sprechen wir von *geordneten Zahlpartitionen*.

Achtung: Geordnete Zahlpartitionen sind also gerade solche, in denen die Summanden in der Regel **nicht** nach ihrer Größe sortiert auftreten. Dagegen kann man ungeordnete Zahlpartitionen (in denen die Reihenfolge egal ist) immer so darstellen, dass die Summanden aufsteigend geordnet sind.

Lemma: Die Anzahl der geordneten k–Zahlpartitionen von n ist $\binom{n-1}{k-1}$.

Beweis: Die Grundidee ist sehr einfach. Wir schreiben die Zahl n als Summe von n Einsen, also mit $n-1$ Pluszeichen. Für eine Zahlpartition $n = n_1 + n_2 + \ldots + n_k$ fassen wir die ersten n_1 Einsen zu einem Block zusammen, markieren das darauf folgende Pluszeichen, fassen die nächsten n_2 Einsen wieder zu einem Block zusammen und markieren das darauf folgende Pluszeichen usw. Bei k Summanden markiert man $k-1$ der $n-1$ Pluszeichen und jede solche Markierung beschreibt eindeutig die entsprechende Zahlpartition.

Für einen formalen Beweis konstruieren wir eine Bijektion von der Menge S der geordneten k–Zahlpartitionen auf die Menge T der $(k-1)$–Untermengen von $\{1, 2, \ldots, n-1\}$ durch:

$n = n_1 + n_2 + \ldots + n_k \ \in S \mapsto \{n_1, n_1 + n_2, \ldots, n_1 + n_2 + \ldots + n_{k-1}\} \in T$. Die Bijektivität dieser Abbildung kann durch die Konstruktion der Umkehrabbildung nachgewiesen werden, und die Behauptung folgt nun aus der Gleichheitsregel.

Stirlingzahlen und surjektive Funktionen

Die hier besprochenen Zählkoeffizienten sind auch sehr gut zum Abzählen von Funktionenmengen geeignet. Sei N eine n–Menge und R eine r–Menge. Aus der Mengenlehre wissen wir, dass die Anzahl der Abbildungen von N nach R gleich r^n ist.
Ist $n \leq r$, dann sind injektive Abbildungen von N in R durch n–Variationen von R charakterisiert, d.h. $|Inj(N,R)| = r^{\underline{n}}$.

Ist $n \geq r$, dann sind surjektive Abbildungen von N auf R durch geordnete r–Partitionen von N charakterisiert, d.h. $|Surj(N,R)| = r!\, S_{n,r}$.

Die Anzahl der bijektiven Abbildungen ist $n!$ (falls $n = r$) oder 0. Klassifizieren wir die Menge aller Abbildungen $f : N \longrightarrow R$ nach ihren Bildern $A = \{f(x) \mid x \in N\} \subseteq R$, so ergibt die Summenregel:

$$\begin{aligned} r^n = \quad |Abb(N,R)| \quad &= \textstyle\sum_{A \subseteq R} |Surj(N,A)| \\ &= \textstyle\sum_{k=0}^{r} \sum_{|A|=k} |Surj(N,A)| \\ &= \textstyle\sum_{k=0}^{r} \binom{r}{k} k!\, S_{n,k} \quad = \quad \sum_{k=0}^{r} r^{\underline{k}}\, S_{n,k} \end{aligned}$$

3.3 Zwölf Arten des Abzählens

Eine zusammenfassende Interpretation der Zählkoeffizienten erhalten wir durch die folgende Fragestellung:
Sei eine Menge von n Bällen gegeben, die in r Fächer verteilt werden sollen. Gesucht ist die Anzahl solcher Verteilungen allgemein und unter der zusätzlichen Bedingung, dass nur injektive (jedes Fach höchstens ein Ball), surjektive (jedes Fach mindestens ein Ball) bzw. bijektive (jedes Fach genau ein Ball) Verteilungen zu betrachten sind. Natürlich hängt die Problemstellung auch davon ab, ob die Bälle bzw. die Fächer unterscheidbar sind oder nicht. Offensichtlich korrespondieren die obigen Abzählungen von Abbildungen zu dem Fall, dass sowohl die Bälle als auch die Fächer unterscheidbar sind. Da die Frage der bijektiven Verteilungen relativ leicht zu beantworten sind, reduziert sich die Problemstellung auf zwölf Varianten.

Sind nur die Bälle unterscheidbar, aber nicht die Fächer, dann korrespondieren surjektive Verteilungen zu r–Partitionen einer n–Menge und die Menge aller Verteilungen kann mit der Summenregel über die Klassifizierung nach Anzahl der belegten Fächer abgezählt werden.

Sind Bälle und Fächer nicht unterscheidbar, dann korrespondieren surjektive Verteilungen zu r–Zahlpartitionen von n und auch hier kann die Menge aller Verteilungen mit der Summenregel über die Klassifizierung nach Anzahl der belegten Fächer abgezählt werden.

Sind nur die Fächer unterscheidbar, aber nicht die Bälle, dann korrespondieren die surjektiven Verteilungen zu geordneten r–Zahlpartitionen von n. Die injektiven Verteilungen sind durch die Auswahl der n belegten Fächer aus der r–Menge aller Fächer vollständig charakterisiert. Eine beliebige Abbildung ist durch eine geordnete

Summenzerlegung der Form $n = n_1 + \ldots + n_r$ mit $n_i \in \mathbb{N}$ charakterisiert. Um daraus eine r–Zahlpartition im Sinne der Definition zu machen (Zusatzbedingung $n_i \geq 1$), addiert man zu jedem Summanden eine 1. Damit erhalten wir eine Bijektion auf die Menge der geordneten r–Zahlpartitionen von $(n + r)$. Diese Menge hat $\binom{n+r-1}{r-1}$ Elemente.

Die folgende Tabelle gibt einen vollständigen Überblick.

	beliebig	injektiv	surjektiv	bijektiv
N unterscheidbar R unterscheidbar	r^n	$r^{\underline{n}}$	$r!S_{n,r}$	0 oder $n!$
N nicht untersch. R unterscheidbar	$\binom{n+r-1}{r-1}$	$\binom{r}{n}$	$\binom{n-1}{r-1}$	0 oder 1
N unterscheidbar R nicht untersch.	$\sum_{k=1}^{r} S_{n,k}$	0 oder 1	$S_{n,r}$	0 oder 1
N nicht untersch. R nicht untersch.	$\sum_{k=1}^{r} P_{n,k}$	0 oder 1	$P_{n,r}$	0 oder 1

3.4 Rekursionen

Pascalsches Dreieck

Oft ist es schwierig, für die zu untersuchenden kombinatorischen Größen $f(n)$ eine geschlossene Formel anzugeben, z.B. für die Stirling Zahlen zweiter Art $S_{n,k}$. In solchen Fällen kann eine Rekursion, das ist eine Formel, die $f(n)$ auf $f(n-1)$ und/oder $f(n-2), \ldots$ zurückführt, ein wirkungsvolles Hilfsmittel sein. Sind dann noch die benötigten Anfangswerte $f(0)$ und/oder $f(1)$ bekannt, kann man jedes $f(n)$ effektiv berechnen.

Wir haben bereits eine Rekursion für die Binomialkoeffizienten kennengelernt:

$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$ für alle $n, k > 0$.

Mit den trivialen Zusatzinformationen $\binom{n}{n} = \binom{n}{0} = 1$ und $\binom{n}{k} = 0$ falls $k > n$, können alle Binomalkoeffizienten berechnet werden. Das dafür verwendete Schema wird *Pascalsches Dreieck* genannt:

n \ k	0	1	2	3	4	5	6 …
0	1						
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
6	1	6	15	20	15	6	1
:							

Es ist klar, dass die Summe aller Einträge in der n–ten Zeile die Anzahl aller Unter-

mengen einer n–Menge ergibt:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Die folgenden zwei Identitäten machen Aussagen über Spalten- und Diagonalsummen, die an einer beliebigen Stelle abzubrechen sind:

$$\begin{aligned} \textstyle\sum_{m=0}^{n} \binom{m}{k} &= \binom{n+1}{k+1} && \text{für alle } n, k \geq 0 \\ \textstyle\sum_{k=0}^{n} \binom{m+k}{k} &= \binom{m+n+1}{n} && \text{für alle } m, n \geq 0 \end{aligned}$$

Beide Formeln kann man mit vollständiger Induktion und auch durch kombinatorische Argumentationen beweisen. Einige Identitäten dieser Art liefern auch wichtige Aussagen für Polynome. Dazu muss man einen oder mehrere Parameter der Zählkoeffizienten durch Polynomvariable (über den komplexen Zahlen) ersetzen. Setzt man $x^{\underline{k}} = x(x-1)\dots(x-k+1)$, dann kann man auch $\binom{x}{k} = \frac{x^{\underline{k}}}{k!}$ als Polynom vom Grad k ansehen. Wie uns bekannt ist, sind zwei Polynome vom Grad k genau dann gleich, wenn ihre Werte an mindestens $k+1$ Stellen übereinstimmen. Da die Rekursion der Binomialkoeffizienten auch für alle natürlichen Zahlen $n \geq k$ bewiesen ist, müssen auch die entsprechenden Polynome gleich sein, also:

$$\binom{x}{k} = \binom{x-1}{k} + \binom{x-1}{k-1} \ .$$

Auch die sogenannte *Vandermonde Identität* kann mit dieser Polynommethode bewiesen werden:

$$\binom{x+y}{n} = \textstyle\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k}$$

Wieder folgt die Gleichheit dieser Polynome aus dem Fakt, dass ihre Werte an unendlich vielen Stellen (nämlich für alle $x, y \in \mathbb{N}$) übereinstimmen. Der Fakt selbst kann mit der Summenregel begründet werden:
Seien $x, y \in \mathbb{N}$ und X, Y zwei disjunkte Mengen mit $|X| = x, |Y| = y$. Auf der linken Seite der Vandermonde Identität steht dann die Anzahl aller n–Kombinationen von $X \cup Y$. Wir klassifizieren solche n–Kombinationen nach der Anzahl k der Elemente aus X. Dann besteht die Klasse S_k aus allen n–Kombinationen mit k Elementen aus X und $n-k$ Elementen aus Y. Nach der Produktregel ist $|S_k| = \binom{x}{k}\binom{y}{n-k}$ und die Behauptung folgt aus der Summenregel.

Binomialsatz: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

Der induktive Beweis nutzt die Rekursion der Binomialkoeffizienten.

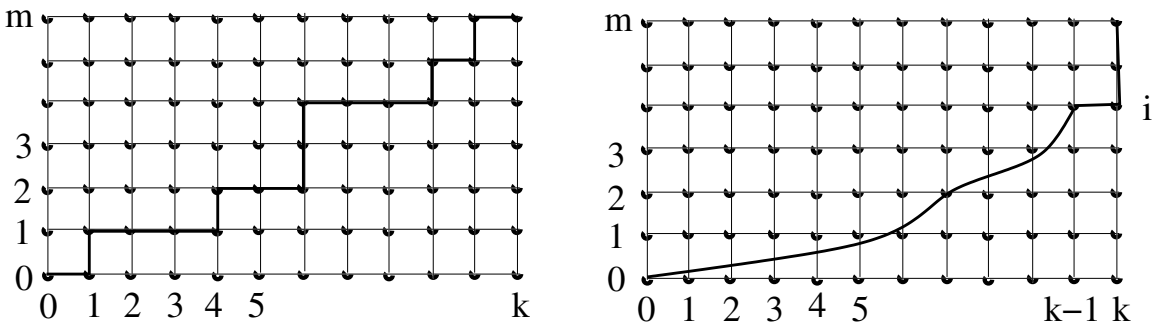
Als unmittelbare Folgerung aus dem Binomialsatz erhalten wir die folgenden zwei Formeln:

$$\begin{aligned} (x+1)^n &= \textstyle\sum_{k=0}^{n} \binom{n}{k} x^k \\ 2^n &= \textstyle\sum_{k=0}^{n} \binom{n}{k} \end{aligned}$$

Monotone Gitterwege

Eine weitere geometrisch-anschauliche Deutung der Binomialkoeffizienten ergibt sich aus der Abzählung von sogenannten Gitterwegen. Dazu betrachten wir die Punktmenge $G_{k,m} = \{0, 1, \ldots, k\} \times \{0, 1, \ldots, m\}$ in der Ebene und verbinden zwei Punkte durch eine Kante, wenn sie auf einer Koordinate übereinstimmen und sich auf der anderen Koordinate um 1 unterscheiden. Dieses Gebilde nennt man ein $k \times m$-Gitter. Unter einem *monotonen Gitterweg* im $k \times m$-Gitter verstehen wir eine Punktfolge, die mit $(0, 0)$ beginnt, mit (k, m) endet und bei der für zwei aufeinanderfolgende Punkte $(a, b), (c, d)$ gilt, dass entweder $c = a$ und $d = b + 1$ oder $c = a + 1$ und $d = b$. Anschaulich bedeutet das, dass ein monotoner Gitterweg von der Ecke links unten in die Ecke recht oben verläuft und jeder Schritt entweder nach rechts oder nach oben erfolgt (Abbildung links)). Der Zusammenhang zwischen monotonen Gitterwegen und Binomialkoeffizienten beruht auf einer simplen Beobachtung: Da sich jeder monotone Gitterweg in einem $k \times m$-Gitter aus $k + m$ Kanten zusammensetzt, von denen genau k horizontal (und die anderen m vertikal) verlaufen, kann man sie als k-Kombinationen aus einer $(k + m)$-Menge beschreiben.

