Mathematik für Informatiker Kombinatorik, Stochastik und Statistik

Vorlesungsmanuskript Sommersemester 2020

Janko Böhm

11. Juli 2025

Inhaltsverzeichnis

Abbildungsverzeichnis

Symbolverzeichnis

0

Einleitung

Wir wollen uns mit den Grundlagen der Kombinatorik, Stochastik und Statistik beschäftigen. Dies sind eng verknüpfte zentrale Teilgebiete der Mathematik, neben Analysis, Zahlentheorie, Algebra, Geometrie, Topologie und Numerik. Während die Stochastik, Statistik und Numerik eher angewandten Charakter haben, werden die anderen genannten Gebiete der reinen Mathematik zugeordnet. Wir wollen zunächst einen kurzen Überblick über diese Teilgebiete und deren Zusammenhänge bekommen:

Beginnen wir mit der **Kombinatorik**, die von allen genannten Bereichen mit den einfachsten Grundstrukturen startet (was aber nicht bedeutet, dass die Kombinatorik einfach wäre): Die Kombinatorik beschäftigt sich mit dem Zählen, basiert also auf den natürlichen Zahlen $\mathbb{N} = \{1, 2, 3, ...\}$. Mit Hilfe der Kombinatorik kann man zum Beispiel berechnen, dass es beim Ziehen der Lottozahlen $\binom{49}{6} \approx 14\,000\,000$ mögliche Ergebnisse gibt. Die Kombinatorik ist also eng mit der Wahrscheinlichkeitstheorie verknüpft, der sogenannten Stochastik: Sind alle Ereignisse beim Lotto gleich wahrscheinlich, dann ist die Wahrscheinlichkeit bei einem Spiel zu gewinnen gleich

$$\frac{1}{\binom{49}{6}} \approx \frac{1}{14\,000\,000}.$$

In der Informatik ist ein Teilgebiet der Kombinatorik besonders wichtig, die Graphentheorie. Graphen werden z.B. verwendet um Netzwerke zu beschreiben. Der Graph in Abbildung 1 beschreibt z.B. auf welche Weise vier Internet-Sites untereinander

verlinkt sind. Solche Graphen werden beispielsweise in Googles Page-Rank-Algorithmus verwendet.

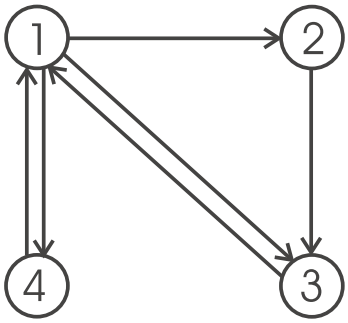


Abbildung 1: Gerichteter Graph von Links zwischen Internetseiten

Es gibt aber mehr ganze Zahlen als nur die in $\mathbb{N}$. Die **Zahlentheorie** untersucht die Eigenschaften der ganzen Zahlen in

$$\mathbb{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\},$$

insbesondere mit der Beziehung zwischen der Verknüpfungen Addition und Multiplikation. Viele zahlentheoretische Probleme können sehr einfach formuliert, aber nur sehr schwer gelöst werden. Das bekannteste Beispiel ist sicherlich Fermats letzter Satz von 1637: Es gibt für $n \geq 3$ keine (nichttriviale) ganzzahlige Lösung der Gleichung

$$x^n + y^n = z^n$$

Fermats letzter Satz wurde erst 1995 (von A. Wiles) bewiesen nach 350-jährigen Vorarbeiten, bei denen viele neue Konzepte in der Mathematik entwickelt wurden. Heute bestehen enge Beziehungen der Zahlentheorie zum Beispiel zur algebraischen Geometrie, Kombinatorik, Kryptographie und Codierungstheorie.

Formeln wie $x^n + y^n - z^n$ sind ein Rezept für die Anwendung von sogenannten Verknüpfungen, wie Addition und Multiplikation von ganzen Zahlen. Eine Gleichung entsteht dann indem wir ein festgelegtes Ergebnis fordern, etwa 0 in der obigen Gleichung $x^n + y^n - z^n = 0$. Die **Algebra** ist ein sehr umfangreiches Gebiet der Mathematik, das sich mit für alle Bereiche der Mathematik grundlegende algebraische Strukturen, wie Gruppen, Ringen und

Körpern beschäftigt, d.h. mit der Frage, wie man auf Mengen Verknüpfungen einführen kann. Die Public-Key Kryptographie verwendet z.B. Ergebnisse aus der Zahlentheorie und der Algebra. Das Konzept des Körpers, einer Menge mit Addition und Multiplikation, sodass es zu jeder Zahl die negative Zahl gibt (etwa -1 zu 1) und zu jeder Zahl $\neq 0$ den Kehrwert (etwa $\frac{1}{2}$ zu 2), spielt eine entscheidende Rolle in Anwendungen: Ein wichtiger Berührungsbereich der Algebra besteht neben der Zahlentheorie mit der algebraischen **Geometrie**. Diese untersucht die Lösungsmengen von polynomialen Gleichungssystemen in mehreren Variablen über einem Körper K (zum Beispiel $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ der Körper der rationalen, reellen oder komplexen Zahlen). Zum Beispiel besteht die gemeinsame Lösungsmenge von $x^2 + 2y^2 = 3$ und $2x^2 + y^2 = 3$, d.h. der Durchschnitt von zwei Ellipsen, aus den 4 Punkten $(1, 1), (-1, 1), (1, -1), (-1, -1)$, siehe Abbildung 2. Bei

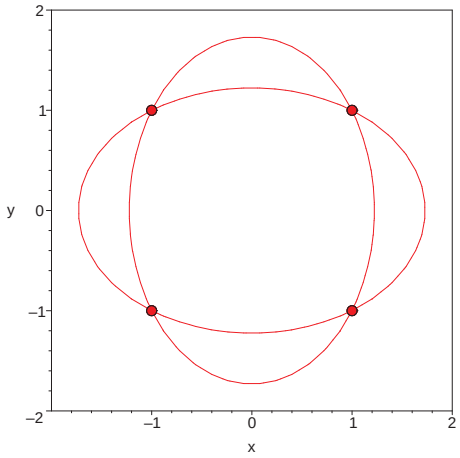


Abbildung 2: Vier Punkte

algebraischer Geometrie über $K = \mathbb{Q}$ kommt wieder die Zahlentheorie ins Spiel.

Der einfachste (aber in der Praxis sehr wichtige) Spezialfall sind lineare Gleichungssysteme über einem Körper K, das

Kernthema der linearen Algebra. Hier lösen wir

$$\begin{aligned} a_{1,1}x_1 + \ldots + a_{1,m}x_m &= b_1 \\ &\vdots \\ a_{n,1}x_1 + \ldots + a_{n,m}x_m &= b_n \end{aligned}$$

mit $a_{ij} \in K$, $b_i \in K$ nach $x_j \in K$ (mit $i = 1, \ldots, n$ und $j = 1, \ldots m$). Der Gaußalgorithmus zum Lösen von linearen Gleichungssystemen ist sicher einer der wichtigsten Algorithmen überhaupt, denn lineare Gleichungssysteme treten in fast allen Anwendungen der Mathematik auf, von der Wettervorhersage bis zur Berechnung eines Page-Ranking für Suchmaschinen aus dem Link-Graphen wie in Abbildung 1.

In der **Topologie** untersucht man Eigenschaften von Objekten, die sich unter stetigen Verformungen nicht ändern. Man sieht etwa, dass sich der Knoten in Abbildung 3 nicht ohne Aufschneiden entwirren läßt. Will man alle möglichen Wege von einer Internetseite zu einer anderen in einem Linkgraphen untersuchen, wird man auch Methoden der Topologie verwenden.

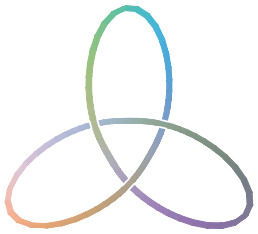


Abbildung 3: Knoten

Ein zentrales Konzept in der **Analysis** wiederum ist die Stetigkeit von Abbildungen. Eine stetige Abbildung ist nichts anderes als eine stetige Verformung einer Geraden in den Graphen einer Funktion. Die Funktion mit dem Graphen in Abbildung 4 ist z.B. stetig, die in Abbildung 5 nicht. Der Begriff der Stetigkeit spielt eine wichtige Rolle in der Analysis und algebraischen Geometrie. Ändert eine stetige Funktion ihr Vorzeichen, dann ist es nicht schwer zu beweisen, dass diese Funktion eine Nullstelle hat, und dieser Beweis liefert tatsächlich einenn Algorithmus zur Bestimmung der Nullstelle. Eine unstetige Funktion kann dagegen den Funktionswert 0 überspringen und muss damit nicht

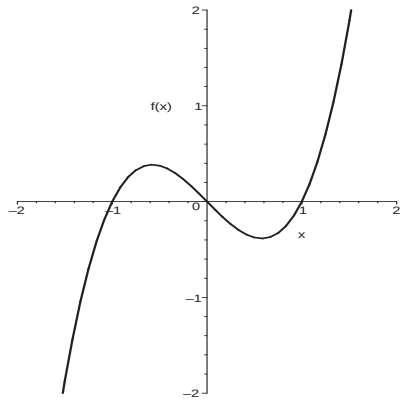


Abbildung 4: Eine stetige Funktion

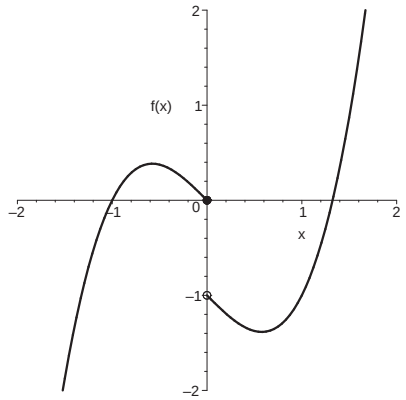


Abbildung 5: Eine unstetige Funktion

unbedingt eine Nullstelle besitzen. Die moderne Analysis geht auf die Infinitesimalrechnung zurück, die von Leibniz und Newton entwickelt wurde. Im Wesentlichen geht es darum, neben dem Stetigkeitsbegriff, den wir schon gesehen haben, auch einen Begriff der Steigung $f'(x)$ einer Funktion $f(x)$ zu entwickeln, indem man die Tangente (Abbildung 6) an einem gegebenen Punkt durch Sekanten (Abbildung 7) approximiert. Dabei verwenden wir, dass zwei verschiedene Punkte in der (x, y)-Ebene eindeutig eine Gerade festlegen. Für die Tangente würden wir gerne beide Punkte gleichsetzen. Durch einen einzelnen Punkt ist aber keine eindeutige Gerade mehr festgelegt. Wir können den zwei-

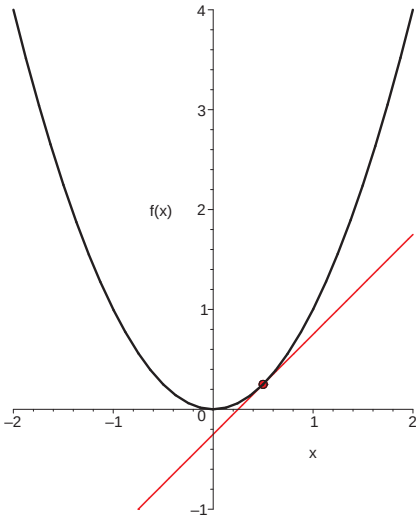


Abbildung 6: Die Tangente an $f(x) = x^2$ in $x = \frac{1}{2}$

ten Punkt also nur beliebig nahe an den Punkt heranführen, an dem wir die Tangente bestimmen wollen. Liefert dieser sogenannte Grenzwertprozess ein eindeutiges Ergebnis (unabhängig davon wie sich der zweite Punkt annähert), dann existiert eine Tangente und die Funktion heißt differenzierbar[1].

Die Funktion in Abbildung 8 hat in $x = 0$ dagegen offenbar keine vernünftige Tangente, denn wenn sich der zweite Punkt von links bzw. von rechts dem Punkt $(0, 0)$ nähert, erhalten wir unterschiedliche Grenzwerte der Sekantensteigung.