Lemma: Die Anzahl der montonen Gitterwege in einem $k \times m$-Gitter ist $\binom{k+m}{k}$.



Viele Fakten über Binomialkoeffizienten lassen sich mit diesem Modell anschaulich erklären. Hier sind einige Beispiele dazu:

- Die Menge der monotonen Gitterwege im $k \times m$-Gitter kann man durch die Auswahl der k horizontalen Schritte aus $k+m$ Gesamtschritten oder alternativ durch die Auswahl von m vertikalen Schritten aus $k+m$ Gesamtschritten generieren. Daraus folgt $\binom{k+m}{k} = \binom{k+m}{m}$. Wenn wir $n = k+m$ setzen, ist $m = n-k$ und wir erhalten die bekannte Formel $\binom{n}{k} = \binom{n}{n-k}$.

- Partitioniert man die Menge der monotonen Gitterwege im $k \times m$-Gitter danach, ob der letzte Schritt horizontal oder vertikal gemacht wird, so ergibt sich die Formel $\binom{k+m}{k} = \binom{k+m-1}{k-1} + \binom{k-1+m}{k}$. Mit der oben verwendeten Ersetzung erhalten wir die bekannte Rekursion $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

- Wir partitionieren die Menge der monotonen Gitterwege im $k \times m$-Gitter danach, in welcher Höhe der letzte horizontale Schritt (von $k-1$ nach k) gemacht wird. Dafür kommen alle y-Werte zwischen 0 und m in Frage. Diese Situation ist in

der Abbildung auf der rechten Seite schematisch dargestellt: Der letzte horizontale Schritt erfolgt von $(k-1,i)$ nach (k,i) und darauf folgen nur noch vertikale Schritte bis (k,m). Die Anfangsstücke aller Wege dieser Klasse sind beliebige monotone Gitterwege im $(k-1)\times i$-Gitter, die im der Abbildung durch eine monotone Kurve symbolisiert sind. Daraus folgt die Identität

$$\binom{k+m}{k} = \sum_{i=0}^{m} \binom{k-1+i}{k-1} = \sum_{i=0}^{m} \binom{k-1+i}{i}.$$

Rekursion für die Stirling Zahlen zweiter Art

Satz: Für alle $n,k \in \mathbb{N}$ mit $n \geq k > 0$ gilt $S_{n,k} = S_{n-1,k-1} + k \cdot S_{n-1,k}$.

Man beachte, dass diese Formel eine große Ähnlichkeit mit der Rekursionsformel für Binomialkoeffizienten aufweist. Auch der Beweis ist sehr ähnlich:

Sei a ein fest gewähltes Element der n–Menge N. Die k–Partitionen von N können nun danach klassifiziert werden, ob $\{a\}$ einen Block der Partition bildet oder nicht. Bei positiver Antwort kann der Partition auf eineindeutige Weise (Streichung von $\{a\}$) eine $(k-1)$–Partition von $N \setminus \{a\}$ zugeordnet werden. Zur Begründung des zweiten Summanden überlegt man sich, das man alle k–Partitionen von N, in denen $\{a\}$ keinen Block bildet, auf folgende Art erhält: Man betrachte alle k–Partitionen von $N \setminus \{a\}$ und ergänze eine ihrer Klassen um a (dafür gibt es jeweils k Möglichkeiten).

Aus der Definition ergeben sich die Anfangsbedingungen $S_{n,0} = 0$ für $n > 0$ und $S_{0,k} = 0$ für $k > 0$. Wie man leicht sieht, muß $S_{0,0} = 1$ gesetzt werden, damit die Formel auch für $S_{1,1}$ ihre Gültigkeit behält.

3.5 Lösung von Rekursionen

In diesem Abschnitt werden zwei Methoden beschrieben, mit deren Hilfe man (in einigen Fällen) aus Rekursionen geschlossene Formeln ableiten kann. Wir beginnen mit Rekursionen der Form

$$a_n = ca_{n-1} + d,\ a_0 = C, \qquad c,d,C \in \mathbb{R}$$

Offensichtlich können wir $c = 1$ als Trivialfall (dann ist $a_n = nd + C$) ausschließen. Für $c \neq 1$ erhalten wir durch elementare Umformungen:

$$\begin{array}{lll} a_n & = d + ca_{n-1} & = d + c(d + ca_{n-2}) \\ & = d + cd + c^2 a_{n-2} & = d + cd + c^2(d + ca_{n-3}) \\ & = \ldots & = d + cd + c^2 d + \ldots + c^{n-1}d + c^n a_0 \\ & & = \frac{c^n - 1}{c-1} d + c^n C \end{array}$$

Wir betrachten jetzt sogenannte *homogene lineare Rekursionen vom Grad k*, das sind Rekursionen der Form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \ldots + c_k a_{n-k}$$

wobei $c_1, \ldots, c_k \in \mathbb{R}$ und $c_k \neq 0$ vorausgesetzt sind. Darüber hinaus sollen die reelle Zahlen $a_0 = C_0, \ldots, a_{k-1} = C_{k-1}$ als Anfangsbedingungen bekannt sein.

Der Basisansatz besteht darin, eine Lösung der Form $a_n = r^n$ für eine Konstante r zu suchen. Gäbe es eine solche Lösung, dann wäre $r^n = c_1 r^{n-1} + c_2 r^{n-2} + \ldots + c_k r^{n-k}$, also wäre r eine Nullstelle des sogenannten *charakteristischen Polynoms* $x^k - c_1 x^{k-1} - c_2 x^{k-2} \ldots - c_k$ der Rekursion. Eine solche Lösung könnte aber im Widerspruch zu den Anfangsbedingungen stehen.

Satz: Angenommen, das charakteristische Polynom der Rekursion
$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \ldots + c_k a_{n-k}$ hat k verschiedene Nullstellen $r_1, \ldots, r_k$. Dann gibt es Konstanten $\alpha_1, \ldots, \alpha_k$, so dass $a_n = \alpha_1 r_1^n + \ldots + \alpha_k r_k^n$ für alle $n \in \mathbb{N}$.
Diese Konstanten sind durch die k Anfangsbedingungen $a_0 = C_0, \ldots, a_{k-1} = C_{k-1}$ eindeutig bestimmt.

Wir verzichten auf den Beweis. Es sei aber angemerkt, dass zur Bestimmung der Konstanten ein lineares Gleichungssystem von k Gleichungen (jeder Anfangswert liefert eine Gleichung) mit den gesuchten Konstanten $\alpha_1, \ldots, \alpha_k$ als Variable gelöst werden muss. Die Verschiedenheit der Nullstellen sichert, dass dieses LGS eine eindeutige Lösung besitzt.

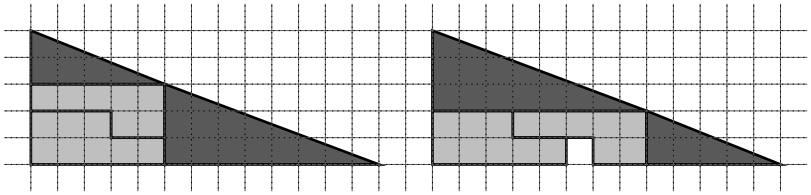
Beispiel: Die Rekursion $f_n = f_{n-1} + f_{n-2}$ mit den Anfangsbedingungen $f_0 = 0,\ f_1 = 1$ beschreibt die sogenannten *Fibonacci Zahlen*.
Das charakteristische Polynom dieser Rekursion hat die Form
$x^2 - x - 1$. Die Nullstellen dieses Polynoms sind $\frac{1+\sqrt{5}}{2}$ und $\frac{1-\sqrt{5}}{2}$.
Zur Bestimmung der Konstanten α_1 und α_2 muss das folgende LGS gelöst werden:

$$\begin{array}{rrl} \alpha_1 + & \alpha_2 & = 0 \\ \frac{1+\sqrt{5}}{2}\alpha_1 + & \frac{1-\sqrt{5}}{2}\alpha_2 & = 1 \end{array}$$

Aus der ersten Gleichung folgt $\alpha_2 = -\alpha_1$. Verwendet man diese Substitution in der zweiten Gleichung, dann ergibt sich $\sqrt{5}\alpha_1 = 1$, also $\alpha_1 = 1/\sqrt{5}$. So erhält man folgende geschlossene Formel für die Fibonacci Zahlen:

$$f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

Aus dieser Formel kann man sehr gut das exponentielle Wachstum der Fibonacci Zahlen und ihr Konvergenzverhalten ablesen. Eine geometrische Demonstration findet man in der folgenden Abbildung:



Aus vier Puzzleteilen wird jeweils ein 5×13 Dreieck zusammengesetzt, aber im rechten Dreieck bleibt ein Feld frei. Wie kann man das erklären?
Beim genaueren Hinsehen oder besser beim Nachrechnen entdeckt man, dass die Hypotenusen der beiden kleinen Dreiecke verschiedene Anstiege haben und somit keine Gerade bilden: $\frac{2}{5} = 0,4$ und $\frac{3}{8} = 0,375$. Unter Berücksichtigung von $2 = f_3,\ 3 = f_4,\ 5 = f_5$ und $8 = f_6$ erhält man eine geometrische Idee davon, dass sich die Quotienten f_3/f_5 und f_4/f_6 nur wenig unterscheiden.

3.6 Das Schubfachprinzip

In den bisherigen Themen zur Kombinatorik ging es immer um die genaue Bestimmung von bestimmten Werten. Es kommt aber auch vor, dass die Informationen über bestimmte Mengen nicht ausreichen, um genaue Angaben über ihre Größe machen zu können. Oft muss man sich in solchen Fällen damit zufrieden geben, möglichst gute obere und untere Schranken für die Größe zu kennen. Eine grundlegende und wichtige Methode zur Herleitung solcher Ungleichungen mit kombinatorischen Mitteln besteht in der Anwendung des sogenannten Schubfachprinzips.

Das *Schubfachprinzip*, in der englischsprachigen Literatur auch als *Taubenschlagprinzip* bezeichnet, geht auf den Mathematiker G. L. Dirichlet zurück:

Verteilt man n Gegenstände in k Schubfächer und ist $n > k$, dann gibt es mindestens ein Schubfach, in dem mindestens 2 Gegenstände liegen.

Auf den ersten Blick scheint dieses Prinzip nur triviale Schlussfolgerungen zuzulassen, wie: "in jeder Gruppe von 367 Menschen gibt es mindestens zwei, die den gleichen Geburtstag haben", aber wie die folgenden eleganten Anwendungen zeigen, trügt der Schein.