Gegeben eine Funktion $x \mapsto f(x)$ stellt sich natürlich die Frage, ob $x \mapsto f'(x)$ wieder eine Funktion ist, wo sie definiert ist und welche Eigenschaften sie hat. Solche Fragen beantwortet die

[1]Konkret haben wir für $f(x) = x^2$ die Steigung $f'(\frac{1}{2}) = 1$ der Tangente in $x = \frac{1}{2}$. Um diesen Wert zu bestimmen, gehen wir von der Steigung der Sekante durch die Punkte $(\frac{1}{2}, \frac{1}{4})$ und (x, x^2) aus. Deren Steigung ist gegeben durch

$$\frac{f(x) - f(\frac{1}{2})}{x - \frac{1}{2}} = \frac{x^2 - \left(\frac{1}{2}\right)^2}{x - \frac{1}{2}} = x + \frac{1}{2}.$$

Nähert sich nun x dem Wert $\frac{1}{2}$ an, dann nähert sich die Sekante der Tangenten an, und somit die Sekantensteigung der Tangentensteigung. Für $x \to \frac{1}{2}$ erhalten wir $f'(\frac{1}{2}) = \frac{1}{2} + \frac{1}{2} = 1$.

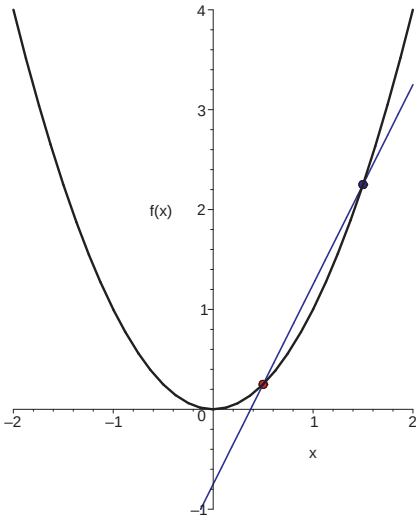


Abbildung 7: Eine Sekante an $f(x) = x^2$ in $x = \frac{1}{2}$

Differentialrechnung. Umgekehrt kann f' gegeben sein und man will f finden. Dies ist ein Problem der Integralrechnung.

Eine zentrale Anwendung von Ableitungen ist die Bestimmung von lokalen Extremwerten: Hat f bei $x = a$ ein lokales Minimum oder Maximum, dann ist

$$f'(a) = 0.$$

Für $f(x) = x^3 - x$ (siehe Abbildung 4) ist $f'(x) = 3x^2 - 1$, also sind $a = \pm\frac{1}{\sqrt{3}}$ Kandidaten. Tatsächlich liegt bei $x = -\frac{1}{\sqrt{3}}$ ein lokales Maximum und bei $x = \frac{1}{\sqrt{3}}$ ein lokales Minimum, wie man mit einem hinreichenden Kriterium sehen kann.

Die ursprüngliche Motivation für die Entwicklung der Analysis war das Newtonsche Kraftgesetz. Die Bewegung einer Masse m an einer Feder (siehe Abbildung 9) wird beschrieben durch die Gleichung

$$m \cdot x''(t) = -c \cdot x(t)$$

zwischen der Position $x(t)$ und der zweiten Ableitung $x''(t)$. Die Rückstellkraft der Feder ist dabei direkt proportional zu der Auslenkung $x(t)$ der Feder (mit Proportionalitätskonstante $c > 0$) und führt zu der Beschleunigung $x''(t)$ der Masse $m > 0$. Man spricht von einem sogenannten harmonischen Oszillator. Eine

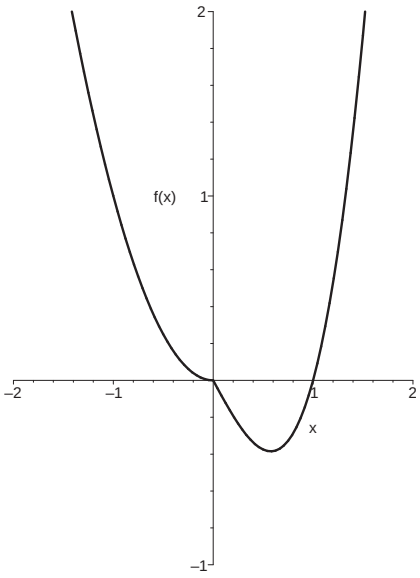


Abbildung 8: Eine Funktion die in $x = 0$ keine Tangente besitzt

Gleichung dieser Form bezeichnet man auch als Differentialgleichung für die Funktion $x(t)$. Eine mögliche Lösung ist

$$x(t) = \sin(\sqrt{\frac{c}{m}} \cdot t),$$

siehe Abbildung 10, denn

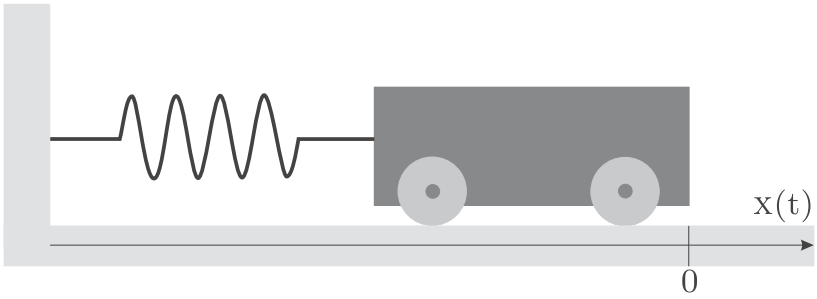


Abbildung 9: Harmonischer Oszillator

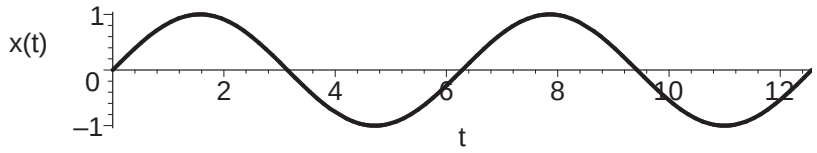


Abbildung 10: Eine Lösung für den harmonischen Oszillator

$$x'(t) = \sqrt{\frac{c}{m}}\cos(\sqrt{\frac{c}{m}} \cdot t)$$
$$x''(t) = -\frac{c}{m}\sin(\sqrt{\frac{c}{m}} \cdot t).$$

Mit den Methoden der Analysis kann man die Menge aller möglichen Lösungen der Differentialgleichung beschreiben.

Es ist oft schwer (oder sogar unmöglich) die Lösungen einer Differenzialgleichung wie oben durch explizite Funktionsausdrücke anzugeben. Das Gleiche gilt schon für die Nullstellen von Gleichungen. Die **Numerik** versucht für Probleme der reinen Mathematik, die mit Hilfe von reellen oder komplexen Zahlen formuliert werden, approximative Lösungen zu finden. Musterbeispiele sind das Lösen von nichtlinearen Gleichungssystemen oder Differenzialgleichung. Eine der wichtigsten Anwendungsfälle in der Numerik ist aber schon das Lösen von linearen Gleichungssystemen. Eine numerische Lösung kann hier oft schneller berechnet werden als ein exakter algebraischer Lösungsausdruck. Nichtlineare Probleme werden oft durch lineare approximiert. Beispielsweise kann man approximativ eine Nullstelle einer Funktion $x \mapsto f(x)$ mit Hilfe des Newtonverfahrens bestimmen: Ausgehend von einem Startwert x_1 berechnet man iterativ

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

d.h. x_{n+1} ist die Nullstelle der Tangente von

$$f(x) = x^2 - 2$$

in x_n, siehe Abbildung 11. Wie wir sehen werden, sind solche Approximationsverfahren eng verknüpft mit der Frage, was eine reelle Zahl wie etwa die Nullstelle $\sqrt{2}$ der obigen Funktion überhaupt ist. Es ist nicht schwer zu beweisen, dass sich $\sqrt{2}$ nicht als

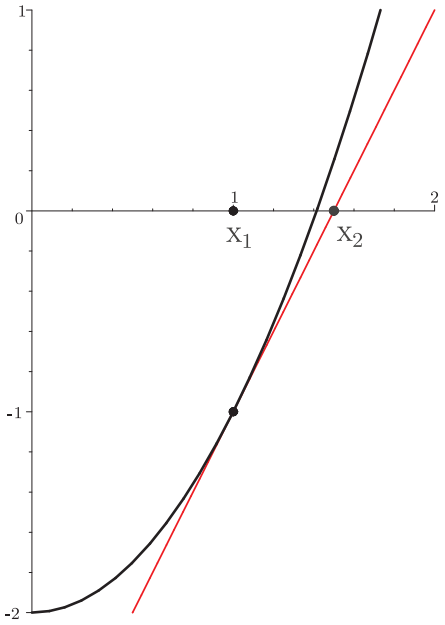


Abbildung 11: Newtonverfahren

Bruch von ganzen Zahlen darstellen lässt, also keine rationale Zahl ist. Andererseits nimmt man in der Praxis in der Numerik den pragmatischen Standpunkt ein, dass wir sowieso oft nur eine Näherung der Nullstelle finden können. Deshalb werden auf dem Computer solche Rechnungen üblicherweise mit Fließkommazahlen durchgeführt, d.h. wir schreiben etwa

$$\sqrt{2} \approx 1.414213562.$$

In diesem Sinne ist dann $\sqrt{2}$ nicht anderes als jede andere Zahl, etwa

$$\frac{2}{3} \approx 0.6666666666.$$

Für praktische Fragestellungen enorm wichtige Anwendungen der Kombinatorik, der Analysis und der linearen Algebra finden sich in der Stochastik und Statistik. Die **Stochastik** (oder auch Wahrscheinlichkeitstheorie) ist die mathematische Sprache zur Quantifizierung von zufälligen Prozessen. Dies reicht von Wurf einer Münze, über die Beschreibung von Lotto über eine

sogenannte Gleichverteilung (d.h. alle Ergebnisse einer Lottoziehung sind gleich wahrscheinlich, siehe oben), bis hin zur Analyse von Algorithmen. Neben der Gleichverteilung ist die bekannteste und wichtigste Zufallsverteilung die Gaußverteilung (oder auch Normalverteilung), die es sogar auf den 10 DM Schein geschafft hat, siehe Abbildung 12. Beispielsweise gehorchen die Körpergrößen von Menschen einer Gaußverteilung. Wie man schon an der Verwendung von Funktionen sieht, baut die Stochastik wesentlich auf den Methoden der Analysis auf.

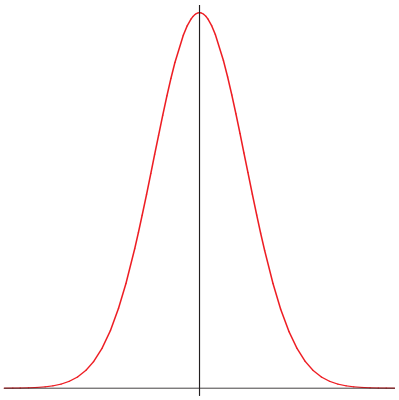


Abbildung 12: Normalverteilung

Die **Statistik** ist ein zur Stochastik eng verwandtes Gebiet und befasst sich mit dem Sammeln und Analysieren von Daten. Während wir also in der Stochastik untersuchen, was wir über die Eigenschaften des Resultats eines gegebenen datenerzeugenden Prozesses sagen können, ist die Kernfragestellung der Statistik das dazu inverse Problem: Gegeben eine Menge an Daten, was können wir über den Prozess sagen, der diese Daten erzeugt hat. Das Gegenstück zur Statistik ist in der Informatik ist das Data Mining und das Machine Learning, wobei hier weniger Gewicht auf die exakte mathematische Beschreibung als vielmehr auf die Effizienz der verwendeten Algorithmen gelegt wird. Ein typisches Problem ist es, in einer Klasse von Funktionen eine Funktion zu finden, deren Funktionsgraph eine gegebene Datenmenge am besten beschreibt. In Abbildung 13 sehen wir die parallele Effizienz eines Computerprogramms, d.h. den Beschleunigungsfaktor geteilt durch die Anzahl der verwendeten Rechenkerne auf einem

High-Performance-Computing Cluster und eine möglichst gute Approximation der Daten durch eine lineare Funktion

$$f(x) = a + b \cdot x$$

mit zwei geeignet gewählten Parametern a und b. Diese Frage-

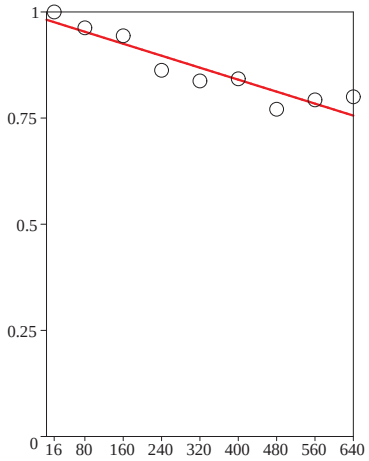


Abbildung 13: Lineare Regression

stellung löst man mit Mitteln der Analysis und linearen Algebra. Betrachten wir ein explizites Beispiel:

Für die Datenpunkte

$$(x, y) = (1, 3), (2, 3), (3, 5)$$

wollen wir eine Ausgleichsgerade $f(x) = a+b\cdot x$ finden. Die Datenpunkte könnten z.B. GPS-Messwerte für ein Fahrzeug auf einer geraden Straße sein. Dazu setzen wir die Datenpunkte in $f(x)$ ein und minimieren die Summe $S(a, b)$ der Quadrate[2] der Differenzen der y-Werte also

$$\begin{aligned} S(a, b) &= (a + b - 3)^2 + (a + 2b - 3)^2 + (a + 3b - 5)^2 \\ &= 12ab - 48b - 22a + 3a^2 + 14b^2 + 43. \end{aligned}$$

[2]Man könnte auch z.B. die Absolutbeträge der Abweichungen addieren, jedoch ist die Betragsfunktion nicht differenzierbar und somit nicht gut geeignet, um mit Hilfe von Ableitungen Extremwerte zu bestimmen.

Wie man in der multivariaten Analysis sieht, ist durch das Verschwinden der partiellen Ableitungen wiederum ein notwendiges Kriterium für ein lokales Minimum gegeben. Wir erhalten also das Gleichungssystem

$$\begin{aligned} 0 &= \frac{\partial S}{\partial a} = 6a + 12b - 22 \\ 0 &= \frac{\partial S}{\partial b} = 12a + 28b - 48. \end{aligned}$$

Da wir eine quadratische Funktion minimieren, erhalten wir durch die Ableitugen sogar ein lineares Gleichungssystem. Dieses ist mit dem Gaußalgorithmus äquivalent zu

$$\begin{aligned} 6a + 12b - 22 &= 0 \\ 4b - 4 &= 0 \end{aligned}$$

und besitzt damit genau die Lösung

$$(a, b) = \left(\frac{5}{3}, 1\right)$$

also ist unsere Ausgleichsgerade

$$f(x) = \frac{5}{3} + x$$

siehe Abbildung 14.

Während ein derartiges univariates lineares Regressionsproblem leicht zu lösen ist, können moderne Algorithmen statistische Fragestellungen mit Millionen von Variablen und Tausenden von Parametern behandeln.

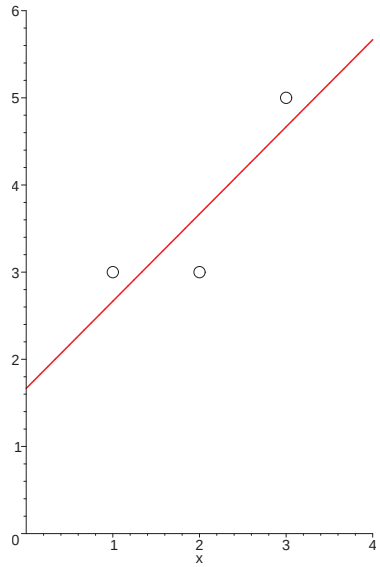


Abbildung 14: Lineare Regression

1

Kombinatorik

1.1 Übersicht

In der Kombinatorik untersucht man endliche oder abzählbar unendliche Strukturen in der Mathematik.

Die abzählende Kombinatorik beschäftigt sich mit der Bestimmung der Anzahl der Elemente von endlichen Mengen. Eine klassische Fragestellung ist: Wieviele Teilmengen hat eine endliche Menge M? Eine derartige Frage haben wir schon im Satz über die Potenzmenge beantwortet:

Definition 1.1.1 *Sei M eine Menge. Die **Potenzmenge** von M ist*

$$2^M = \mathfrak{P}(M) = \{A \mid A \subset M\}.$$

Satz 1.1.2 *Sei M eine endliche Menge. Dann gilt*

$$\left|2^M\right| = 2^{|M|}.$$

Beispiel 1.1.3 *Potenzmengen:*

$$\begin{aligned} 2^{\varnothing} &= \{\varnothing\} \\ 2^{\{1\}} &= \{\varnothing, \{1\}\} \\ 2^{\{1,2\}} &= \{\varnothing, \{1\}, \{2\}, \{1,2\}\}. \end{aligned}$$

Wir beweisen Satz 1.1.2 mit Hilfe von vollständiger Induktion:

Beweis. Durch Nummerieren der Elemente von M können wir ohne Einschränkung der Allgemeinheit annehmen, dass $M = \{1, \ldots, n\}$, wobei wir die Konvention verwenden, dass $\{1, \ldots, 0\} = \varnothing$. Wir müssen also zeigen, dass die Aussage

$$\left|2^{\{1,\ldots,n\}}\right| = 2^n$$

für alle $n \in \mathbb{N}_0$ gilt.
Induktionsanfang $n = 0$: Es ist $2^{\varnothing} = \{\varnothing\}$, also $|2^{\varnothing}| = 1 = 2^0$.
Induktionsschritt $n-1$ nach n: Die Vereinigung

$$\begin{aligned} 2^{\{1,\ldots,n\}} &= \{A \subset \{1, ..., n\} \mid n \notin A\} \dot\cup \\ &\quad \{A \subset \{1, ..., n\} \mid n \in A\} \\ &= \{A \mid A \subset \{1, ..., n-1\}\} \dot\cup \{A' \cup \{n\} \mid A' \subset \{1, ..., n-1\}\} \end{aligned}$$

ist disjunkt, also folgt aus der Induktionsvoraussetzung

$$\left|2^{\{1,\ldots,n-1\}}\right| = 2^{n-1},$$

dass

$$\left|2^{\{1,\ldots,n\}}\right| = 2^{n-1} + 2^{n-1} = 2^n.$$

■

Die abzählende Kombinatorik ist von zentraler Bedeutung für das Design und die Analyse von Algorithmen in der Informatik. Um die Performance oder den Speicherverbrauch eines Algorithmus (z.B. zur Bestimmung von 2^M) abzuschätzen, ist es etwa wichtig zu verstehen, wieviele Schritte er benötigt, um das Ergebnis zu liefern. Eine andere Anwendung liegt in der Stochastik. Zum Beispiel ist (unter der Voraussetzung, dass alle Ergebnisse gleich wahrscheinlich sind) die Gewinnwahrscheinlichkeit beim Lotto

$$\frac{1}{\binom{49}{6}}$$

wobei $\binom{49}{6}$ die Anzahl der möglichen Ergebnisse bezeichnet.

Ein anderer Teilbereich der Kombinatorik ist die Graphentheorie. Graphen wie in Abbildung 1 sind eine der wichtigsten Datenstrukturen in der Informatik. Sie bestehen aus Ecken und Kanten (eventuell mit einer Länge). In einem Graphen (etwa dem Schienennetz der Bahn) will man z.B. herausfinden, welcher Weg der kürzeste zwischen zwei gegebenen Ecken ist.

Viele weitere Teilbereiche der Kombinatorik, die wir hier nicht ansprechen können, sind ebenfalls relevant für die Informatik, etwa Matroide und Designs.

1.2 Binomialkoeffizienten

Die Anzahl der Teilmengen einer n-elementigen Menge haben wir schon bestimmt. Aber wieviele Teilmengen mit einer vorgegebenen Anzahl k von Elementen gibt es?

Definition 1.2.1 *Seien $n, k \in \mathbb{N}_0$. Wir bezeichnen mit $\binom{n}{k}$ die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge.*

Für $k \in \mathbb{Z}$ negativ setzen wir $\binom{n}{k} = 0$.

Beispiel 1.2.2 $\binom{3}{2} = 3$, $\binom{3}{0} = 1$, $\binom{0}{0} = 1$, $\binom{2}{3} = 0$.

Definition 1.2.3 *Ist $k \in \mathbb{N}_0$ und M eine Menge, dann schreiben wir*

$$\binom{M}{k} := \{A \subset M \mid |A| = k\}$$

für die Menge der k-elementigen Teilmengen von M.

Beispiel 1.2.4 $\binom{\{1,2,3\}}{2} = \{\{2,3\},\{1,3\},\{1,2\}\}$.

Zunächst eine grundlegende Symmetrieeigenschaft von Binomialkoeffizienten:

Proposition 1.2.5 *Es gilt $\binom{n}{k} = \binom{n}{n-k}$.*

Beweis. Für $k < 0$ oder $k > n$ sind beide Seiten 0. Anderenfalls sei M eine n-elementige Menge. Die Abbildung

$$\begin{array}{rccc} \alpha: & \binom{M}{k} & \to & \binom{M}{n-k} \\ & U & \mapsto & M \backslash U \end{array}$$

ist bijektiv:

- injektiv: Falls $M\backslash U_1 = M\backslash U_2$ für $U_i \subset M$, dann $U_1 = U_2$.
- surjektiv: Sei $V \in \binom{M}{n-k}$. Es gilt $\alpha(M\backslash V) = M\backslash(M\backslash V) = V$.

■

Der Beweis sagt nichts anderes als, dass die Auswahl von k Elementen aus n die verbleibenden $n-k$ Elemente festlegt und umgekehrt.

Beispiel 1.2.6 *Wir illustrieren den Beweis an einem Beispiel: Die Abbildung*

$$\begin{aligned}\binom{\{1,2,3\}}{1} &\to \binom{\{1,2,3\}}{2}\\ \{1\} &\mapsto \{2,3\}\\ \{2\} &\mapsto \{1,3\}\\ \{3\} &\mapsto \{1,2\}\end{aligned}$$

ist bijektiv, also $\binom{3}{1} = \binom{3}{2}$.

Beispiel 1.2.7 *Beim Lotto-Glücksspiel werden aus einem Topf von* 49 *nummerierten Kugeln* 6 *Kugeln gezogen. Da die Kugeln unterscheidbar sind, ist die Menge der möglichen Lottoergebnisse*

$$\binom{\{1,\ldots,49\}}{6} = \{\{1,2,3,4,5,6\},\ldots\}$$

und die Anzahl der möglichen Ergebnisse die Anzahl der 6*-elementigen Teilmengen einer* 49*-elementigen Menge, d.h.*

$$\binom{49}{6}.$$

Wie groß ist diese Zahl?

Um diese Frage zu beantworten, leiten wir im Folgenden eine geschlossene Formel für $\binom{n}{k}$ her.

Proposition 1.2.8 *Für alle* $n, m, k \in \mathbb{N}_0$ *gilt*

$$(n+1)\binom{n}{k} = (k+1)\binom{n+1}{k+1}$$

Beweis. Sei M eine Menge mit $|M| = n+1$. Die Menge

$$F = \left\{(m,U) \in M \times \binom{M}{k+1} \mid m \in U\right\}$$

können wir anschaulich interpretieren als die Menge aller $(k+1)$-elementigen Teilmengen $U \subset M$, wobei ein $m \in U$ markiert wird. Wir können die Elemente von F auf zwei Weisen abzählen:

- Wähle eine Teilmenge $U \subset M$ mit $|U| = k+1$ und wähle dann ein $m \in U$ aus. Dies zeigt, dass

$$|F| = \binom{n+1}{k+1}(k+1).$$

- Wähle $m \in M$, wähle dann ein $V \subset M \backslash \{m\}$ mit $|V| = k$, und bilde daraus $U = \{m\} \cup V$. Dies zeigt, dass

$$|F| = (n+1)\binom{n}{k}.$$

■

Beispiel 1.2.9 *Wir illustrieren den Beweis an einem Beispiel: Sei $n = 3$ und $k = 2$. Wir können $M = \{1,2,3,4\}$ annehmen. Im Folgenden stellen wir die Elemente $(m,U) \in F$ dar als U mit einer Markierung $m \in U$.*

Wählen wir zunächst $m \in M$ und ergänzen zu einer 3-elementigen Teilmenge von M, so erhalten wir folgende Abzählung der Elemente von F

m	1	2	3	4
	$\{\mathbf{1},2,3\}$	$\{\mathbf{2},1,3\}$	$\{\mathbf{3},1,2\}$	$\{\mathbf{4},1,2\}$
	$\{\mathbf{1},2,4\}$	$\{\mathbf{2},1,4\}$	$\{\mathbf{3},1,4\}$	$\{\mathbf{4},1,3\}$
	$\{\mathbf{1},3,4\}$	$\{\mathbf{2},3,4\}$	$\{\mathbf{3},2,4\}$	$\{\mathbf{4},2,3\}$

mit insgesamt $4 \cdot \binom{3}{2}$ Elementen.