Satz: In jeder $n+1$ elementigen Untermenge M von $\{1, 2, \ldots, 2n\}$ gibt es zwei Zahlen a und b, so dass a Teiler von b ist.

Beweis: Sei $M = \{a_1, a_2, \ldots, a_{n+1}\}$. Jedes Element dieser Menge hat eine eindeutige Produktzerlegung der Form $a_i = 2^{k_i} q_i$, wobei q_i eine ungerade Zahl ist. Da die Anzahl der ungeraden Zahlen zwischen 1 und $2n$ gleich n ist, müssen mindestens zwei Elemente aus M bei der Zerlegung die gleiche ungerade Zahl liefern, und es ist klar, dass dann die kleinere von beiden ein Teiler der größeren ist.

Satz von Erdös/Szekeres: Jede Folge von $n^2 + 1$ verschiedenen Zahlen enthält eine monoton wachsende oder monoton fallende Unterfolge der Länge $n + 1$.

Beweis: Ist die Folge $a_1, a_2, \ldots, a_{n^2+1}$ gegeben, dann hat eine Unterfolge der Länge m die Form $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ wobei $1 \leq i_1 < i_2 < \ldots < i_m \leq n^2 + 1$ gelten muss. Die Unterfolge ist monoton wachsend (bzw. fallend), wenn auch $a_{i_1} < a_{i_2} < \ldots < a_{i_m}$ (bzw. $a_{i_1} > a_{i_2} > \ldots > a_{i_m}$) gilt.
Angenommen die gegebene Folge hat keine monoton wachsende oder monoton fallende Unterfolge der Länge $n + 1$, dann ordnen wir jedem $k = 1, 2, \ldots, n^2 + 1$ ein Paar (w_k, f_k) zu, wobei w_k (f_k) die maximale Länge einer monoton wachsenden (fallenden) Unterfolge ist, die mit a_k beginnt. Da nur n^2 Paare möglich sind, muss mindestens

ein Paar mehrfach auftreten, d.h. es gibt Zahlen $1 \leq s < t \leq n^2 + 1$, so dass $w_s = w_t$ und $f_s = f_t$. Das ist aber bereits ein Widerspruch, denn ist $a_s < a_t$, dann könnte jede mit a_t beginnende monoton wachsende Unterfolge von vorn durch a_s verlängert werden, also wäre $w_s > w_t$. Analog folgt aus $a_s > a_t$ auch $f_s > f_t$ und $a_s = a_t$ ist nach Voraussetzung nicht möglich. □

Verallgemeinerung des Schubfachprinzips:
Verteilt man n Gegenstände in k Schubfächer, dann gibt es mindestens ein Schubfach, in dem mindestens $\lceil \frac{n}{k} \rceil$ Gegenstände liegen.

Beispiele:

- In jeder Gruppe von 100 Menschen gibt es mindestens 9, die das gleiche Sternzeichen haben.

- In jeder Menge A, B, C, D, E, F von 6 Besuchern einer Party gibt es eine Gruppe von mindestens drei Personen, die sich entweder vorher schon paarweise kannten oder die sich vorher paarweise noch nicht kannten.

  Das ist nicht ganz so offensichtlich. Für A gibt es nach dem Schubfachprinzip mindestens $\lceil \frac{5}{2} \rceil = 3$ andere Personen, die A entweder alle kennt (Fall 1) oder alle nicht kennt (Fall 2).
  Wenn sich im Fall 1 unter diesen Bekannten von A zwei Personen bereits kennen, bilden diese zusammen mit A die gesuchte Gruppe, sonst sind die Bekannten von A eine Gruppe von Personen, die sich nicht kennen. Der zweite Fall ist analog. Wenn unter den für A fremden Personen sich zwei nicht kennen, bilden sie zusammen mit A die gesuchte Gruppe, sonst bilden sie selbst eine Gruppe von Personen, die sich paarweise kennen.

4 Diskrete Wahrscheinlichkeitsverteilungen

4.1 Wahrscheinlichkeitsräume, Ereignisse und Unabhängigkeit

Definition: Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar $(\Omega, \Pr)$, wobei Ω eine endliche oder abzählbar unendliche Menge von elementaren Ereignissen und $\Pr : \Omega \longrightarrow [0,1]$ eine *Wahrscheinlichkeitsverteilung* ist, die jedem $a \in \Omega$ seine *Wahrscheinlichkeit* $\Pr(a)$ zuordnet. Außerdem wird gefordert, dass $\sum_{a\in\Omega} \Pr(a) = 1$ ist. Wenn $|\Omega| = n$ endlich ist und für alle $a \in \Omega$ auch $\Pr(a) = \frac{1}{n}$ gilt, dann wird $\Pr$ eine Gleichverteilung genannt.

Beispiele:

1. Für einen (fairen) Würfel ist $\Omega = \{1,2,3,4,5,6\}$ und $\Pr(1) = \Pr(2) = \ldots = \Pr(6) = \frac{1}{6}$ eine Gleichverteilung.

2. Der diskrete Wahrscheinlichkeitsraum für einen Münzwurf besteht aus $\Omega = \{0,1\}$ mit $\Pr(0) = \Pr(1) = 1/2$ (Gleichverteilung), wobei in der Regel die 0 den Kopf und 1 die Zahl der Münze bezeichnet.

3. Die Ziehung von Kugeln aus einer Urne bildet die Grundlage für ein anderes Standardmodell der Wahrscheinlichkeitsrechnung. So kann man eine Urne mit n nummerierten (also unterscheidbaren) Kugeln als ein physikalisches Modell für eine Gleichverteilung ansehen. Alternativ kann man eine Urne mit k weißen und m schwarzen Kugeln betrachten. Auch wenn die weißen Kugeln nicht unterscheidbar sind, kann man sich eine unsichtbare Nummerierung der Kugeln vorstellen, bei denen die weißen von 1 bis k und die schwarzen von $k+1$ bis $n = k+m$ nummeriert sind. Die elementaren Ereignisse sind die Ziehungen einer Kugel i jeweils mit der Wahrscheinlichkeit $\frac{1}{n}$. Da es aber letztlich egal ist, welche weiße bzw. welche schwarze Kugel gezogen wurde, betrachtet man hier nur zwei Ereignisse, nämlich die Ziehung einer weißen Kugel mit der Wahrscheinlichkeit $\frac{k}{n}$ und die Ziehung einer schwarzen mit der Wahrscheinlichkeit $\frac{m}{n}$.

4. Der diskrete Wahrscheinlichkeitsraum für zwei Münzwürfe besteht aus der Menge $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$ mit der Gleichverteilung $\Pr((i,j)) = 1/4$. Dabei geht man davon aus, dass die beiden Würfe voneinander unabhängig sind und entweder hintereinander oder gleichzeitig mit zwei unterscheidbaren Münzen erfolgen.

5. Beim gleichzeitigen Werfen von zwei ununterscheidbaren Münzen müsste man $\Omega = \{\{0,0\},\{0,1\},\{1,1\}\}$ setzen. In diesem Fall wäre $\Pr$ (faire Münzen vorausgesetzt) keine Gleichverteilung, denn $\Pr(\{0,0\}) = \Pr(\{1,1\}) = 1/4$ und $\Pr(\{0,1\}) = 1/2$. Dieses Modell ist eher als Gedankenspiel zu betrachten, denn in der Realität sind es zwei verschiedene Münzen, auch wenn sie äußerlich nicht unterscheidbar sind. Wenn wir im Weiteren über die Wahrscheinlichkeitsräume von k Münzwürfen oder k Würfeln sprechen, wird deshalb immer die k-fache

Wiederholung eines Experiments oder eine Ausführung mit k unterscheidbaren und unabhängigen Objekten gemeint sein.

6. Definiert man eine Folge von Münzwürfen, bis das erste Mal Zahl (also 1) fällt, als elementares Ereignis, dann gibt es offensichtlich für jedes $n \in \mathbb{N}^+$ ein elementares Ereignis, nämlich die Folge $a_n = \underbrace{0, 0, \ldots, 0}_{(n-1) \text{ mal}}, 1$. Eine faire Münze vorausgesetzt, gilt $\Pr(a_n) = \left(\frac{1}{2}\right)^n$ und damit ist auch für diesen unendlichen Raum die Summenbedingung erfüllt:

$$\sum_{n=1}^{\infty} \Pr(a_n) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = 1$$

Definition: Eine beliebige Untermenge $A \subseteq \Omega$ wird *Ereignis* genannt, und man definiert $\Pr(A) = \sum_{a \in A} \Pr(a)$. Dadurch wird die Verteilungsfunktion zu einem Wahrscheinlichkeitsmaß über der Menge $\mathcal{P}(\Omega)$ aller Ereignisse erweitert. Die Verwendung der gleichen Bezeichnung Pr für die Verteilung und das Wahrscheinlichkeitsmaß ist eine kleine technische Unsauberkeit, die aber in der Literatur weit verbreitet ist.

Beispiele:

- Bei der Urne mit k weißen und m schwarzen Steinen wurde dieser Schritt schon vorweg genommen, indem die Ziehungen der Kugeln 1 bis k (bzw. $k+1$ bis $k+m$) zum Ereignis der Ziehung einer weißen (bzw. schwarzen) Kugel zusammengefasst wurden.

- Die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, ist gleich der Summe der Wahrscheinlichkeiten, eine $2, 4$ oder 6 zu würfeln, also $1/2$.

- Das Ereignis A, bei drei Münzwürfen genau zweimal Zahl zu erhalten, setzt sich aus den drei elementaren Ereignissen $(0,1,1), (1,0,1)$ und $(1,1,0)$ zusammen und somit ist $\Pr(A) = 3 \cdot \frac{1}{8} = \frac{3}{8}$.

Für jedes Ereignis $A \subseteq \Omega$ bezeichnet $\overline{A} = \Omega \setminus A$ das sogenannte *Komplementärereignis* von A.
Aus der Definition und dem kombinatorischen Basiswissen kann man die folgenden Eigenschaften von Wahrscheinlichkeitsmaßen ableiten:

1. $\Pr(\Omega) = 1$ und $\Pr(\emptyset) = 0$
2. $A \subseteq B \Longrightarrow \Pr(A) \leq \Pr(B)$
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
4. $\Pr(A) = \sum_{i=1}^{k} \Pr(A_i)$ für jede Partition $A = \bigcup_{i=1}^{k} A_i$

5. $\Pr\left(\overline{A}\right) = 1 - \Pr(A)$

6. $\Pr\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} \Pr(A_i)$.

Bedingte Wahrscheinlichkeit und unabhängige Ereignisse

Oft taucht das Problem auf, die Wahrscheinlichkeit eines Ereignisses A unter der zusätzlichen Voraussetzung zu bestimmen, dass ein anderes Ereignis B eingetreten ist. So ist offensichlich unter der Voraussetzung, dass eine gerade Zahl gewürfelt wurde, die Wahrscheinlichkeit der 1 gleich Null und die Wahrscheinlichkeit der 2 gleich 1/3. Dieses Phänomen wird durch den Begriff der bedingten Wahrscheinlichkeit beschrieben.

Definition: Seien A und B Ereignisse und $\Pr(B) > 0$, dann nennt man den Ausdruck $\Pr(A|B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$ die *bedingte Wahrscheinlichkeit von A unter B.*

Beispiele:

1. Im Wahrscheinlichkeitsraum eines fairen Würfels sei A das Ereignis, eine gerade Zahl zu werfen, B das Ereignis, eine Zahl größer als drei zu werfen, und C das Komplementärereignis zu B. Offensichtlich gilt $\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{2}$. Man bestimmt die Durchschnitte $A \cap B = \{4, 6\}$, $A \cap C = \{2\}$, $B \cap C = \emptyset$ und ihre Wahrscheinlichkeiten $\Pr(A \cap B) = \frac{2}{6}$, $\Pr(A \cap C) = \frac{1}{6}$ sowie $\Pr(B \cap C) = 0$ und berechnet damit die bedingten Wahrscheinlichkeiten:
   $\Pr(B|A) = \frac{2}{3}$, $\Pr(C|A) = \frac{1}{3}$ und $\Pr(B|C) = 0$.