Wählen wir zunächst eine 3-elementige Teilmenge $U \subset M$ und markieren dann ein Element $m \in U$, bekommen wir folgende Abzählung der Elemente von F

U	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$
	$\{\mathbf{1},2,3\}$	$\{\mathbf{1},2,4\}$	$\{\mathbf{1},3,4\}$	$\{\mathbf{2},3,4\}$
	$\{1,\mathbf{2},3\}$	$\{1,\mathbf{2},4\}$	$\{1,\mathbf{3},4\}$	$\{2,\mathbf{3},4\}$
	$\{1,2,\mathbf{3}\}$	$\{1,2,\mathbf{4}\}$	$\{1,3,\mathbf{4}\}$	$\{2,3,\mathbf{4}\}$

mit insgesamt $\binom{4}{3} \cdot 3$ Elementen. Dies zeigt, dass

$$4 \cdot \binom{3}{2} = |F| = \binom{4}{3} \cdot 3.$$

Corollar 1.2.10 *Für $0 \leq k \leq n$ gilt*

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

wobei

$$n! = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot \ldots \cdot n$$

***n*-Fakultät** *bezeichnet.*

Beweis. Induktion nach n:
Induktionsanfang $n = 0$: $\binom{0}{0} = 1$
Induktionsschritt: $n-1 \mapsto n$: Proposition 1.2.8 und die Induktionsvoraussetzung geben

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} = \frac{n}{k}\frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n!}{k!(n-k)!}.$$

■

Beispiel 1.2.11 *Beim Lottospiel gibt es*

$$\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13\,983\,816$$

Möglichkeiten.

Die Binomialkoeffizienten lassen sich auch rekursiv berechnen. Dazu verwenden wir:

Proposition 1.2.12 (Vandermonde Identität) *Für alle $n, m, k \in \mathbb{N}_0$ gilt*

$$\sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$$

Beweis. Seien A und B disjunkte Mengen mit $|A| = n$ und $|B| = m$. Die Anzahl der k-elementigen Teilmengen von $A \cup B$ ist $\binom{n+m}{k}$. Andererseits ist $\binom{n}{j}\binom{m}{k-j}$ die Anzahl der k-elementigen Teilmengen $U \subset A \cup B$ mit $|U \cap A| = j$. ■

Beispiel 1.2.13 *Wir illustrieren den Beweis an einem Beispiel: Sei* $A = \{1,2,3\}$, $B = \{4,5\}$ *und* $k = 4$. *Die 4-elementigen Teilmengen von* $A \cup B$ *sind*

$$\{1,2,3,4\}, \{1,2,3,5\}, \{1,2,4,5\}, \{1,3,4,5\}, \{2,3,4,5\}.$$

Diese sortieren wir nach der Anzahl j *der Elemente aus* A: *Für* $j = 2$ *erhalten wir*

$$\{1,2,4,5\}, \{1,3,4,5\}, \{2,3,4,5\}$$

und für $j = 3$

$$\{1,2,3,4\}, \{1,2,3,5\}.$$

Die Summanden $j = 0, 1, 4$ *leisten keinen Beitrag. Dies zeigt, dass*

$$\binom{5}{4} = \binom{3}{2}\binom{2}{2} + \binom{3}{3}\binom{2}{1}.$$

Corollar 1.2.14 *Für alle* $n, k \in \mathbb{N}_0$ *gilt*

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

Beweis. Proposition 1.2.12 gibt

$$\binom{n+1}{k+1} = \sum_{j=0}^{k+1} \binom{n}{j}\binom{1}{k+1-j} = \binom{n}{k} \cdot 1 + \binom{n}{k+1} \cdot 1.$$

■

Siehe auch Übung 1.9 und 1.10.

Bemerkung 1.2.15 *Aus den Anfangswerten* $\binom{0}{0} = 1$ *und* $\binom{0}{k} = 0$ *für* $k \neq 0$ *erhalten wir alle anderen Binomialkoeffizienten mittels der Rekursionsgleichung aus Corollar 1.2.14. In dem* ***Pascalschen Dreieck***

$$\begin{array}{ccccccc}
 & & & \binom{0}{0} & & & \\
 & & \binom{1}{0} & & \binom{1}{1} & & \\
 & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \\
\binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} \\
\iddots\ \ddots & & \iddots\ \ddots & & \iddots\ \ddots & & \iddots\ \ddots
\end{array}$$

ist also jeder Eintrag (ausser den Anfangswerten) die Summe der beiden über ihm liegenden Einträge (wobei wir Binomialkoeffizienten gleich 0 *nicht schreiben):*

$$\begin{array}{ccccccccccc}
 & & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & \\
 & & 1 & & 2 & & 1 & & & & \\
 & 1 & & 3 & & 3 & & 1 & & & \\
1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 5 & & 10 & & 10 & & 5 & & 1 &
\end{array}$$

Beispielsweise gilt

$$3 = \binom{3}{1} = \binom{2}{0} + \binom{2}{1} = 1 + 2.$$

Abschließend zeigen wir noch einen wichtigen Satz, der den Binomialkoeffizienten ihren Namen gegeben hat. Dazu verwenden wir:

Definition und Satz 1.2.16 *Der* ***Polynomring*** $K[X]$ *über einem Körper* K *in der Unbestimmten* X *ist die Menge aus* 0 *und allen Ausdrücken (****Polynome****)*

$$f = a_0 + a_1X^1 + \ldots + a_nX^n$$

mit $n \in \mathbb{N}_0$, $a_i \in K$, $a_n \neq 0$. *Diese Elemente können wir auch als unendliche Listen*

$$f = (a_0, \ldots, a_n, 0, 0, \ldots)$$

von Elementen aus K *auffassen, in denen nur endlich viele Einträge* $\neq 0$ *sind. Dann ist*

$$0 = (0, 0, \ldots)$$

Wir nennen $\deg(f) := n$ *den* ***Grad*** *von* f *und setzen* $\deg(0) = -\infty$.
Mit der Addition

$$\begin{aligned}
&\left(a_0 + a_1X^1 + \ldots + a_nX^n\right) + \left(b_0 + b_1X^1 + \ldots + b_mX^m\right) \\
&= \sum_{i=0}^{\max(n,m)} \left(a_i + b_i\right) X^i
\end{aligned}$$

und der Multiplikation

$$\begin{aligned}&\left(a_0 + a_1X^1 + ... + a_nX^n\right) \cdot \left(b_0 + b_1X^1 + ... + b_mX^m\right)\\ &= \sum_{i=0}^{n+m} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) X^i,\end{aligned}$$

wird $K[X]$ ein kommutativer Ring mit 1.

Für den Beweis siehe Aufgabe 1.14.

Beispiel 1.2.17 *In $\mathbb{Q}[X]$ gilt*

$$\begin{aligned}(1 + 2X + X^2) \cdot (1 + X) &= (1 + 2X + X^2) + (X + 2X^2 + X^3)\\ &= 1 + 3X + 3X^2 + X^3.\end{aligned}$$

Mit MAPLE *können wir diese Rechnung folgendermaßen durchführen:*

```
f:=(1+2*X+X^2)*(1+X);
```

$(1 + 2X + X^2) \cdot (1 + X)$

```
expand(f);
```

$1 + 3X + 3X^2 + X^3$

Summanden mit $a_i = 0$ in $f = a_0 + a_1X^1 + ... + a_nX^n$ schreibt man üblicherweise nicht. Ein Polynom der Form $f = X^n$ bezeichnen wir auch als **Monom**, $f = a_nX^n$ als **Term**, und $f = a_mX^m + a_nX^n$ als **Binom**.

Bemerkung 1.2.18 *In der Informatik stellt man ein Polynom $f = a_0 + a_1X^1 + ... + a_nX^n$ meist durch die Liste*

$$(a_0, ..., a_n) \in K^{n+1}$$

*seiner Koeffizienten a_i dar (sogenannte **dicht besetzte Darstellung** von Polynomen). Haben die betrachteten Polynome allerdings nur wenige Koeffizienten $a_i \neq 0$ ist es effizienter das Polynom als die Menge von Tupeln*

$$\{(i, a_i) \mid a_i \neq 0\} \subset \mathbb{N}_0 \times K$$

*zu speichern (**dünn besetzte Darstellung**).*

Beispielsweise würden wir das Polynom $f = 7 + 13 \cdot X^{10}$ *darstellen als*

$$f = (7, 0, 0, 0, 0, 0, 0, 0, 0, 0, 13)$$

oder als

$$f = \{(0, 7), (10, 13)\}.$$

Für die Implementierung der Polynomarithmetik siehe Aufgabe 1.15.

Bemerkung 1.2.19 *Jedem Polynom*

$$p = a_0 + a_1 X^1 + \ldots + a_n X^n \in K[X]$$

ist durch ***Einsetzen*** *eines Werts* $c \in K$ *für die Variable* X *ein Wert*

$$p(c) = a_0 + a_1 c^1 + \ldots + a_n c^n \in K$$

zugeordnet.

Durch Einsetzen kann somit zu jedem Polynom $p \in K[X]$ *eine Abbildung*

$$K \to K,\ c \mapsto p(c)$$

assoziiert werden.

Beispiel 1.2.20 *Die durch das Polynom* $p = X^2 \in \mathbb{R}[X]$ *gegebene Abbildung ist die Parabelfunktion*

$$\mathbb{R} \to \mathbb{R},\ c \mapsto c^2$$

aus Abbildung 1.1.

Bemerkung 1.2.21 *Für alle* $p, q \in K[X]$ *und* $c \in K$ *gilt*

$$(p \cdot q)(c) = p(c) \cdot q(c)$$

$$(p + q)(c) = p(c) + q(c).$$

Es ist also egal, ob wir erst einsetzen und dann Elemente aus K *multiplizieren/addieren, oder erst Polynome multiplizieren/addieren und dann einsetzen. Für den (leichten) Beweis siehe Aufgabe 1.13.*

Man sagt dazu auch: Für jedes $c \in K$ *ist die Einsetzabbildung* $K[X] \to K,\ p \mapsto p(c)$ *ein* ***Ringhomomorphismus****.*

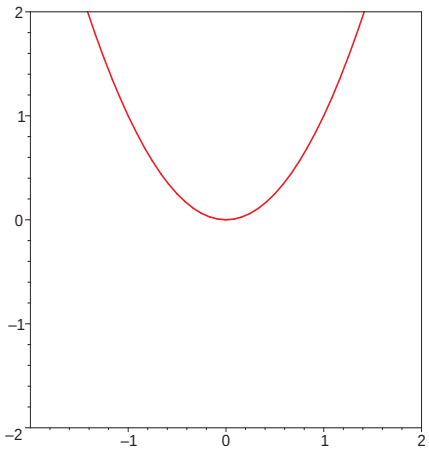


Abbildung 1.1: Graph der Parabel

Wir können also mit Formeln in der abstakten Variablen X rechnen und dann z.B. die erhaltene Formel für viele verschiedene Werte c für X verwenden, anstatt für jedes einzelne c dieselbe Rechnung durchzuführen.

Der Binomialsatz beschreibt, wie man Potenzen von Binomen berechnet:

Satz 1.2.22 (Binomialsatz) *Für alle $n \in \mathbb{N}_0$ gilt*

$$(X+1)^n = \sum_{k=0}^{n} \binom{n}{k} X^k$$

Beweis. Ausmultiplizieren mit dem Distributivgesetz gibt

$$\overbrace{(X+1)\cdot\ldots\cdot(X+1)}^{n} = \sum_{T\subset\{1,\ldots,n\}} X^{|T|}$$

denn jeder Faktor $(X+1)$ auf der linken Seite trägt zu jedem Summanden auf der rechten Seite mit X oder 1 bei. Wir numerieren die Faktoren von $1,\ldots,n$ und interpretieren T als die Menge der Faktoren die mit X beitragen und das Komplement von T als die Menge der Faktoren die mit 1 beitragen.