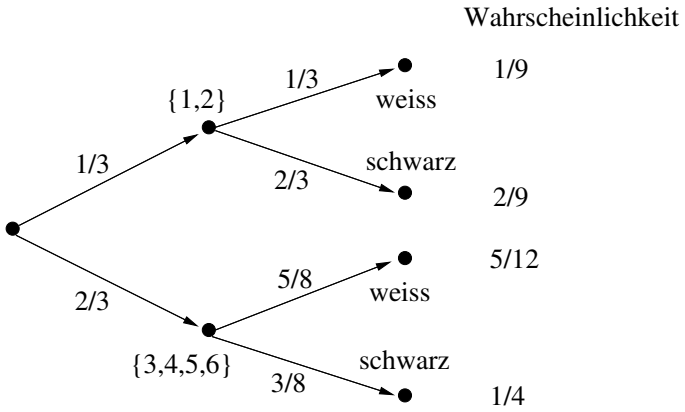
2. Würfelt man mit zwei (unabhängigen und unterscheidbaren!) Würfeln, dann ist $\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\}$ und $\Pr((i, j)) = 1/36$ für jedes Paar (i, j) aus Ω. Sei $A = \{(4, 6), (5, 5), (6, 4)\}$ das Ereignis, dass die Augensumme gleich 10 ist, und $B = \{(1, 1), (2, 2), \ldots, (6, 6)\}$ das Ereignis, dass beide Würfel auf die gleiche Zahl fallen. Dann ist $\Pr(A) = \frac{3}{36} = \frac{1}{12}$, $\Pr(A) = \frac{6}{36} = \frac{1}{6}$ und $\Pr(A \cap B) = \frac{1}{36}$. Daraus ergibt sich $\Pr(A|B) = \frac{1}{3}$ und $\Pr(B|A) = \frac{1}{6}$.

3. Wir bleiben im Wahrscheinlichkeitsraum von zwei Würfeln. Bezeichne $A_1 = \{(i, j) \in \Omega \mid i = 1\}$ das Ereignis, dass der erste Würfel eine 1 zeigt und der zweite beliebig ist. Analog lassen sich Ereignisse $A_2, \ldots, A_6$ beschreiben. Auf der anderen Seite können durch Festlegung der Augenzahl des zweiten Würfels Ereignisse $B_1 = \{(i, j) \in \Omega \mid j = 1\}$, $B_2, \ldots, B_6$ definiert werden. Offensichtlich ist die Wahrscheinlichkeit dieser Ereignisse jeweils 1/6 und Durchschnitte wie $A_1 \cap B_3 = \{(1, 3)\}$ haben die Wahrscheinlichkeit 1/36. Folglich ist

   $$\Pr(A_1|B_3) = \frac{\Pr(A_1 \cap B_3)}{\Pr(B_3)} = \frac{1/36}{1/6} = \frac{1}{6}$$

   und analog ist die bedingte Wahrscheinlichkeit von jedem A_i unter einem beliebigen B_j gleich 1/6, also gleich der Wahrscheinlichkeit von A_i selbst. Diese Eigenschaft ist durch die Unabhängigkeit der beiden Würfe begründet.

4. In einigen Fällen kann es hilfreich sein, die Sachverhalte durch ein Diagramm zu veranschaulichen. Angenommen wir haben eine rote Urne mit zwei weißen und vier schwarzen Kugeln und eine blaue Urne mit fünf weißen und drei schwarzen Kugeln. Zuerst wird gewürfelt und wenn das Ergebnis 1 oder 2 ist, wird eine Kugel aus der roten Urne, sonst eine Kugel aus der blauen Urne gezogen. Die elementaren Ereignisse sind Paare aus einer Zahl zwischen 1 und 6 und einer Farbe Schwarz oder Weiß. Das folgende Diagramm zeigt, wie sich die Ergebnisse des Experiments aufspalten. Die Werte auf den Kanten geben bedingte Wahrscheinlichkeiten an, die Wahrscheinlichkeiten der Ereignisse auf der rechten Seite ergeben sich als Produkt der Wahrscheinlichkeiten auf den Kanten des jeweiligen Wegs. Die Wahrscheinlichkeit, eine weiße Kugel zu ziehen, ist demnach $\frac{1}{9} + \frac{5}{12} = \frac{19}{36}$.



Definition: Zwei Ereignisse A und B werden *unabhängig* genannt, falls $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.

Lemma: Sind A und B zwei unabhängige Ereignisse mit
$\Pr(A \cap B) > 0$, dann gilt $\Pr(A|B) = \Pr(A)$ und $\Pr(B|A) = \Pr(B)$.

Ereignisse mit positver Wahrscheinlichkeit, die einander ausschließen, wie zum Beispiel A_1 und A_2 aus dem letzten Beispiel sind voneinander nicht unabhängig, also abhängig. Weitere Beispiele für abhängige Ereignisse beim Spiel mit einem Würfel ist der Wurf einer geraden Zahl und der Wurf einer 2 oder auch der Wurf einer geraden Zahl und der Wurf einer Zahl ≤ 3. Auch im folgenden Beispiel geht es um die Abhängigkeit von Ereignissen.

Beispiel: Ein Krebstest fällt mit 96% Wahrscheinlichkeit positiv aus, wenn der Patient Krebs hat und mit 94% Wahrscheinlichkeit negativ, wenn der Patient keinen Krebs hat. Bei einem Patienten in dessen Altersgruppe 0.5% aller Personen Krebs haben, fällt der Test positiv aus. Wie wahrscheinlich ist es, dass er tatsächlich Krebs hat?
Wir betrachten die folgenden Ereignisse

K	zufälliger Patient hat Krebs	$\Pr(K) = 0.005$
N	zufälliger Patient hat keinen Krebs	$\Pr(N) = 0.995$
T	Test positiv bei zufälligem Patienten	$\Pr(T) = ??$

Gesucht ist die bedingte Wahrscheinlichkeit $\Pr(K|T) = \dfrac{\Pr(K \cap T)}{\Pr(T)}$.

Aus $\Pr(T|K) = \dfrac{\Pr(K \cap T)}{\Pr(K)} = 0.96$ kann man $\Pr(K \cap T) = 0.005 \cdot 0.96$ ausrechnen.
Dagegen ist $\Pr(T|N) = 1 - \Pr\big(\overline{T}|N\big) = 1 - 0.94 = 0.06$. Zur Bestimmung von $\Pr(T)$ überlegt man sich, dass das Ereignis T die disjunkte Vereinigung $(T \cap K) \cup (T \cap N)$ ist und sich die Wahrscheinlichkeiten der beiden Komponenten wieder durch bedingte Wahrscheinlichkeiten ausdrücken lassen:
$\Pr(T) = \Pr(T|K) \cdot \Pr(K) + \Pr(T|N) \cdot \Pr(N) = 0.96 \cdot 0.005 + 0.06 \cdot 0.995$. Letztlich erhalten wir

$$\Pr(K|T) = \frac{\Pr(K \cap T)}{\Pr(T)} = \frac{0.005 \cdot 0.96}{0.96 \cdot 0.005 + 0.06 \cdot 0.995} = 0.0744\ldots$$

Damit ist die Wahrscheinlichkeit, dass der Patient mit positivem Test wirklich Krebs hat, kleiner als 7.5%.

Definition: Sind $(\Omega_1, \Pr_1)$ und $(\Omega_2, \Pr_2)$ zwei Wahrscheinlichkeitsräume, dann ist der Produktraum $(\Omega_1 \times \Omega_2, \Pr)$ definiert durch $\Pr((a_1, a_2)) = \Pr_1(a_1) \cdot \Pr_2(a_2)$.

Diese Definition impliziert Unabhängigkeiten der folgenden Form: für alle $A \subseteq \Omega_1$ und $B \subseteq \Omega_2$ sind die Ereignisse $A \times \Omega_2$ und $\Omega_1 \times B$ voneinander unabhängig. Ein Beispiel dafür wurde bereits behandelt, denn der Wahrscheinlichkeitsraum für zwei Würfel ist das Produkt des Raums eines Würfels mit sich selbst. Im folgenden werden endliche Produkte des Münzwurfraums genauer analysiert:

- Eine faire Münze wird sechsmal geworfen. Der zugehörige Wahrscheinlichkeitsraum besteht aus allen 6-Tupeln (oder 0/1–Folgen der Länge 6) mit Gleichverteilung, d.h. jedes Tupel hat die Wahrscheinlichkeit $\frac{1}{2^6}$. Wie groß ist die Wahrscheinlichkeit des Ereignisses, dass genau zwei Einsen in dem Tupel auftreten? Die Anzahl dieser Tupel stimmt offensichtlich mit der Anzahl der 2–Kombinationen einer 6–Menge überein. Folglich hat das Ereignis die Wahrscheinlichkeit $\binom{6}{2}/2^6$. Allgemein ist die Wahrscheinlichkeit für k mal Zahl bei n Münzwürfen gleich $\binom{n}{k}\frac{1}{2^n}$.

- Sei eine "unfaire" Münze gegeben, die mit Wahrscheinlichkeit p $(0 < p < 1)$ Zahl (d.h. 1) und mit Wahrscheinlichkeit $q = 1 - p$ Kopf (d.h. 0) liefert. Betrachtet man den Wahrscheinlichkeitsraum für n Münzwürfe, dann ist die Wahrscheinlichkeit einer festen Folge der Länge n mit genau k Einsen gleich $p^k(1-p)^{n-k}$. Die Wahrscheinlichkeit des Ereignisses, dass eine Folge genau k Einsen enthält, ist $\binom{n}{k}p^k(1-p)^{n-k}$. Deshalb werden Verteilungen dieser Art auch *Binomialverteilungen* genannt.

4.2 Zufallsvariable und Erwartungswert

Definition: Sei $(\Omega, \Pr)$ ein Wahrscheinlichkeitsraum. Eine *Zufallsvariable* oder *Zufallsgröße* ist eine Funktion $X : \Omega \longrightarrow \mathbb{R}$.

Im Fall eines Würfels oder einer zufällig gezogenen Zahl aus einer Urne fällt der Unterschied zwischen einem elementaren Ereignis und einer Zufallsvariable nicht auf, weil das Ereignis selbst schon eine Zahl ist, d.h. man kann die identische Abbildung als eine Zufallsvariable ansehen. Die folgenden Beispiele zeigen, dass man viele Zufallsexperimente auf sehr natürliche Weise mit Zufallsvariablen beschreiben kann.

Beispiele:

1) Auf dem diskreten Wahrscheinlichkeitsraum $\Omega = \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$ für zwei Münzwürfe sind Zufallsvariablen für die Summe und die Differenz der beiden Werte interessant: $X(a, b) = a + b \qquad Y(a, b) = |a - b|$.

2) Auf dem Wahrscheinlichkeitsraum für n Münzwürfe $\Omega = \{0, 1\}^n$ interessiert man sich vor allem für die Zufallsvariablen, welche die Anzahl der Zahl- bzw. Kopfergebnisse in einer solchen Serie zählen, d.h.

$$\begin{aligned} X(a_1, a_2, \ldots, a_n) &= |\{i \,|\, 1 \leq i \leq n \text{ und } a_i = 1\}| \\ Y(a_1, a_2, \ldots, a_n) &= |\{i \,|\, 1 \leq i \leq n \text{ und } a_i = 0\}| = n - X(a_1, a_2, \ldots, a_n) \end{aligned}$$

3) Bei der Wiederholung von Münzwürfen bis zur ersten Zahl (oder bis zum ersten Kopf) ist offensichtlich die Länge des Experiments eine interessante zufällige Größe:

$$X(\underbrace{0, 0, \ldots, 0}_{(n-1) \text{ mal}}, 1) = n$$

Definition: Der *Erwartungswert* der Zufallsvariablen X ist definiert als $E(X) = \sum_{a \in \Omega} X(a)\Pr(a)$.

Im ersten und dritten Beispiel kann man Erwartungswerte relativ leicht ausrechnen:

1) $E(X) = 7$ und $E(Y) = \frac{70}{36}$;

3) $E(X) = \sum_{n=1}^{\infty} n \cdot \frac{1}{2^n} = 2$ (ist etwas schwerer).

Im zweiten Beispiel ist $E(X) = \frac{n}{2}$ die richtige Antwort. Intuitiv ist das klar, aber eine direkte Berechnung durch konkrete Auswertung der Definition ist relativ aufwändig. Der folgende Fakt macht die Bestimmung des Erwartungswerts sehr einfach.

Lemma: Der Erwartungswert ist *additiv*, d.h. für beliebige Zufallsvariable $X, Y : \Omega \longrightarrow \mathbb{R}$ gilt $E(X + Y) = E(X) + E(Y)$, wobei die neue Variable $X + Y$ in naheliegender Weise durch $(X + Y)(a) = X(a) + Y(a)$ definiert ist.

Anwendungen:

- Wir betrachten die binomialverteilte Zufallsvariable X aus Beispiel 2, die jedem 0/1–Tupel der Länge n die Anzahl seiner Einsen zuordnet. Um das Beispiel noch etwas allgemeiner zu machen, betrachten wir eine unfaire Münze, die mit Wahrscheinlichkeit p eine 1 (Zahl) und mit Wahrscheinlichkeit $q = 1 - p$ eine 0

(Kopf) liefert. Gruppiert man die Folgen derart, dass alle Folgen mit genau k Einsen einer Gruppe angehören, so erhält man $E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$. Es ist relativ aufwändig, daraus $E(X) = np$ abzuleiten. Durch Nutzung der Additivität der Erwartungswerte kann man diese Formel einfacher erhalten: Offensichtlich ist $X = X_1 + X_2 + \ldots + X_n$, wobei die Variable X_i angewendet auf eine Folge von Münzwurfergebnissen $(a_1, \ldots, a_n)$ das Ergebnis des i-ten Wurfs auswertet, also $X_i((a_1, \ldots, a_n)) = a_i$. Da die Erwartungswerte dieser Summanden jeweils p sind, folgt $E(X) = E(X_1) + \ldots + E(X_n) = np$.

- An einem Tisch sitzen 12 Personen. Von jeder Person wurde ein Foto gemacht. Diese 12 Bilder liegen verdeckt auf dem Tisch. Jede Person zieht zufällig ein Foto (zufällig bedeutet hier, dass alle möglichen Zuordnungen der Bilder gleich wahrscheinlich sein sollen). Wie hoch ist die erwartete Anzahl der Personen, die ihr eigenes Foto gezogen haben?

  Wieder kann man die Zufallsvariable X, welche die Anzahl dieser Personen zählt, als Summe von zwölf Variablen $X_1, \ldots, X_{12}$ darstellen, wobei X_i den Wert 1 hat, wenn Person i ihr eigenes Foto gezogen hat, und 0 sonst. Offensichtlich ist $E(X_i) = \frac{1}{12}$ für alle i und folglich ist $E(X) = 12 \cdot \frac{1}{12} = 1$.
  Aus dieser Betrachtung ergibt sich außerdem die (vielleicht etwas überraschende) Tatsache, dass der Erwatungswert nicht von der Anzahl der Personen abhängt - auch bei 100 Personen zieht im Erwartungswert eine Person ihr eigenes Foto.

- Eine Versicherung macht folgende Kalkulation: Jeder Versicherte zahlt 100 Euro ein. Mit Wahrscheinlichkeit von 1/20 (also 5 %) muss eine Leistung von 1000 Euro und mit Wahrscheinlichkeit 1/10 muss eine Leistung von 400 Euro erbracht werden. Wie hoch ist der Erwartungswert für den Gewinn (pro Versicherungsnehmer)?

  Der zu betrachtende Wahrscheinlichkeitsraum hat drei Elementarereignisse a_1, a_2 und a_3 mit $\Pr(a_1) = 1/20$, $\Pr(a_2) = 1/10$ und $\Pr(a_3) = 1-1/20-1/10 = 17/20$. Die Zufallsvariable X, die den Gewinn der Versicherung beschreibt, ist die Differenz aus einer Zufallsvariablen Y für die Einnahmen und einer Zufallsvariablen Z für die Ausgaben in den einzelnen Ereignissen: Dabei hat Y in jedem Fall den Wert 100 und $Z(a_1) = 1000$, $Z(a_2) = 400$ und $Z(a_3) = 0$. Nach Definition erhalten wir für den Erwartungswert

$$E(X) = E(Y) - E(Z) = 100 - (1000 \cdot \frac{1}{20} + 400 \cdot \frac{1}{10} + 0) = 100 - 90 = 10$$

In vielen praktischen Anwendungen entstehen Zufallsvariable durch Messung von zufälligen Größen, die auf sehr schwer zu analysierende Wahrscheinlichkeitsräume zurückgehen. Beispiele dafür sind die Zeit, die man für seinen täglichen Arbeitsweg braucht oder die Körpergröße einer Person. Deshalb spricht man häufig auch von der Verteilung einer Variablen, d.h. man interessiert sich für die Wahrscheinlichkeit,

dass die Variable einen bestimmten Wert (oder einen Wert aus einem bestimmten Intervall) annimmt und weniger für die Wahrscheinlichkeiten der einzelnen elementaren Ereignisse. Die Wahrscheinlichkeit, dass eine Zufallsvariable X den Wert x_0 bzw. einen Wert kleiner oder gleich x_0 annimmt wird mit $\Pr(X = x_0)$ bzw. $\Pr(X \leq x_0)$ notiert. In den gerade genannten Beispielen sind das sogenannte Normalverteilungen, die man nicht mit einem diskreten Modell beschreiben kann. Es gibt aber auch eine Reihe von diskreten Zufallsvariablen, die häufig auftreten und deshalb in bestimmte Typen eingeteilt werden.

- Eine *Gleichverteilung* liegt vor, wenn eine Zufallsvariable k verschiedene Werte annehmen kann und jeder Wert die Wahrscheinlichkeit $1/k$ hat.

- Eine *Bernoulli–Verteilung* mit dem Parameter p liegt vor, wenn eine Zufallsvariable X nur die Werte 0 und 1 annimmt und $\Pr(X = 1) = p$, $\Pr(X = 0) = 1 - p$ gilt.

- Eine Zufallsvariable X ist *binomialverteilt* mit den Parametern n und p, wenn sie die Werte $0, 1, 2, \ldots, n$ annimmt und $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für alle k aus dem Wertebereich gilt. Mit anderen Worten entsteht eine Binomialverteilung immer dann, wenn man ein Bernoulli–Experiment n–mal wiederholt und die Zufallsvariable die Anzahl der 1–Ergebnisse zählt.

- Eine Zufallsvariable X ist *geometrisch verteilt* mit dem Parameter p, wenn sie alle Werte aus $\mathbb{N}^+$ annehmen kann und $\Pr(X = k) = (1-p)^{k-1} p$ für alle $k \in \mathbb{N}^+$ gilt. Mit anderen Worten entsteht eine geometrische Verteilung immer dann, wenn man ein Bernoulli–Experiment so lange wiederholt, bis zum ersten Mal eine 1 fällt und die Zufallsvariable die Anzahl der Versuche zählt.

Wichtige Anwendungen von Erwartungswerten in der Informatik sind die *Analyse der durchschnittlichen Laufzeit* von Algorithmen (Ω besteht dann aus den Eingaben einer bestimmten Größe und $X(a)$ ist die Laufzeit für eine konkrete Eingabe a) und die Analyse *randomisierter Algorithmen*. Im letzten Fall wird die Arbeit des Algorithmus durch eingebaute Zufälle gesteuert (z.B. Auswahl eines zufälligen Pivotelements bei Quicksort) und somit ist die Laufzeit für jede konkrete Eingabe eine Zufallsvariable, deren Wert von den zufälligen Entscheidungen im Ablauf des Algorithmus abhängt.

5 Grundlagen der Graphentheorie

5.1 Graphen und ihre Darstellungen

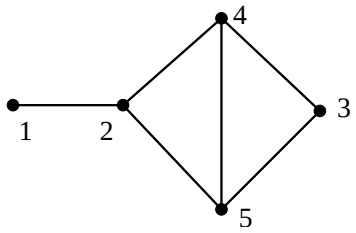
Ein *Graph* beschreibt Beziehungen zwischen den Elementen einer Menge von Objekten. Die Objekte werden als *Knoten* des Graphen bezeichnet – besteht zwischen zwei Knoten eine Beziehung, so sagen wir, dass es zwischen ihnen eine *Kante* gibt.

Definition: Für eine Menge V bezeichne $\binom{V}{2}$ die Menge aller zweielementigen Untermengen von V. Ein *einfacher, ungerichteter Graph* $G = (V, E)$ (kurz *Graph* genannt) besteht aus einer endlichen Menge V von *Knoten* , auch *Ecken (Vertex)* genannt, und einer Menge $E \subseteq \binom{V}{2}$ von *Kanten (Edge)*. Hier sind Kanten ungeordnete Paare von Knoten, d.h. $\{u, v\}$ und $\{v, u\}$ sind zwei verschiedene Schreibweisen für ein und dieselbe Kante. Im Gegensatz dazu ist ein *gerichteter Graph* G ein Paar (V, E) bestehend aus einer endlichen Knotenmenge V und einer Kantenmenge E von geordneten Knotenpaaren $e = (u, v)$, mit $u, v \in V$.

Ist $e = \{u, v\}$ eine Kante von G, dann nennt man die Knoten u und v zueinander *adjazent* oder *benachbart* und man nennt sie *inzident* zu e. Die Menge $N(v) = \{u \in V \mid \{u, v\} \in E\}$ der zu einem Knoten v benachbarten Knoten wird die *Nachbarschaft* von v genannt. Der *Grad* eines Knotens v wird durch $deg(v) = |N(v)|$ definiert. In gerichteten Graphen bezeichnet man die Anzahl der von einem Knoten v ausgehenden (die Anzahl der zu v führenden Kanten) als *Ingrad* (*Ausgrad*) von v.
Die Anzahl der Knoten $|V|$ bestimmt die *Ordnung* und die Anzahl der Kanten $|E|$ die *Größe* eines Graphen.

Im Folgenden werden verschiedene Darstellungen von Graphen an einem Beispiel demonstriert.

1) Darstellung als Zeichnung:



2) Darstellung als Adjazenzmatrix. Jedem Knoten wird eine Zeile und eine Spalte zugeordnet und der Eintrag in der Zeile von u und der Spalte von v wird 1 gesetzt, wenn $\{u, v\}$ eine Kante des Graphen ist (sonst 0):

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

3) Darstellung als Adjazenzliste. Für jeden Knoten wird die Liste seiner Nachbarn

angegeben (Liste von Listen):
$1:2;\ 2:1,4,5;\ 3:4,5;\ 4:2,3,5;\ 5:2,3,4$; oder mit anderer Syntax
$(2),(1,4,5),(4,5),(2,3,5),(2,3,4)$

4) Darstellung als Inzidenzmatrix. Jedem Knoten wird eine Zeile und jeder Kante eine Spalte zugeordnet und der Eintrag in der Zeile von u und der Spalte von e wird 1 gesetzt, wenn u eine Ecke der Kante e ist (sonst 0):

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 9 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Für die Behandlung algorithmischer Probleme sind oft Adjazenzlisten die Datenstruktur der Wahl, weil sie die wahre Größe eines Graphen (Anzahl der Kanten) widerspiegeln, während Adjazenzmatrizen per Definition $|V|^2$ Einträge haben. Andererseits haben Adjazenzmatrizen den Vorteil, dass Anfragen, ob zwei Knoten u und v durch eine Kante verbunden sind in konstanter Zeit (ein Speicherzugriff) beantwortet werden können, während man Adjazenzlisten durchsuchen muss.
Während die Adjazenzmatrix eines ungerichteten Graphen symmetrisch ist (Diagonale als Symmetrieachse), sind Adjazenzmatrizen von gerichteten Graphen im allgemeinen nicht symmetrisch.

Da Graphen über eine sehr einfache Struktur verfügen, finden sie bei der Modellierung und algorithmischen Lösung vieler praktischer Probleme Anwendung, wie z.B.