Da es $\binom{n}{k}$ Teilmengen $T \subset \{1,\ldots,n\}$ mit $|T| = k$ gibt, folgt

$$\sum_{T\subset\{1,\ldots,n\}} X^{|T|} = \sum_{k=0}^{n} \binom{n}{k} X^k.$$

■

Siehe auch Aufgabe 1.11 zur Berechnung von Ausdrücken der Form $(x+y)^n$ mit Hilfe des Binomialsatzes.

Beispiel 1.2.23 *Für $n = 2$ interpretieren wir im Beweis von Satz 1.2.22 die Menge $T \subset \{1,2\}$ als*

T	$\{1,2\}$	$\{1\}$	$\{2\}$	$\varnothing$
Summand	$X \cdot X$	$X \cdot 1$	$1 \cdot X$	$1 \cdot 1$

und erhalten

$$\begin{aligned}(X+1)^2 &= X \cdot X + X \cdot 1 + 1 \cdot X + 1 \cdot 1 \\ &= X^2 + 2X + 1\end{aligned}$$

Beispiel 1.2.24 *Satz 1.2.22 gibt*

$$\begin{aligned}(X+1)^1 &= X + 1 \\ (X+1)^2 &= X^2 + 2X + 1 \\ (X+1)^3 &= X^3 + 3X^2 + 3X + 1 \\ (X+1)^4 &= X^4 + 4X^3 + 6X^2 + 4X + 1 \\ (X+1)^5 &= X^5 + 5X^4 + 10X^3 + 10X^2 + 5X + 1\end{aligned}$$

mit den Binomialkoeffizienten aus Bemerkung 1.2.15.

Beispiel 1.2.25 *Bei einer jährlichen Verzinsung $0 < x < 1$ des Kapitals m, erhält man nach n Jahren von der Bank (hoffentlich)*

$$m \cdot (1+x)^n = m \cdot \sum_{k=0}^{n} \binom{n}{k} x^k.$$

Für kleines x erhalten wir mit dem konstanten und linearen Term der Binomialformel die Approximation

$$m \cdot (1+x)^n \approx m \cdot (1 + n \cdot x).$$

In der Praxis bedeutet dies die Vernachlässigung von Zinseszinsen. Durch Hinzufügen weiterer Terme ansteigender x-Potenz in der Binomialformel lässt sich die Näherung verbessern, etwa zu

$$m \cdot (1+x)^n \approx m \cdot (1 + n \cdot x + \frac{n(n-1)}{2} x^2).$$

Beispielsweise für $x = \frac{1}{100}$ *und* $n = 3$ *wird*

$$(1 + \frac{1}{100})^3 = 1.030301$$

durch

$$1 + 3 \cdot \frac{1}{100} = 1.03$$

bzw.

$$1 + 3 \cdot \frac{1}{100} + 3 \cdot \frac{1}{10000} = 1.0303$$

approximiert. Die exakte Formel erhalten wir mit dem Binomialsatz als

$$1 + 3 \cdot \frac{1}{100} + 3 \cdot \frac{1}{10000} + 1 \cdot \frac{1}{1000000} = 1.030301$$

Für den Beweis der Siebformel im folgenden Abschnitt zeigen wir noch ein Corollar zum Binomialsatz:

Corollar 1.2.26 *Für alle* $n \in \mathbb{N}_0$ *gilt*

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0.$$

Beweis. Sei $f = (X + 1)^n$ und $g = \sum_{k=0}^{n} \binom{n}{k} X^k$. Mit Bemerkung 1.2.21 ist $f(-1) = (-1 + 1)^n = 0$. Andererseits ist $g(-1) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k$. Wegen Satz 1.2.22 gilt $f = g$ also auch $f(-1) = g(-1)$. ■

Beispiel 1.2.27 *Es gilt*

$$\binom{4}{0} - \binom{4}{1} + \binom{4}{2} - \binom{4}{3} + \binom{4}{4} = 1 - 4 + 6 - 4 + 1 = 0$$

1.3 Siebformel

Bevor wir als Anwendung von Binomialkoeffizienten im nächsten Abschnitt die Catalanzahlen diskutieren, leiten wir noch als eine wichtige Folgerung aus Corollar 1.2.26 die Siebformel her. Für die Vereinigung von zwei endlichen Mengen M_1, M_2 gilt die bekannte Formel

$$|M_1 \cup M_2| = |M_1| + |M_2| - |M_1 \cap M_2|$$

(siehe Übung 1.2). Diese Beziehung bezeichnet man auch als das Prinzip der Inklusion und Exklusion. Die Siebformel verallgemeinert diese Formel auf eine beliebige Anzahl n endlicher Mengen $M_1, ..., M_n$: Sie setzt die Anzahl der Elemente von $M_1 \cup ... \cup M_n$ mit der Anzahl der Elemente der Durchschnitte

$$M_T = \bigcap_{i \in T} M_i$$

für alle $T \subset \{1, ..., n\}$ in Beziehung.

Satz 1.3.1 (Siebformel) *Für endliche Mengen $M_1, ..., M_n$ gilt*

$$|M_1 \cup ... \cup M_n| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{|T|=k} |M_T|$$

Beispiel 1.3.2 *Für drei Mengen erhalten wir*

$$\begin{aligned} |M_1 \cup M_2 \cup M_3| = {} & |M_1| + |M_2| + |M_3| \\ & - |M_1 \cap M_2| - |M_1 \cap M_3| - |M_2 \cap M_3| \\ & + |M_1 \cap M_2 \cap M_3| \end{aligned}$$

siehe auch Abbildung 1.2.

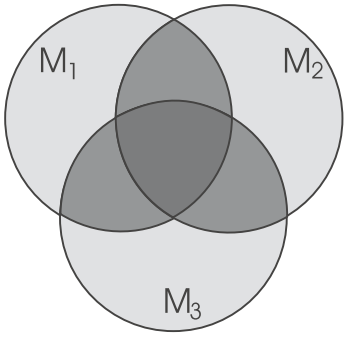


Abbildung 1.2: Siebformel für drei Mengen.

Nun zum Beweis von Satz 1.3.1:

Beweis. Sei $x \in M_1 \cup ... \cup M_n$. Wir wollen zeigen, dass x zu der rechten Seite genau mit 1 beiträgt. Angenommen x liegt in genau r der Mengen M_i, ohne Einschränkung $x \in M_1 \cap ... \cap$

M_r. Dann wird x in $\sum_{|T|=k} |M_T|$ genau $\binom{r}{k}$-mal gezählt, in jedem Durchschnitt von k der $M_1, ..., M_r$ genau 1-mal. Insgesamt trägt x also zu der rechten Seite mit

$$a = \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k}$$

bei. Da mit Corollar 1.2.26

$$0 = \sum_{k=0}^{r} (-1)^k \binom{r}{k} = 1 - a$$

gilt, ist $a = 1$. ■

Beispiel 1.3.3 *Wir illustrieren den Beweis für $n = 3$: Sei z.B. $r = 2$ also OE $x \in M_1 \cap M_2$ und $x \notin M_1 \cap M_2 \cap M_3$, siehe Abbildung 1.3. Es gibt folgende Möglichkeiten für Teilmengen $T \subset \{1, ..., n\}$*

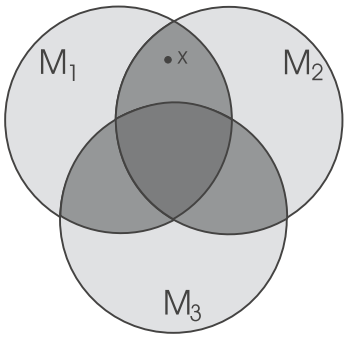


Abbildung 1.3: Beitrag zur Siebformel für $r = 2$.

mit $x \in M_T$:

k	$(-1)^{k-1}$	T *mit* $x \in M_T$
1	1	$\{1\}, \{2\}$
2	−1	$\{1, 2\}$
3	1	

Somit trägt x zu der rechten Seite mit

$$(1 + 1) - 1 + 0 = \binom{2}{1} - \binom{2}{2} + \binom{2}{3} = 1$$

bei. Genauso geht man für $r = 1$ bzw. $r = 3$ vor und erhält

$$1 - 0 + 0 = \binom{1}{1} - \binom{1}{2} + \binom{1}{3} = 1$$

bzw.

$$(1+1+1)-(1+1+1)+1=\binom{3}{1}-\binom{3}{2}+\binom{3}{3}=1.$$

Beispiel 1.3.4 *Mit der Siebformel können wir die Anzahl der Primzahlen* ≤ 40 *bestimmen: In der Primfaktorisierung einer Zahl* $n \leq 40$ *ist der kleinste Primfaktor* $p \leq 6$ *(also* $2, 3$ *oder* 5*), denn gilt* $n = p \cdot q$ *mit* $p \leq q$*, dann ist* $p^2 \leq p \cdot q = n$*.*

Sei T_m *die Menge der durch* m *teilbaren Zahlen* ≤ 40*, also*

$$T_m = \{a \cdot m \mid a \in \mathbb{N} \text{ mit } a \cdot m \leq 40\}.$$

Somit ist

$$|T_m| = \left\lfloor \frac{40}{m} \right\rfloor$$

wobei $\lfloor q \rfloor$ *die Abrundung von* q*, also die größte ganze Zahl* $\leq q$ *bezeichnet. Für* $\mathrm{ggT}(m_1, m_2) = 1$ *haben wir*

$$T_{m_1} \cap T_{m_2} = T_{m_1 \cdot m_2}$$

denn eine Zahl ist durch m_1 *und* m_2 *teilbar genau dann, wenn sie durch* $\mathrm{kgV}(m_1, m_2) = \frac{m_1 \cdot m_2}{\mathrm{ggT}(m_1, m_2)} = m_1 \cdot m_2$ *teilbar ist. Beispielsweise ist eine Zahl durch* 6 *teilbar genau dann, wenn sie durch* 2 *und* 3 *teilbar ist. Somit gilt*

$$T_2 \cap T_3 = T_6 \qquad T_2 \cap T_5 = T_{10} \qquad T_3 \cap T_5 = T_{15}$$
$$T_2 \cap T_3 \cap T_5 = T_{30}$$

Die Siebformel liefert dann

$$\begin{aligned} |T_2 \cup T_3 \cup T_5| &= |T_2| + |T_3| + |T_5| \\ &\quad - |T_6| - |T_{10}| - |T_{15}| \\ &\quad + |T_{30}| \\ &= (20 + 13 + 8) - (6 + 4 + 2) + 1 \\ &= 30 \end{aligned}$$

Es gibt also 30 *Zahlen die ein Vielfaches von* 2*,* 3 *oder* 5 *sind. Somit gibt es genau*

$$30 - 3 = 27$$

zusammengesetzte Zahlen ≤ 40 *(denn* $2,3,5 \in T_2 \cup T_3 \cup T_5$ *sind prim). Alle anderen Zahlen* ≤ 40 *ausser der* 1 *sind prim, also gibt es genau*

$$40 - 27 - 1 = 12$$

Primzahlen ≤ 40.

In MAPLE *erhalten wir diese Primzahlen wie folgt:*

```
L:=[];
for j from 1 to 40 do
    if isprime(j) then L:=[op(L),j];fi;
od:
L;
[2,3,5,7,11,13,17,19,23,29,31,37]
```

Bemerkung 1.3.5 *Die* MAPLE*-Funktion* `isprime` *ist ein* ***probabilistischer Primzahltest****, d.h für* $n \in \mathbb{Z}$ *beweist das Ergebnis* `isprime(n)=false`*, dass* n *echt zusammengesetzt ist. Andererseits bedeutet* `isprime(n)=true` *nur, dass* n *mit sehr hoher Wahrscheinlichkeit eine Primzahl ist.*

Es ist keine Zahl n *bekannt, für die* `isprime` *fälschlicherweise* `true` *liefert, und man vermutet, dass ein solches* n *mehrere hundert Dezimalstellen haben muss.*

1.4 Anwendung: Vollständige Klammerungen und Catalan-Zahlen

Im Folgenden diskutieren wir noch eine Anwendung von Binomialkoeffizienten in der Informatik genauer. Nehmen wir an, wir wollen im Computer $2 \cdot 3 \cdot 4 = 24$ berechnen. Prozessoren können stets in jedem Schritt nur eine arithmetische Operation ausführen. Auch im Sinn der Mathematik ist die Addition und die Multiplikation in einem Ring eine Abbildung mit zwei Argumenten

$$+ : R \times R \to R$$
$$\cdot : R \times R \to R$$

Wir müssen den Ausdruck also so klammern, dass stets nur zwei Zahlen verknüpft werden. Man spricht dann auch von einer **vollständigen Klammerung**. Da die Multiplikation in $\mathbb{Z}$ assoziativ

ist, spielt die Wahl der Klammerung für das Ergebnis keine Rolle:

$$(2 \cdot (3 \cdot 4)) = 2 \cdot 12 = 24 = 6 \cdot 4 = ((2 \cdot 3) \cdot 4).$$

Beinhaltet der Ausdruck sowohl Additionen als auch Multiplikationen, dann ist die der Klammerung auch für seine syntaktische Analyse im Computer wichtig, denn das Ergebnis hängt im Allgemeinen von der Klammerung ab z.B.

$$((2 \cdot 3) + 4) \neq (2 \cdot (3 + 4)).$$

In unserem Beispiel $2 \cdot 3 \cdot 4$ gibt es offenbar zwei Möglichkeiten das Produkt zu klammern. Im Folgenden wollen wir die Frage beantworten, wieviele vollständige Klammerungen es für ein Produkt

$$x_1 \cdot \ldots \cdot x_m$$

aus m Faktoren x_i in einem Ring R gibt.

Beispiel 1.4.1 *Für 4 Faktoren gibt es folgende Klammerungen*

$$\begin{gathered}(x_1 \cdot (x_2 \cdot (x_3 \cdot x_4))) \\ (x_1 \cdot ((x_2 \cdot x_3) \cdot x_4)) \\ ((x_1 \cdot x_2) \cdot (x_3 \cdot x_4)) \\ ((x_1 \cdot (x_2 \cdot x_3)) \cdot x_4) \\ (((x_1 \cdot x_2) \cdot x_3) \cdot x_4)\end{gathered}$$

Definition 1.4.2 *Für $n \in \mathbb{N}_0$ ist die **Catalan-Zahl** c_n die Anzahl der vollständigen Klammerungen eines Produkts $x_1 \cdot \ldots \cdot x_{n+1}$ aus $n+1$ Faktoren.*

Offenbar gilt $c_0 = 1$, $c_1 = 1$ und wie gerade gesehen ist $c_2 = 2$ und $c_3 = 5$. Über die folgende Rekursionsgleichung können wir alle c_n berechnen:

Satz 1.4.3 *Es gilt $c_0 = 1$ und*

$$c_n = \sum_{j=0}^{n-1} c_j c_{n-1-j}$$

für $n \geq 1$.

Beispiel 1.4.4 *Nach dem Satz gilt also z.B.*

$$\begin{aligned} c_0 &= 1 \\ c_1 &= c_0^2 = 1 \\ c_2 &= c_0c_1 + c_1c_0 = 2 \\ c_3 &= c_0c_2 + c_1^2 + c_2c_0 = 5 \\ c_4 &= c_0c_3 + c_1c_2 + c_2c_1 + c_0c_3 = 14 \end{aligned}$$

Wir zeigen nun Satz 1.4.3:

Beweis. Sei K_n die Menge der vollständig geklammerten Produkte aus $n+1$ beliebigen Faktoren, also $c_n = |K_n|$. Dann ist

$$\begin{aligned} \bigcup_{j=0}^{n-1} K_j \times K_{n-1-j} &\to K_n \\ (p,q) &\mapsto (p \cdot q) \end{aligned}$$

eine bijektive Abbildung, denn sie hat eine Umkehrabbildung: Jedes Element von K_n (mit $n+1$ Faktoren) lässt sich eindeutig in die zwei Produkte $p \in K_j$ (mit $j+1$ Faktoren) und $q \in K_{n-1-j}$ (mit $n-j$ Faktoren) in der äußersten Klammer zerlegen.

Die Formel folgt dann, da die Vereinigung disjunkt ist, mit Übung 1.2. ■

Beispiel 1.4.5 *Wir illustrieren die Zerlegung im Beweis an Beispiel 1.4.1:*

$(p \cdot q)$	p	q
$(x_1 \cdot (x_2 \cdot (x_3 \cdot x_4)))$	x_1	$(x_2 \cdot (x_3 \cdot x_4))$
$(x_1 \cdot ((x_2 \cdot x_3) \cdot x_4))$	x_1	$((x_2 \cdot x_3) \cdot x_4)$
$((x_1 \cdot x_2) \cdot (x_3 \cdot x_4))$	$(x_1 \cdot x_2)$	$(x_3 \cdot x_4)$
$((x_1 \cdot (x_2 \cdot x_3)) \cdot x_4)$	$(x_1 \cdot (x_2 \cdot x_3))$	x_4
$(((x_1 \cdot x_2) \cdot x_3) \cdot x_4)$	$(x_1 \cdot x_2)$	x_4

Man erhält also

$$c_3 = 5 = 1 \cdot 2 + 1 \cdot 1 + 2 \cdot 1 = c_0c_2 + c_1^2 + c_2c_0.$$

Können wir eine geschlossene Formel für die Catalan-Zahlen herleiten? Zunächst bemerken wir:

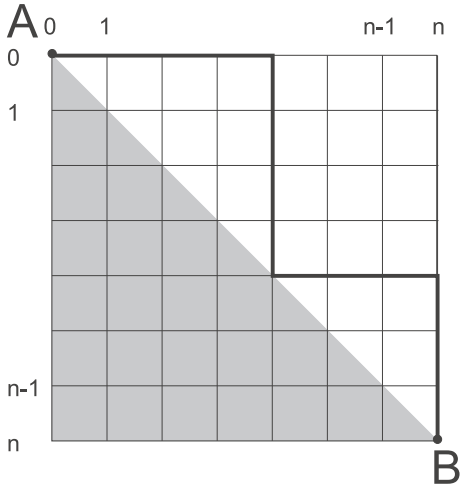


Abbildung 1.4: Kürzeste Wege überhalb der Winkelhalbierenden in einem quadratischen Gitter

Satz 1.4.6 *Es gibt eine Bijektion zwischen der Menge der vollständigen Klammerungen von* $x_1 \cdot \ldots \cdot x_{n+1}$ *und der Menge der kürzesten, überhalb der Winkelhalbierenden verlaufenden Wege in einem* $(n+1) \times (n+1)$*-Gitter (Abbildung 1.4).*

Beweis. Offenbar ist durch folgende Vorschrift eine Abbildung gegeben:

1) Streiche in der Klammerung die Symbole x_i und die Klammern (.

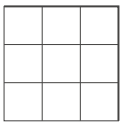
2) Durchlaufe die verbleibenden Symbole von links nach rechts und gehe für jedes $\cdot$ in dem Gitter nach rechts und für jede Klammer) nach unten.

Eine solche Abbildungsvorschrift, die ein Wort liest und abhängig von den Buchstaben Operationen ausführt, bezeichnet man in der Informatik auch als einen **Automaten**. Auf Automaten werden wir in Abschnitt 1.6 etwas näher eingehen.

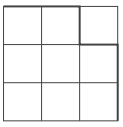
Die Abbildung ist wohldefiniert, da wir jeder Klammer) eine Multiplikation links davon zuordnen können. Um zu zeigen, dass die Abbildung bijektiv ist, konstruiere man als Übung die Umkehrabbildung. ■

Beispiel 1.4.7 *In Beispiel 1.4.1 ordnen wir zu*

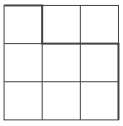
$$(x_1\cdot(x_2\cdot(x_3\cdot x_4))) \quad\mapsto\quad \cdots))) \quad\mapsto$$



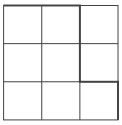
$$(x_1\cdot((x_2\cdot x_3)\cdot x_4)) \quad\mapsto\quad \cdot\cdot)\cdot)) \quad\mapsto$$



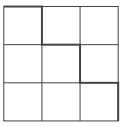
$$((x_1\cdot x_2)\cdot(x_3\cdot x_4)) \quad\mapsto\quad \cdot)\cdot\cdot)) \quad\mapsto$$



$$((x_1\cdot(x_2\cdot x_3))\cdot x_4) \quad\mapsto\quad \cdot\cdot))\cdot) \quad\mapsto$$



$$(((x_1\cdot x_2)\cdot x_3)\cdot x_4) \quad\mapsto\quad \cdot)\cdot)\cdot) \quad\mapsto$$



Satz 1.4.8 *Die Anzahl der überhalb der Winkelhalbierenden verlaufenden Wege in einem $(m+1)\times(n+1)$-Gitter mit $n\geq m$ ist*

$$\frac{n+1-m}{n+1}\binom{n+m}{m}.$$

Beweis. In Übung 1.19 zeigen wir, dass die Anzahl gleich

$$\binom{n+m}{n}-\binom{n+m}{n+1}=\left(1-\frac{m}{n+1}\right)\cdot\binom{n+m}{n}$$

ist, wobei die Gleichheit mit Corollar 1.2.10 folgt. ■

Corollar 1.4.9 *Es gilt*

$$c_n=\frac{1}{n+1}\binom{2n}{n}$$

Beweis. Folgt sofort aus Satz 1.4.6 und Satz 1.4.8 mit $n=m$. ■

Beispiel 1.4.10 *In* MAPLE *können wir die Catalan-Zahlen $c_0,\ldots,c_{10}$ berechnen durch:*

```
seq(binomial(2*n,n)/(n+1),n=0..10);
```

```
1,1,2,5,14,42,132,429,1430,4862,16796
```

Erhält man eine solche Folge c_n von Zahlen durch Experimente, kann man in der *Online Encyclopeadia of Integer Sequences* [17] überprüfen, welche kombinatorischen Interpretationen der Folge bekannt sind. Diese Datenbank enthält Beschreibungen von über 200000 Folgen von ganzen Zahlen. Insbesondere findet man dort noch viele weitere Interpretationen der Catalan-Zahlen.

1.5 Abzählen von Abbildungen

Viele wichtige Klassen von Objekten in der Informatik sind im mathematischen Sinne Abbildungen. Das wichtigste Beispiel ist eine **Liste** $L = (L_1, ..., L_n) \in M^n$ der Länge n mit Einträgen $L_i \in M$, die wir auch als Abbildung

$$\begin{aligned} \{1, ..., n\} &\to M \\ i &\mapsto L_i \end{aligned}$$

auffassen können (in manchen Programmiersprachen beginnt die Indizierung der Liste auch mit 0, d.h. wir betrachten Abbildungen $\{0, ..., n-1\} \to M$). Eine **Matrix**, oder in der Informatik ein **Array**, ist eine Abbildung

$$\{1, ..., m\} \times \{1, ..., n\} \to M.$$

Die Einträge werden also durch zwei Zahlen indiziert.

Beispiel 1.5.1 *Sei $M = \{a, ..., z\}$. Die Liste*

$$(a, h, a)$$

entspricht der Abbildung

$$\begin{aligned} \{1, 2, 3\} &\to M \\ 1 &\mapsto a \\ 2 &\mapsto h \\ 3 &\mapsto a \end{aligned}$$

Das Array

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

wird durch die Abbildung

$$\begin{aligned}\{1,2\}\times\{1,2,3\} &\to M\\ (1,1) &\mapsto a\\ (1,2) &\mapsto b\\ (1,3) &\mapsto c\\ (2,1) &\mapsto d\\ (2,2) &\mapsto e\\ (2,3) &\mapsto f\end{aligned}$$

dargestellt.

Die Frage nach der Anzahl solcher Listen oder Arrays übersetzt sich also in die Frage nach der Anzahl der entsprechenden Abbildungen. Diese Frage können wir allgemein beantworten:

Satz 1.5.2 *Sind N und M endliche Mengen mit $|N| = n$ und $|M| = m$, dann gibt es*

$$m^n$$

Abbildungen $N \to M$.

Beweis. Sei $f : N \to M$ eine Abbildung und schreibe $N =: \{x_1, ..., x_n\}$. Für jedes $f(x_i)$ gibt es m Möglichkeiten, insgesamt also

$$\overbrace{m \cdot ... \cdot m}^{n} = m^n$$

Abbildungen f. ■

Definition 1.5.3 *Wir schreiben kurz M^N für die Menge aller Abbildungen $f : N \to M$.*

Notation 1.5.4 *Für $N = \{x_1, ..., x_n\}$ schreiben wir die Abbildungsvorschrift für $f : N \to M$ auch als kurz als*

$$f = \begin{pmatrix} x_1 & \cdots & x_n \\ f(x_1) & \cdots & f(x_n) \end{pmatrix}$$

d.h. wir notieren in der ersten Zeile der Tabelle die Elemente der Quelle und in der zweiten Zeile jeweils deren Bild.