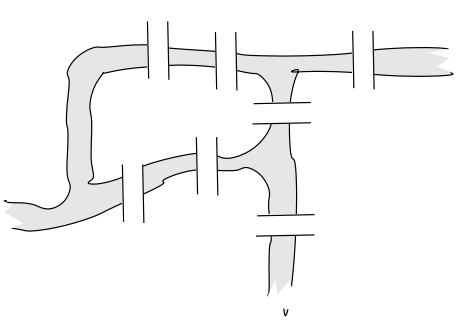
- Modellierung von Straßen-, Flug- und Telefonnetzen
- Darstellung von Molekülen
- Interpretation von Relationen (Beziehungsgeflechten)
- Gerüste von Polyedern (lineare Optimierung)
- Entwurf von Mikrochips (VLSI-Design)

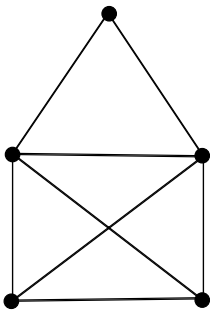
Die folgenden Beispiele für graphentheoretische Aufgabenstellungen haben die Entwicklung der Graphentheorie stark beeinflusst und unterstreichen die praktische Relevanz dieser Struktur:

1. Das 4-Farben-Problem: Man stelle sich die Welt mit einer beliebigen politischen Landkarte vor. Wir definieren einen Graphen, indem wir jedem Land einen Knoten zuordnen und zwei Knoten mit einer Kante verbinden, wenn sie einen gemeinsamen Grenzabschnitt haben. Wie viele Farben braucht man, um die Länder so einzufärben, dass benachbarte Länder verschiedene Farben haben? Man hat (mit Computerhilfe) bewiesen, dass vier Farben immer ausreichen! Einen Beweis ohne Computer gibt es bis heute nicht.

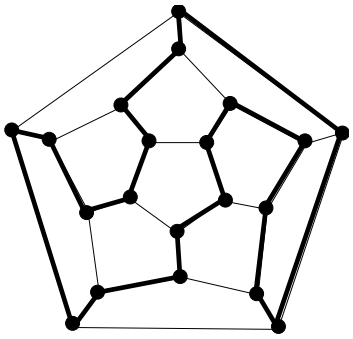
2. Welche Graphen treten als Ecken-Kanten-Gerüst von Polyedern auf? Solche Graphen spielen insbesondere bei der Lösung von linearen Optimierungsproblemen eine große Rolle.

3. Eulersche Kreise: Man charakterisiere jene Graphen, bei denen man die Kanten so durchlaufen kann, dass man jede Kante einmal benutzt und man am Schluss wieder am Ausgangspunkt steht. Der Ausgangspunkt für diese Frage war das von Euler gelöste sogenannte Königsberger Brückenproblem, bei dem es um die Frage geht, ob man einen Rundgang durch die Stadt machen kann, bei dem jede Brücke genau einmal überquert wird. Die Antwort ist negativ, weil auf jeder Flussseite und auf der Insel eine gerade Anzahl von Brücken ankommen müsste, um eine solchen Rundgang realisieren zu können. Wenn man auf die Bedingung der Gleichheit von Anfangs- und Endpunkt verzichtet, können auch zwei Knoten ungeraden Grads auftreten. Man spricht dann von Eulerschen Wegen. Auf der rechten Seite ist ein solcher Graph abgebildet, dessen Eulerscher Weg aus dem "Haus vom Nikolaus"-Prinzip bekannt sein sollte.





4. Hamiltonsche Graphen:
   Dies sind Graphen, die man so durchlaufen kann, dass man jeden Knoten genau einmal besucht, bis man zum Ausgangsknoten zurückkehrt. Während man für das vorherige Problem effiziente algorithmische Lösungen kennt, ist dieses algorithmisch schwer (NP-vollständig).



5. Travelling Salesman Problem (TSP): Oft hat man es mit bewerteten Graphen zu tun, das heißt Kanten und/oder Knoten haben zusätzliche Informationen wie Gewichte, Längen, Farben etc.
   Ein Beispiel ist das TSP. Wir haben n Städte. Für jedes Paar $\{u, v\}$ von Städten kennt man die Kosten, um von u nach v zu kommen. Man entwerfe für den Handelsreisenden eine geschlossene Tour, die alle Städte besucht und minimale Gesamtkosten hat. Auch dies ist ein algorithmisch schweres Problem.

6. Planare Graphen: Welche Graphen lassen sich so in der Ebene zeichnen, dass sich Kanten nicht schneiden, also sich höchstens in Knoten berühren? Wie kann man sie charakterisieren und algorithmisch schnell erkennen?

7. Netzwerke: Welchen Graphen sollte man der Architektur eines Rechnernetzes zu Grunde legen, wenn die Kanten beschränkte Kapazität haben, aber trotzdem schneller Informationsaustausch gewährleistet werden soll?

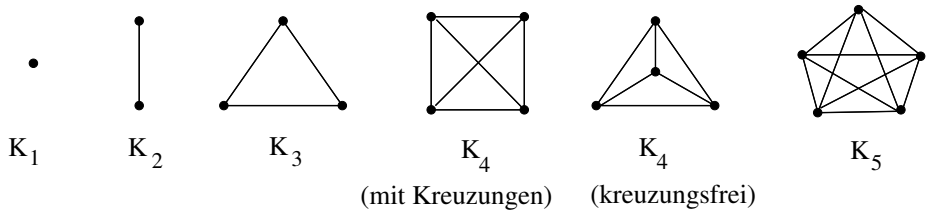
Satz (Handschlaglemma): Für jeden Graph $G = (V, E)$ gilt $\sum_{v \in V} deg(v) = 2|E|$, d.h. $\sum_{v \in V} deg(v)$ ist eine gerade Zahl.
Beweis: Bei Betrachtung der Inzidenzstruktur zwischen Knoten und Kanten ergibt sich die Aussage durch doppeltes Abzählen: Für jeden Knoten v ist die Anzahl inzidenter Kanten $deg(v)$ und jede Kante ist zu ihren zwei Eckknoten inzident. Damit erhalten wir $\sum_{v \in V} deg(v) = \sum_{e \in E} 2 = 2|E|$.

Folgerung: Die Anzahl der Knoten mit ungeradem Grad ist in jedem Graphen gerade.

Definition: Seien $G = (V, E)$ und $G' = (V', E')$ zwei Graphen. Eine Abbildung $\varphi : V \longrightarrow V'$ wird *Graphhomomorphismus* genannt, falls für alle Kanten $\{u, v\} \in E$ auch $\{\varphi(u), \varphi(v)\} \in E'$ gilt. Ist darüber hinaus φ eine bijektive Abbildung und φ^{-1} auch ein Graphhomomorphismus, so nennt man φ einen *Graphisomorphismus* (und G, G' zueinander *isomorph*).

Die folgenden Standardbeispiele beschreiben formal gesehen nicht einzelne Graphen, sondern Isomorphieklassen.

1. Mit K_n ($n \geq 1$) bezeichnet man den *vollständigen Graphen* der Ordnung n, d.h. eine Knotenmenge V mit $|V| = n$ und der vollen Kantenmenge $\binom{V}{2}$.

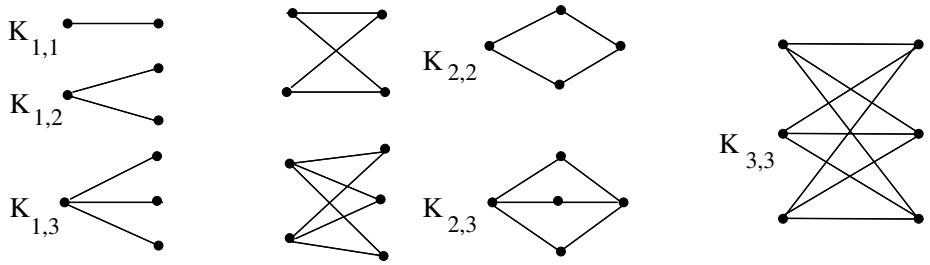


   Die vollständigen Graphen K_1, K_2, K_3 und K_4 sind planar, alle weiteren nicht. Dabei hat K_5 die Eigenschaft, dass er durch Streichung einer beliebigen Kante planar wird.

2. Mit C_n ($n \geq 3$) bezeichnet man den *Kreis* der Länge n, d.h. eine Knotenmenge $V = \{v_1, v_2, \ldots, v_n\}$ mit der Kantenmenge $E = \{\{v_1, v_2\}, \ldots, \{v_{n-1}, v_n\}, \{v_n, v_1\}\}$.

3. Mit Q_n ($n \geq 1$) bezeichnet man den *n-dimensionalen Würfel* mit Knotenmenge $\{0,1\}^n$ (Menge aller n-Tupel über $\{0,1\}$), wobei zwei Tupel dann und nur dann adjazent sind, wenn sie sich an genau einer Stelle unterscheiden.
   Die Bestimmung der Anzahl der Kanten des Q_n ist eine schöne Anwendung des Handschlaglemmas. Wie man leicht sieht, hat jeder Knoten in Q_n den Grad n und folglich ist die Gradsumme aller Knoten gleich $n \cdot 2^n$. Damit ist $2|E| = n \cdot 2^n$, also $|E| = n \cdot 2^{n-1}$.

4. Mit $K_{n,m}$ ($n, m \geq 0$) bezeichnet man den *vollständigen, bipartiten Graphen*, dessen Eckenmenge V die disjunkte Vereinigung von zwei Mengen A und B mit $|A| = n,\ |B| = m$ ist und dessen Kantenmenge aus allen Paaren $\{a, b\}$ mit $a \in A,\ b \in B$ besteht.
   Die Graphen $K_{1,m}$ und $K_{2,m}$ sind planar. Dagegen ist der $K_{3,3}$ und alle $K_{n,m}$ mit $n, m \geq 3$ nicht planar. Der $K_{3,3}$ wird planar, wenn man eine beliebige Kante streicht.



Definition: Man nennt $G' = (V', E')$ einen *Untergraph* von $G = (V, E)$, wenn $V' \subseteq V$ und $E' \subseteq E$ gilt. Ist außerdem $E' = E \cap \binom{V}{2}$, so wird G' als *induzierter Untergraph* von G bezeichnet.

Definition: Ein Graph $G = (V, E)$ wird *bipartit* genannt, wenn er Untergraph eines vollständigen, bipartiten Graphen ist. Etwas anschaulicher kann man formulieren, dass ein Graph genau dann bipartit ist, wenn man die Knoten so mit zwei Farben einfärben kann, dass keine gleichfarbigen Ecken benachbart sind.

Definition: Das Komplement eines Graphen $G = (V, E)$ ist der Graph $\overline{G} = (V, \binom{V}{2} \setminus E)$.

Definition: Eine Folge von paarweise verschiedenen Ecken $v_1, v_2, \ldots, v_k$ eines Graphen $G = (V, E)$ repräsentiert einen *Weg der Länge* $k-1$, falls $\{v_i, v_{i+1}\} \in E$ für alle $1 \leq i < k$. Ist außerdem $\{v_k, v_1\} \in E$, so repräsentiert die Folge auch einen *Kreis der Länge* k. Man sagt, dass v von u *erreichbar* ist, falls ein Weg $v_1, v_2, \ldots, v_k$ mit $v_1 = u$ und $v_k = v$ in G existiert.

Lemma: Die Relation der Erreichbarkeit in der Knotenmenge V eines Graphen ist eine Äquivalenzrelation.

Definition: Die Äquivalenzklassen der Erreichbarkeitsrelation nennt man die *Zusammenhangskomponenten* (kurz *Komponenten*) des Graphen. G wird zusammenhängend

genannt, wenn er genau eine Komponente hat.

Stellt man G durch eine Zeichnung dar, so kann man diesen Begriff anschaulich erklären: Zwei Knoten gehören zur selben Komponente, wenn man sie durch einen geschlossenen Kantenzug verbinden kann, wobei die Kanten nur in Knoten und nicht auf Kantenschnitten in der Zeichnung gewechselt werden dürfen.

Satz: Sind zwei Graphen isomorph, so sind jeweils die Ordnung, die Größe, die sortierten Gradfolgen und die Anzahl der Komponenten der beiden Graphen gleich.

Definition: Seien u, v Knoten in einem ungerichteten Graphen $G = (V, E)$. Sind u und v in einer gemeinsamen Zusammenhangskomponente von G, so definieren wir ihren *Abstand* $d_G(u, v)$ (oder kurz $d(u, v)$) als Länge eines kürzesten Weges (Anzahl der Kanten des Wegs) von u nach v. Gehören sie zu verschiedenen Komponenten, so setzen wir $d(u, v) = \infty$.

Definition: Der *Durchmesser* $D(G)$ des Graphen ist definiert als das Maximum über alle paarweisen Abstände zwischen Knoten.

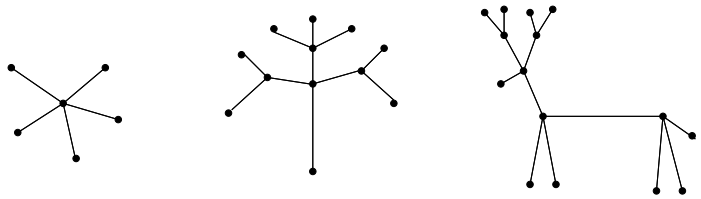
Satz: Ein Graph ist genau dann bipartit, wenn alle in ihm als Untergraph enthaltenen Kreise gerade Länge haben.

Beweis: Zunächst überlegt man sich, dass wir den Graphen als zusammenhängend voraussetzen können, ansonsten führt man den folgenden Beweis für jede Zusammenhangskomponente.
Sei $G = (V, E)$ bipartit, das heißt, $V = A \cup B$ mit $A \cap B = \emptyset$ und Kanten verlaufen nur zwischen Knoten aus A und Knoten aus B, und sei C ein Kreis in G. Dann liegen auf C alternierend Knoten aus A und B und somit haat C gerade Länge.

Zum Beweis der anderen Richtung fixieren wir einen beliebigen Knoten $u \in V$. Wir definieren: $A = \{v \in V |\ d(u, v) \text{ gerade}\}$, $B = V \setminus A$ Wir müssen zeigen, dass es keine Kanten zwischen Knoten aus A (bzw. zwischen Knoten aus B) gibt und führen einen indirekten Beweis indem aus der Annahme, dass es eine Kante $\{v, w\}$ mit $v, w \in B$ (für A analog) gibt, ein Widerspruch abgeleitet wird zur Voraussetzung, dass alle Kreise gerade Länge haben. Wir betrachten kürzeste Wege von u zu v und zu w. Diese Wege müssen die gleiche Länge haben (wegen der Kante zwischen v und w). Sei x der letzte gemeinsame Knoten auf beiden Wegen. Dann bilden die beiden Wegabschnitte von x nach v bzw. nach w zusammen mit der Kante $\{v, w\}$ einen Kreis ungerader Länge.

5.2 Bäume

Definition: Ein Graph G heißt ein *Baum*, wenn er zusammenhängend ist und keine Kreise enthält. Ein Graph, dessen Komponenten jeweils Bäume sind, wird ein Wald genannt. Ein Graph ist also genau dann ein Wald, wen er keine Kreise enthält.



Die Abbildung zeigt drei Bäume, die zusammen einen Wald bilden.

Satz: Sei $G = (V, E)$ ein Graph, dann sind die folgenden drei Bedingungen äquivalent:

1. G ist ein Baum.
2. Je zwei Ecken von G sind durch genau einen Weg verbunden.
3. G ist zusammenhängend und $|E| = |V| - 1$.

Beweis: (1) $\Rightarrow$ (2) und (2) $\Rightarrow$ (1) sind einfache indirekte Schlüsse. Die erste Implikation sieht man z.B. wie folgt. Angenommen es gibt zwei Knoten u, v, zwischen denen nicht genau ein Weg verläuft. Dies könnte kein Weg sein, dann ist der Graph nicht zusammenhängend, oder mindestens zwei Wege, dann entsteht ein Kreis. In jedem Fall ist aber G kein Baum.

Wir zeigen (1) $\Rightarrow$ (3):
Zunächst hat jeder Baum, der nicht nur ein einzelner Knoten ist, Knoten vom Grad 1, diese nennt man *Blätter*. Das sieht man wie folgt. Seien $u_1, u_2, \ldots, u_i$ die Knoten eines längsten Weges in G. Alle Nachbarn von u_1 liegen auf diesem Weg, sonst wäre er nicht längster Weg. Nur u_2 kann Nachbar sein, sonst gäbe es einen Kreis.
Wir entfernen u_1 und die Kante $\{u_1, u_2\}$ aus G und erhalten einen zusammenhängenden Restgraphen G'. Dieser hat genau einen Knoten und eine Kante weniger als G. Wenn wir dies iterieren, bleibt zum Schluss genau ein Knoten ohne Kanten übrig. Also $|E| = |V| - 1$.

(3) $\Rightarrow$ (1): Sei $T = (V, E')$ aufspannender Baum von G, damit $|V| - |E'| = 1$. Aber nach Voraussetzung gilt auch $|V| - |E| = 1$. Allerdings ist $E' \subseteq E$ und die einzige Möglichkeit hierfür ist $E = E'$.

Definition: Sei $G = (V, E)$ ein zusammenhängender Graph. Ein *aufspannender Baum* von G ist ein Untergraph mit gleicher Knotenmenge, der ein Baum ist. Analog ist ein *aufspannender Wald* eines beliebigen (auch unzusammenhängenden) Graphen G ein Untergraph mit den gleichen Komponenten wie G, der ein Wald ist.

Für den Informatiker sind Bäume ein sehr wichtiges Werkzeug. Dabei stellt die Tatsache, dass sich eine Reihe praktischer Problemstellungen auf die Konstruktion von aufspannenden Bäumen (bzw. Wäldern) zurückführen lassen, nur einen Teilaspekt des gesamten Anwendungsspektrums dar. In der theoretischen Informatik dienen sogenannte Entscheidungsbäume als Modell für Berechnungen. Schließlich spielen Bäume als Datenstruktur für Such- und Sortierprozesse eine wichtige Rolle.

Oft ist es nützlich, einen Baum als *gerichteten Graphen* anzusehen, d.h. es werden gerichtete Kanten (Formalisierung durch geordnete Paare) betrachtet. Ist $G = (V, E)$

ein Baum und $r \in V$ ein beliebiger, aber fixierter Knoten, den wir die *Wurzel (Root)* nennen werden, so kann man die Kanten von G in eindeutiger Weise so richten, dass für alle Ecken $v \neq r$ die Kanten auf dem (eindeutigen!) Weg von v nach s zur Wurzel hin gerichtet werden.

Es gibt zwei Verfahren, die Breitensuche (BFS - Breadth-first search) und die Tiefensuche (DFS -Depth-first search), die für einen gegebenen zusammenhängenden Graphen $G = (V, E)$ und einen gegebenen Knoten $r \in V$ einen aufspannenden Baum von G mit der Wurzel r ausgeben. Beide Verfahren durchsuchen G beginnend von r und unterscheiden sich nur durch die für die Zwischenspeicherung verwendete Datenstruktur, durch die festgelegt wird, von welcher bereits erreichten Ecke u die Suche fortgesetzt wird: Eine Warteschlange Q (Prinzip: first in first out) bei BFS und ein Kellerspeicher (Stapelprinzip: last in first out) bei DFS. Der zu bestimmende Baum wird durch ein Feld (Array) π repräsentiert, das für jeden Knoten $v \in V$ den ersten Nachfolger von v auf dem gerichteten Weg zur Wurzel r speichert. Für die Wurzel selbst wird der Wert NIL eingetragen.

Zuerst werden alle Knoten weiß (unberührt) markiert. Dann wird r in den Zwischenspeicher (Warteschlange/Kellerspeicher) geladen und grau (berührt) markiert. In jedem weiteren Schritt wird der nächste Knoten u aus dem Zwischenspeicher (also grau) ausgelesen und getestet, ob er noch einen weißen Nachbarn besitzt. Bei negativer Antwort entfernt man u durch schwarze Markierung (erledigt) aus dem Speicher. Hat u einen weißen Nachbarn v, so wird dieser auch gespeichert (grau gefärbt) und $\pi(v) = u$ gesetzt. Dieser Schritt ist damit gleichzusetzen, dass die Kante $\{v, u\}$ in die ungerichtete Version des Baums aufgenommen wird.

Beide Verfahren stoppen, wenn keine grauen Ecken mehr vorhanden sind.

Hier sind die Pseudocodes der beiden Algorithmen:

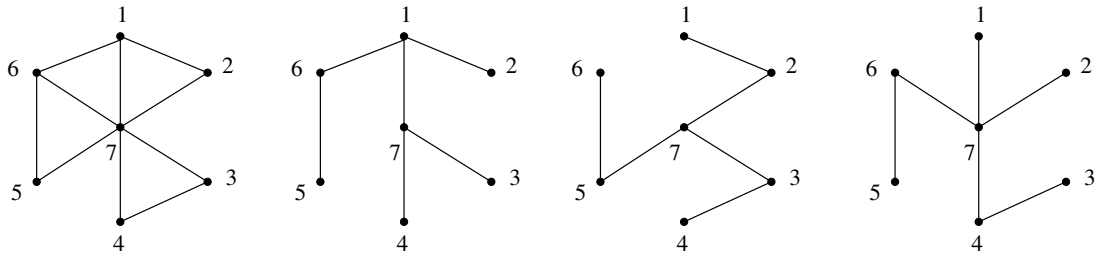
```
BFS(G, r)
for all u ∈ V
    do Farbe[u] ← weiß                          Initialisierung
Farbe[r] ← grau
π[r] ← nil                                      r als Wurzel festlegen
Insert r in Q                                   Wurzel in Q eintragen
while Q≠ ∅                                      sonst abbrechen
    do ū ← Kopf[Q]                              u an erster Stelle in Q
        for all v ∈ Adj[u]
            do if Farbe[v] == weiß              v noch nicht berührt
                then Farbe[v] ← grau            v wird berührt
                     π[v] ← u                   Kante von v zu u
                     Insert v in Q              v in Q eintragen
        Delete Head(Q)                          u aus Q streichen
        Farbe[u] ← schwarz
```

```
DFS(G, r)
for all u ∈ V Farbe[u] ← weiß      Initialisierung
π[r] ← nil                         r als Wurzel festlegen
DFS-visit(r)                       Aufruf einer rekursiven Prozedur
```

```
Procedure DFS-visit(u)
Farbe[u] ← grau
for all v ∈ Adj[u]
    do if Farbe[v] == weiß         v noch nicht berührt
       then π[v] ← u               Kante von v zu u
            DFS-visit(v)           gehe von v weiter in die Tiefe
Farbe[u] ← schwarz
```

Beide Algorithmen können auch zur Erzeugung eines aufspannenden Waldes von unzusammenhängenden Graphen verwendet werden. Dazu ist nur die folgende Änderung notwendig: Hat man einen aufspannenden Baum für eine Komponente erzeugt (keine grauen Punkte mehr), so wählt man, falls noch vorhanden, einen weißen Knoten als Wurzel für eine weitere Komponente aus und startet neu.

Die Ergebnisse der beiden Methoden sind allgemein sehr unterschiedlich: Bei der Tiefensuche DFS werden in der Regel sehr lange Wege erzeugt. Es ist zu beachten, dass die berechneten DFS–Bäume nicht nur von der Wahl der Wurzel r abhängen, sondern auch die Reihenfolge der zu u benachbarten Knoten in der Adjazenzliste $Adj[u]$ einen Einfluss auf das Ergebnis hat. So gibt es für Graphen mit einem Hamiltonkreis bei geeigneter Ordnung der Adjazenzlisten immer einen DFS-Baum der Tiefe $|V|-1$ (also ein Weg). Eine Änderung dieser Ordnung kann (eventuell) zu DFS-Bäumen geringerer Tiefe führen. Das folgende Beispiel zeigt den BFS– und den DFS–Baum eines Graphen, bei Start im Knoten 1 und aufsteigend geordneten Adjazenzlisten. Ganz rechts sieht man einen anderen DFS–Baum, der bei gleichem Startknoten und absteigend geordneten Adjazenzlisten entsteht.