Beispiel 1.5.5 *Für $N = \{1,2,3\}$ und $M = \{a,b\}$ sind alle Abbildungen $f : N \to M$ gegeben durch*

$$\begin{pmatrix}1&2&3\\a&a&a\end{pmatrix}, \begin{pmatrix}1&2&3\\a&b&b\end{pmatrix}, \begin{pmatrix}1&2&3\\b&a&b\end{pmatrix}, \begin{pmatrix}1&2&3\\b&b&a\end{pmatrix},$$
$$\begin{pmatrix}1&2&3\\b&b&b\end{pmatrix}, \begin{pmatrix}1&2&3\\b&a&a\end{pmatrix}, \begin{pmatrix}1&2&3\\a&b&a\end{pmatrix}, \begin{pmatrix}1&2&3\\a&a&b\end{pmatrix}.$$

1.6 Anwendung: Worte

In Abschnitt 1.4 haben wir schon von endlichen Sequenzen aus Symbolen · und) gesprochen. Was ist das eigentlich im mathematischen Sinne?

Definition 1.6.1 *Sei A eine endliche Menge. Ein **Wort** mit n Buchstaben über dem **Alphabet** A ist ein Element von A^n. Wir schreiben für $(a_1,...,a_n) \in A^n$ auch kurz*

$$a_1...a_n$$

Im Sinne der Informatik ist ein Wort also einfach eine endliche Liste.

Beispiel 1.6.2 *Über dem Alphabet $A = \{a,...,z\}$ schreiben wir*

$$hallo = (h,a,l,l,o) \in A^5.$$

Beispiel 1.6.3 *Eine 8-bit Zahl ist ein Wort in $\{0,1\}^8$.*

Chinesische Worte sind oft in $\{1,...,3000\}^2$, d.h. sie haben oft 2 Buchstaben allerdings in einem Alphabet von etwa 3000 Zeichen.

Bemerkung 1.6.4 *Ein Wort $(a_1,...,a_n) \in A^n$ können wir auch als die Abbildung*

$$\begin{aligned}\{1,...,n\} &\to A\\ i &\mapsto a_i\end{aligned}$$

auffassen.

*Damit ist auch klar, was das **leere Wort** sein soll. Es ist die (eindeutige) Abbildung $\varnothing \to A$.*

Beispiel 1.6.5 *Das Wort aha enspricht der Abbildung*

$$\begin{aligned} \{1,2,3\} &\to \{a,...,z\} \\ 1 &\mapsto a \\ 2 &\mapsto h \\ 3 &\mapsto a \end{aligned}$$

Da Worte der Länge n in dem Alphabet A dasselbe wie Abbildungen $\{1,...,n\} \to A$ sind, gilt mit Satz 1.5.2:

Satz 1.6.6 *Die Anzahl der Worte der Länge n in einem Alphabet A mit $|A| = m$ Elementen ist*

$$m^n$$

Wir beschreiben noch jeweils eine zentrale Anwendung von Worten in der Informatik und der Mathematik:

Bemerkung 1.6.7 *In der Informatik spielen Worte eine wichtige Rolle in der Berechenbarkeitstheorie. Ein* ***Automat*** *nimmt als Eingabe ein Wort $(a_1,...,a_n) \in A^n$ und liest die Buchstaben von links nach rechts. Ausgehend von seinem* ***Ausgangszustand*** *wechselt er in jedem Schritt i abhängig von a_i und seinem aktuellen* ***Zustand*** *in einen neuen Zustand. Am Ende prüft er, ob sein* ***Endzustand*** *in einer gegebenen Menge von zulässigen Endzuständen ist.*

Zum Beispiel können wir einen Parkautomaten betrachten. Sein Anfangszustand ist $0\,€$, zulässig sei nur der Endzustand der exakten Parkgebühr $3\,€$. Wir werfen 2 Münzen ein, $1\,€$ oder $2\,€$. Zulässig sind dann

Wort	*Zustandsfolge*
$(1\,€, 2\,€)$	$0\,€, 1\,€, 3\,€$
$(2\,€, 1\,€)$	$0\,€, 2\,€, 3\,€$

unzulässig dagegen

Wort	*Zustandsfolge*
$(1\,€, 1\,€)$	$0\,€, 1\,€, 2\,€$
$(2\,€, 2\,€)$	$0\,€, 2\,€, 4\,€$

Von den $2^2 = 4$ möglichen Worten sind also 2 zulässig und 2 nicht.

Wir skizzieren noch kurz eine wichtige Anwendung von Worten in der Mathematik:

Bemerkung 1.6.8 *Sind $w = a_1...a_n$ und $v = b_1...b_m$ Worte, dann definiert man die Verknüpfung "Hintereinanderschreiben" durch*

$$w \circ v = a_1...a_n b_1...b_m$$

Die Menge W aller Worte (beliebiger Länge) in dem Alphabet A ist zusammen mit $\circ$ ein Monoid. Die Assoziativität ist klar und das neutrale Element ist das leere Wort.

Fügen wir zu dem Alphabet zusätzliche Buchstaben a^{-1} für $a \in A$ mit der Rechenregel

$$aa^{-1} = a^{-1}a = e$$

*hinzu, dann erhalten wir die **freie Gruppe** F erzeugt von A.*

Bemerkung 1.6.9 *Sei*

$$A = \{g_1, ..., g_n\}$$

eine endliche Menge und F die freie Gruppe erzeugt von A (mit neutralem Element e). Seien $r_1, ..., r_s$ Elemente von F und N der kleinste Normalteiler von F, der $r_1, ..., r_s$ enthält. Dann heißt

$$\langle g_1, ..., g_n \mid r_1 = e, ..., r_s = e \rangle := F/N$$

*die Gruppe mit **Erzeugern** g_i **und Relationen** r_i.*

Beispiel 1.6.10 *Durch*

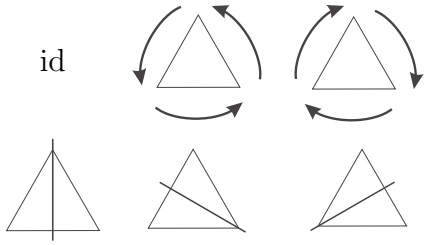
$$\begin{aligned} \langle g \mid g^5 = e \rangle &\to \mathbb{Z}/5 \\ g &\mapsto \bar{1} \end{aligned}$$

ist ein Gruppenisomorphismus gegeben: Sei F die freie Gruppe erzeugt von g. Der Kern von

$$\begin{aligned} F &\to \mathbb{Z}/5 \\ g &\mapsto \bar{1} \end{aligned}$$

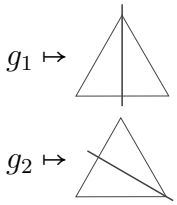
ist offenbar die Untergruppe $\langle g^5 \rangle$ erzeugt von g^5. Somit folgt die Behauptung aus dem Homomorphiesatz.

Beispiel 1.6.11 *Das gleichseitige Dreieck D hat* 6 ***Symmetrien*** *(d.h. abstandserhaltende Abbildungen die D wieder auf sich selbst abbilden): die Identität, zwei Drehungen (um* 120° *und* 240°*) und* 3 *Spiegelungen (an einer Geraden durch eine Ecke und eine Seitenmitte). Diese Elemente wollen wir schematisch schreiben als:*



Die Menge $\mathrm{Sym}(D)$ *dieser Symmetrien bildet eine Gruppe bezüglich der Komposition (die Hintereinanderausführung von zwei Symmetrien ist eine Symmetrie und jede Symmetrie hat eine inverse Symmetrie). Ähnlich wie im vorangegangenen Beispiel kann man zeigen, dass*

$$\langle g_1, g_2 \mid g_1^2 = e,\ g_2^2 = e,\ (g_1 g_2)^3 = e \rangle \to \mathrm{Sym}(D)$$



einen Gruppenisomorphismus definiert.

1.7 Abzählen von injektiven Abbildungen

In Abschnitt 1.5 haben wir schon die Menge aller Abbildungen $N \to M$ zwischen zwei endlichen Mengen abgezählt. Wieviele der Abbildungen sind injektiv, d.h. auf wieviele Weisen kann man N als Teilmenge von M auffassen?

Satz 1.7.1 *Sind N und M endliche Mengen mit $|N| = n$ und $|M| = m$, dann gibt es*

$$\prod_{i=0}^{n-1}(m-i) = \underbrace{m \cdot (m-1) \cdot \ldots \cdot (m-n+1)}_{n}$$

injektive Abbildungen $N \to M$.

Beweis. Sei $f : N \to M$ eine injektive Abbildung und schreibe $N = \{x_1, ..., x_n\}$. Für $f(x_1)$ gibt es m Möglichkeiten, für $f(x_2)$ noch $m-1$, induktiv für $f(x_i)$ noch $m-i+1$ Möglichkeiten, falls $n \leq m$.

Für $n > m$ gibt es nach dem Schubfachprinzip (siehe Aufgabe 1.7(1)) keine injektive Abbildung $N \to M$. Andererseits ist auch das Produkt gleich 0, denn der Faktor für $i = m$ verschwindet. ■

Beispiel 1.7.2 *Sei $N = \{1, 2\}$ und $M = \{a, b, c\}$. Dann sind die injektiven Abbildungen $f : N \to M$ gegeben durch*

$$\begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ a & c \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ b & a \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ b & c \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ c & a \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ c & b \end{pmatrix}.$$

Im Satz erhalten wir

$$\prod_{i=0}^{1}(3-i) = 3 \cdot 2 = 6.$$

Für $N = \{1, 2, 3, 4\}$ und $M = \{a, b\}$ gibt es keine injektive Abbildung $f : N \to M$ und im Satz erhalten wir

$$\prod_{i=0}^{3}(2-i) = 2 \cdot 1 \cdot 0 \cdot (-1) = 0.$$

Nach Übung 1.7 kann es eine bijektive Abbildung zwischen den endlichen Mengen N und M nur geben, wenn $|N| = |M|$.

In diesem Fall sind nach Aufgabe 1.4 die Eigenschaften injektiv, surjektiv und bijektiv äquivalent. Somit folgt wegen

$$\prod_{i=1}^{n}(n-i+1) = n \cdot \ldots \cdot 1 = n!$$

aus Satz 1.7.1:

Corollar 1.7.3 *Sind N und M endliche Mengen mit $|N| = |M| = n$, dann gibt es*

$$n!$$

bijektive Abbildungen $N \to M$.

Bemerkung 1.7.4 *Bijektive Abbildungen $M \to M$ bezeichnet man auch als **Permutationen** von M. Die Menge der bijektiven Abbildungen*

$$S(M) = \{f : M \to M \text{ bijektiv}\}$$

bildet mit der Komposition als Verknüpfung eine Gruppe, denn die Komposition ist assoziativ (siehe Aufgabe 1.5) und die Komposition von zwei bijektiven Abbildungen ist wieder bijektiv (siehe Aufgabe 1.6).

Speziell für $M = \{1, \ldots, n\}$ heißt

$$S_n = S(\{1, \ldots, n\})$$

*die **symmetrische Gruppe**. Elemente $f \in S_n$ schreibenn wir wie oben kurz als*

$$f = \begin{pmatrix} 1 & \cdots & n \\ f(1) & \cdots & f(n) \end{pmatrix}$$

Die erste Zeile ist in dieser Situation eigentlich überflüssig (da die Argumente $1, \ldots, n$ sind, und wir somit f als Liste $(f(1), \ldots, f(n))$ auffassen könnten), wird aber traditionell der Übersichtlichkeit halber geschrieben. Dies ist besonders nützlich bei der Verknüpfung von Permutationen:

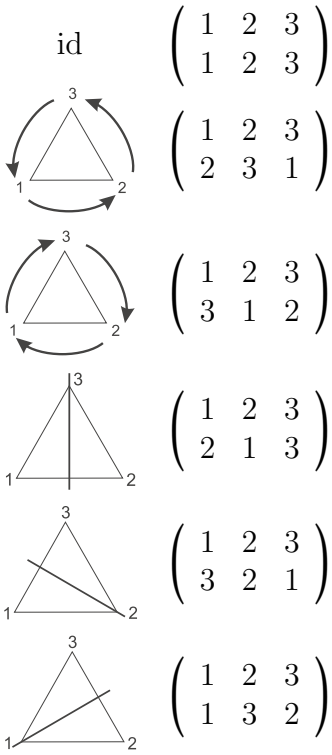
Beispiel 1.7.5 *In der S_3 gilt*

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

denn

$$1 \mapsto 1 \mapsto 2$$
$$2 \mapsto 3 \mapsto 3$$
$$3 \mapsto 2 \mapsto 1$$

Beispiel 1.7.6 *Die Symmetrie des gleichseitigen Dreiecks (d.h. die Gruppe der Drehungen und Spiegelungen, die das Dreieck wieder auf sich selbst abbilden, mit der Komposition als Verknüpfung) ist isomorph zur S_3, da die Lage des Dreiecks durch die Lage der Eckpunkte festgelegt ist. Als Permutationen der Ecken aufgefasst sind die Elemente*



Analog dazu lässt sich jede Symmetrie des Quadrats durch Nummerieren der Ecken als Element der S_4 auffassen (siehe Abbildung 1.7). Jedoch ist nicht jedes Element der S_4 eine Symmetrie des Quadrats (siehe Aufgabe 1.22 und Aufgabe 1.23 analog für das regelmäßige Fünfeck).