Obwohl auch das Ergebnis der Breitensuche von der Ordnung der Adjazenzlisten abhängt, haben für eine festgelegte Wurzel r alle möglichen BFS–Bäume eine gemeinsame Eigenschaft, die im folgenden Lemma formuliert wird.

Lemma: In einem BFS–Baum T ist der Abstand eines beliebigen Knotens $v \in V$ zur Wurzel r genau so groß wie im ursprünglichen Graphen G, kurz geschrieben:

$$d_T(v, r) = d_G(v, r).$$

Man kann dieses Lemma mit vollständiger Induktion über $d_G(v, r)$ beweisen.

In vielen Anwendungen sind die Kanten eines Graphen $G = (V, E)$ mit positiven Zahlen gewichtet, d.h. zusätzlich ist eine Kostenfunktion $w : E \longrightarrow \mathbb{R}^+$ gegeben. Nun besteht die Aufgabe darin, einen aufspannenden Baum (bzw. Wald) zu bestimmen, so dass das Gesamtgewicht der Kanten des Baums (Waldes) minimal ist (Minimum Spanning Tree - MST). Algorithmen zur Lösung dieses Problems werden in der Vorlesung ALP 3 besprochen.

5.3 Matchings

In vielen Anwendungsaufgaben, die sich mit Graphen modellieren lassen, geht es darum, möglichst große Paarungen von adjazenten Knoten zu bilden.
Beispiel: Wir betrachten eine Menge $P = \{p_1, p_2, \ldots, p_n\}$ von Personen und eine Menge $S = \{s_1, s_2, \ldots, s_m\}$ von Jobs und beschreiben durch einen bipartiten Graphen, welche Personen für welche Jobs geeignet sind. Ziel ist es, eine maximale Anzahl von Personen mit Jobs zu versorgen. Wir suchen also eine maximale Anzahl von Kanten M, so dass jedes p_i und jedes s_j zu höchstens einer Kante aus M inzident sind, denn jeder Job kann nur einmal vergeben werden und jede Person kann nur einen Job übernehmen. Diese Aufgabenstellung ist ein typisches Beispiel für ein Matching–Problem.

Definition: Sei $G = (V, E)$ ein Graph. Eine Untermenge $M \subseteq E$ wird *Matching* von G genannt, wenn jeder Knoten $v \in V$ zu höchstens einer Kante aus M inzident ist. Mit $m(G)$ bezeichnen wir die maximale Größe eines Matchings von G und nennen M ein *Maximum–Matching*, wenn $|M| = m(G)$.

Definition: Ist M ein Matching von $G = (V, E)$, dann nennt man einen Knoten *M–saturiert* (bzw.*M–unsaturiert*), wenn er zu einer (bzw. zu keiner) Kante $e \in M$ inzident ist. Ein Weg p in G wird *M–augmentierend* genannt, wenn er mit M–unsaturierten Knoten beginnt und endet und wenn sich auf dem Weg Kanten aus M und aus $E \setminus M$ gegenseitig ablösen (alternieren).

Lemma: Ist M ein Matching von G und ist p ein M–augmentierender Weg, dann ist M **kein** Maximum–Matching.

Beweis: Tauscht man alle Kanten aus M, die auf p liegen gegen die Nichtmatching–Kanten auf p aus, so erhält man ein neues Matching M' mit $|M'| = |M| + 1$. Damit erklärt sich auch der Begriff augmentierend (= erweiternd).

Wir betrachten zunächst Matchings in bipartiten Graphen. Für eine Teilmenge $X \subseteq V$ bezeichne $N(X)$ die Menge aller zu X benachbarten Knoten, d.h. $N(X) = \bigcup_{v \in X} N(v)$.

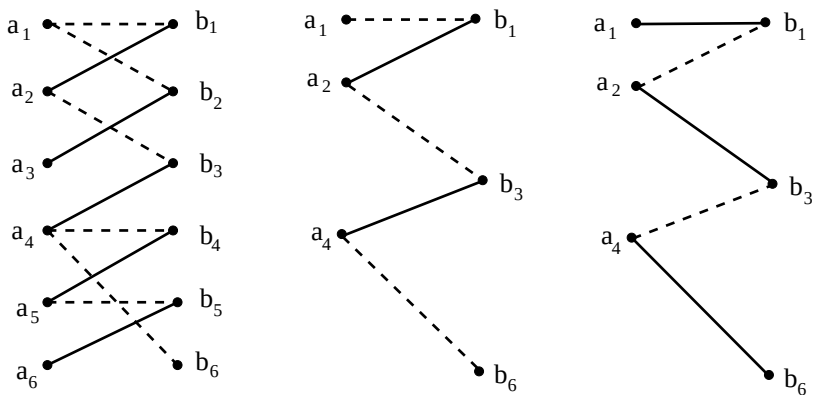
Heiratssatz: Sei $G = (A \cup B, E)$ ein bipartiter Graph. Dann ist $m(G) = |A|$ genau dann, wenn für jede Teilmenge $X \subseteq A$ die Ungleichung $|X| \leq |N(X)|$ gilt.

Beweis: Ist $m(G) = |A|$, dann können alle Knoten aus A gematcht werden und jeder Teilmenge $X \subseteq A$ steht mindestens die Menge der entsprechenden Matchingpartner aus B gegenüber.
Nun gehen wir von $|X| \leq |N(X)|$ für alle $X \subseteq A$ aus, betrachten ein Maximum–Matching M und müssen $|M| = |A|$ zeigen. Das beweisen wir mit Widerspruch, indem aus der Annnahme $|M| < |A|$ die Existenz eines M–augmentierenden Wegs p abgeleitet wird (Widerspruch zur Maximalität von M):

Angenommen $|M|$ ist kleiner als $|A|$, dann gibt es ein M–unsaturiertes $a_1 \in A$, das aber wegen $1 = |\{a_1\}| \leq |N(\{a_1\})|$ mindestens einen Nachbarn $b_1 \in B$ hat. Wäre b_1 M–unsaturiert, dann bildet bereits die Kante $\{a_1, b_1\}$ einen M–augmentierenden Weg.

Anderenfalls ist b_1 M–saturiert und folglich gibt es eine Kante $\{a_2, b_1\} \in M$. Wegen $2 = |\{a_1, a_2\}| \leq |N(\{a_1, a_2\})|$ gibt es ein von b_1 verschiedenes $b_2 \in N(\{a_1, a_2\})$. Ist b_2 M–saturiert, dann gibt es eine Kante $\{a_3, b_2\} \in M$ und man kann wegen $3 = |\{a_1, a_2, a_3\}| \leq |N(\{a_1, a_2, a_3\})|$ ein weiteres $b_3 \in N(\{a_1, a_2, a_3\})$ finden, u.s.w.
Da G endlich ist, findet man irgendwann ein M–unsaturiertes $b_k \in N(\{a_1, a_2, \ldots, a_k\})$ und beginnt den Aufbau des M–augmentierenden Wegs p von dort. Sei $i_1 \leq k$ der kleinste Index, so dass $\{a_{i_1}, b_k\} \in E$ ist (diese Kante ist nicht aus M!). Im Fall $i_1 = 1$, ist der Weg $p = b_k, a_1$ bereits M–augmentierend. Andernfalls setzen wir mit der Kante $\{a_{i_1}, b_{i_1-1}\}$ aus M fort und finden den kleinsten Index i_2, so dass $\{a_{i_2}, b_{i_1-1}\} \in E$ ist. Diese Kante ist nicht aus M und $i_2 \leq i_1 - 1$. Wiederum ist man im Fall $i_2 = 1$ bereits fertig. Andernfalls wird dieser Schritt wiederholt, bis man bei $i_l = 1$ stoppen kann. Wegen $k \geq i_1 > i_2 > \ldots$ muss dieser Prozess terminieren und damit ist die Existenz eines M–augmentierenden Wegs bewiesen.

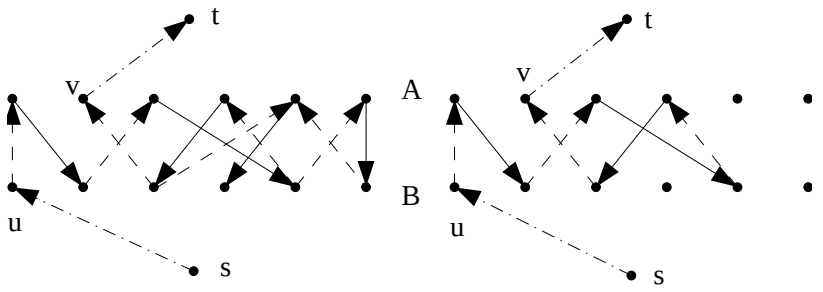
Die folgende Abbildung illustriert diesen Beweis. Auf der linken Seite findet man die Grundkonstruktion, wobei die Knoten in der Reihenfolge $a_1, b_1, a_2, b_2, a_3, \ldots, b_6$ bestimmt werden. Die durchgezeichneten Kanten sind aus M, die gestrichelten aus $E \setminus M$. In der Mitte ist der M–augmentierende Weg dargestellt. Durch Austausch von Matchingkanten gegen Nichtmatchingkanten (rechts) wird M vergrößert.



Die Verwendung von M–augmentierenden Wegen zur schrittweisen Vergrößerung von Matchings kann man zur Berechnung von Maximum–Matchings in Graphen verwenden. Man beginnt mit einem beliebigen Matching M ($M = \emptyset$ geht immer) und sucht einen M–alternierenden Weg. Solange ein solcher Weg existiert, kann man durch Austausch von Matching–Kanten gegen Nichtmatching–Kanten das Matching vergrößern (augmentieren) und rekursiv fortfahren. Der folgende Satz garantiert, dass man im anderen Fall fertig ist.

Satz: Sei M ein beliebiges Matching und M' ein Maximum–Matching von einem Graphen $G = (V, E)$. Ist $|M| < |M'|$, dann gibt es einen M– augmentierenden Weg in G.

Beweis: Wir betrachten den Graphen $H = (V, F)$, wobei F die symmetrische Differenz von M und M' ist (das sind alle Kanten, die entweder in M oder in M' liegen). Da die Knoten in H nur vom Grad $0, 1$ oder 2 sein können, sind die Zusammenhangskomponenten von H isolierte Punkte, gerade Kreise mit M/M'–alternierenden Kanten oder Wege mit M/M'–alternierenden Kanten. Wegen $|M| < |M'|$ muss mindestens einer dieser Wege mehr Kanten aus M' als aus M enthalten und ist folglich M–augmentierend.

Leider sind effiziente Suchverfahren nach M–augmentierenden Wegen im allgemeinen sehr kompliziert, wobei ungerade Kreise das Haupthindernis darstellen. In bipartiten Graphen $G = (A \cup B, E)$ kann man das Problem auf die folgende Art und Weise lösen: Wir führen zwei zusätzliche Knoten s, t ein und bauen einen gerichteten Graphen G_M auf, indem alle Kanten aus M von A nach B und alle Kanten aus $E \setminus M$ von B nach A gerichtet werden. Zusätzlich wird für alle M–unsaturierten Knoten $v \in A$ (bzw. $u \in B$) eine Kante von v nach t (bzw. von s nach u) eingefügt. Offensichtlich existiert ein M–augmentierender Weg genau dann, wenn ein gerichteter Weg von s nach t in G_M existiert. Ein solcher Weg kann duch Tiefensuche (gerichtete Variante) gefunden werden.

Die folgende Abbildung illustriert den vorgestellten Algorithmus. Auf der linken Seite ist die Konstruktion des gerichteten Graphen dargestellt. Der gerichtete Weg von s nach t auf der rechten Seite liefert unmittelbar einen M–augmentierenden Weg, wenn man die erste (von s ausgehende) Kante und die letzte (zu t führende) Kante streicht.