1.8 Abzählen von surjektiven Abbildungen

Schließlich zählen wir noch die surjektiven Abbildungen ab. Als Anwendung werden wir im nächsten Abschnitt herleiten, wieviele Partitionen bzw. Äquivalenzrelationen es auf einer endlichen Menge gibt.

Satz 1.8.1 *Sind N und M endliche Mengen mit $|N| = n$ und $|M| = m$, dann gibt es*

$$\sum_{k=0}^{m}(-1)^k\binom{m}{k}(m-k)^n$$

surjektive Abbildungen $N \to M$.

Beweis. Ohne Einschränkung ist $M = \{1, ..., m\}$. Für $i \in M$ sei

$$A_i = \{f : N \to M \mid i \notin f(N)\}$$

die Menge der Abbildungen, die i nicht treffen. Die Menge der nicht surjektiven Abbildungen ist also $A_1 \cup ... \cup A_m$. Mit der Siebformel (Satz 1.3.1) erhalten wir also

$$|A_1 \cup ... \cup A_m| = \sum_{k=1}^{m}(-1)^{k-1}\sum_{|T|=k}|A_T|$$

wobei für $T \subset \{1, ..., m\}$

$$A_T = \bigcap_{i\in T} A_i$$

die Menge der Abbildungen ist, die T nicht treffen. Für festes k gibt es $\binom{m}{k}$ Wahlen für T. Für jedes solche T gilt

$$|A_T| = (m-k)^n,$$

denn für jedes $f(x)$, $x \in N$ gibt es $m-k$ Möglichkeiten in $M\backslash T$.

Die Zahl der surjektiven Abbildungen ist dann die Anzahl aller Abbildungen minus die Anzahl der nicht surjektiven, also

$$\begin{aligned} &m^n - \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}(m-k)^n \\ &= \sum_{k=0}^{m}(-1)^k\binom{m}{k}(m-k)^n \end{aligned}$$

■

Beispiel 1.8.2 *Für $N = \{1,2,3\}$ und $M = \{a,b\}$ sind die surjektiven Abbildungen $f : N \to M$, gegeben durch*

$$\begin{pmatrix}1&2&3\\a&a&b\end{pmatrix}, \begin{pmatrix}1&2&3\\a&b&a\end{pmatrix}, \begin{pmatrix}1&2&3\\b&a&a\end{pmatrix},$$
$$\begin{pmatrix}1&2&3\\b&b&a\end{pmatrix}, \begin{pmatrix}1&2&3\\b&a&b\end{pmatrix}, \begin{pmatrix}1&2&3\\a&b&b\end{pmatrix}.$$

Dagegen sind die Abbildungen

$$\begin{pmatrix}1&2&3\\a&a&a\end{pmatrix}, \begin{pmatrix}1&2&3\\b&b&b\end{pmatrix}$$

nicht surjektiv.

Im Satz erhalten wir

$$\sum_{k=0}^{2}(-1)^k\binom{2}{k}(2-k)^3$$
$$= 2^3 - \binom{2}{1}1^3 + 0 = 8 - 2 = 6$$

Dabei ist der $k = 0$ Term die Anzahl aller Abbildungen und der $k = 1$ Term die Anzahl der Abbildungen, die genau ein Element von M nicht treffen.

Siehe auch Übungsaufgabe 1.24.

1.9 Anwendung: Partitionen von Mengen und Äquivalenzrelationen

Um die Äquivalenzrelationen auf einer endlichen Menge N abzuzählen, setzen wir diese zunächst mit den Partitionen von N in Beziehung:

Definition 1.9.1 *Eine **Partition** einer Menge N ist eine Menge $P = \{P_1, ..., P_m\}$ von Teilmengen $\varnothing \neq P_i \subset N$ sodass*

1) die P_i paarweise disjunkt sind, d.h. $P_i \cap P_j = \varnothing$ für alle $i \neq j$, und

2) $N = P_1 \cup ... \cup P_m$

Es ist auch gebräuchlich P *durch den Ausdruck*

$$N = P_1 \cup ... \cup P_m$$

darzustellen.

Beispiel 1.9.2 *Für* $N = \{1, 2, 3\}$ *sind die Partitionen*

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

d.h.

$$N = \{1, 2, 3\}$$
$$N = \{1, 2\} \cup \{3\}$$
$$N = \{1, 3\} \cup \{2\}$$
$$N = \{2, 3\} \cup \{1\}$$
$$N = \{1\} \cup \{2\} \cup \{3\}$$

Beispiel 1.9.3 *Die leere Menge* $N = \varnothing$ *hat als Teilmenge nur* $\varnothing$ *also keine nichtleere Teilmenge. Somit ist* $P = \varnothing$ *die einzige Partition von* N*: Da* P *keine Elemente enthält, sind trivialerweise alle Elemente* $\neq \varnothing$ *und paarweise disjunkt. Weiter gibt die leere Vereinigung* $\varnothing = N$*.*

Satz 1.9.4 *Es gibt eine bijektive Abbildung zwischen der Menge der Äquivalenzrelationen auf* N *und der Menge der Partitionen von* N*.*

Beweis. Jede Äquivalenzrelation auf der Menge N gibt eine Partition von N in die disjunkten Äquivalenzklassen. Ist umgekehrt $P = \{P_1, ..., P_n\}$ eine Partition von N, dann ist durch

$$x \sim y \Leftrightarrow \exists i \text{ mit } \{x, y\} \subset P_i$$

eine eindeutige Äquivalenzrelation gegeben: Sie ist reflexiv, da jedes x in einem P_i liegt, die Symmetrie ist klar aus der Definition. Zur Transitivität: Ist $\{x,y\} \subset P_i$ und $\{y,z\} \subset P_j$, dann muss $i = j$ sein (da die P_i paarweise disjunkt sind und $y \in P_i \cap P_j$). Somit erhalten wir $\{x,z\} \subset \{x,y,z\} \subset P_i$. ■

Beispiel 1.9.5 *Die Partition*

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

entspricht der Äquivalenzrelation auf $M = \{1,2,3\}$ *definiert durch*

$$\begin{array}{llll} 1 \sim 1 & 2 \sim 2 & 1 \sim 2 & 2 \sim 1 \\ 3 \sim 3 & & & \end{array}$$

oder als Relation $R \subset M \times M$ *geschrieben*

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

Definition 1.9.6 *Für* $n,m \in \mathbb{N}_0$ *sei die* ***Stirlingzahl*** *(zweiter Art)* $S(n,m)$ *die Anzahl der Partitionen einer* n*-elementigen Menge in* m *nichtleere Teilmengen. Man schreibt auch* $S(N,m)$ *für die Menge der Partitionen von* N *in* m *Teilmengen.*

Die Anzahl aller Partitionen einer n*-elementigen Menge ist die* ***Bellsche Zahl***

$$B_n = \sum_{m=0}^{n} S(n,m)$$

Beispiel 1.9.7 *Gemäß Beispiel 1.9.2 ist*

$$\begin{aligned} S(3,0) &= 0 \\ S(3,1) &= 1 \\ S(3,2) &= 3 \\ S(3,3) &= 1 \end{aligned}$$

und

$$B_3 = 0 + 1 + 3 + 1 = 5$$

Aus Satz 1.9.4 folgt sofort:

Corollar 1.9.8 *Die Anzahl aller Äquivalenzrelationen auf einer* n*-elementigen Menge ist* B_n*.*

Wie bestimmt man also die Stirlingzahlen $S(n,m)$? Zunächst handeln wir einige Randfälle ab:

Satz 1.9.9 *Es gilt*

1) $S(0,0) = 1$,

2) $S(n,0) = 0$ *für* $n > 0$.

Beweis. Es gilt:

1) Die leere Menge hat genau 1 Partition (siehe Beispiel 1.9.3).

2) Es gibt keine Möglichkeit eine nichtleere Menge in 0 Teilmengen zu partitionieren.

■

Ausgehend davon können wir alle verbleibenden Stirlingzahlen rekursiv berechnen:

Satz 1.9.10 *Für alle* $n < m$ *gilt*

$$S(n,m) = 0$$

und für alle $n \geq m$ *gilt*

$$S(n+1,m+1) = \sum_{k=m}^{n} \binom{n}{k} S(k,m)$$

Beweis. Sie erste Aussage ist klar: Eine n-elementige Menge kann nicht in $m > n$ Teile partitioniert werden.

Zum Beweis der zweiten Aussage zählen wir die Partitionen von $N = \{1,...,n+1\}$ in $m+1$ Teilmengen ab. Dazu zählen wir für jedes $0 \leq k \leq n$ die Partitionen, in denen genau k der Elemente von N nicht in derselben Teilmenge wie $n+1$ liegen. Eine solche Partition $P = \{P_1,...,P_{m+1}\}$ können wir wie folgt konstruieren:

1) Wähle eine k-elementige Teilmenge $M \subset \{1,...,n\}$. Dafür gibt es $\binom{n}{k}$ Möglichkeiten.

2) Setze $P_{m+1} = N \backslash M$. Dann ist $n+1 \in P_{m+1}$.

3) Partitioniere M in m Teilmengen

$$M = P_1 \cup ... \cup P_m.$$

Dafür gibt es $S(k,m)$ Möglichkeiten.

Jede dieser Wahlen liefert eine andere Partition

$$N = P_1 \cup ... \cup P_m \cup P_{m+1}$$

von N und wir erhalten alle Partitionen auf diese Weise. Für festes k gibt es also

$$\binom{n}{k} \cdot S(k,m)$$

Partitionen. Die Summe über alle k ist die Gesamtzahl aller Partitionen. Summanden mit $k < m$ tragen nicht bei, da dann $S(k,m) = 0$. ■

Beispiel 1.9.11 *Wir illustrieren den Beweis an einem Beispiel. Sei z.B.* $n+1 = 4$ *und* $m+1 = 3$*, betrachte also Partitionen von* $N = \{1,2,3,4\}$ *in* 3 *Teilmengen. Der Beweis sortiert die Partitionen nach der Zahl* k *der Elemente von* N*, die nicht in derselben Menge wie* 4 *liegen. Für* M *und damit* P_3 *haben wir folgende Möglichkeiten:*

k	2			3
M	$\{2,3\}$	$\{1,3\}$	$\{1,2\}$	$\{1,2,3\}$
P_3	$\{1,4\}$	$\{2,4\}$	$\{3,4\}$	$\{4\}$

Hier ist $k < 2$ *nicht möglich, da sich dann* M *nicht in* $m = 2$ *Mengen partitionieren lässt. Im Fall* $k = 2$ *existieren* $\binom{3}{2} = 3$ *Wahlen für* M*, im Fall* $k = 3$ *gibt es nur* $\binom{3}{3} = 1$ *Möglichkeit.*

Im Fall $k = 2$ *existieren* $S(2,2) = 1$ *Partitionen von* M *in* 2 *Teilmengen, für* $k = 3$ *gibt es* $S(3,2) = 3$ *solche Partitionen:*

k	2			3		
P_1, P_2	$\{2\},\{3\}$	$\{1\},\{3\}$	$\{1\},\{2\}$	$\{1,2\},\{3\}$	$\{1,3\},\{2\}$	$\{2,3\},\{1\}$
P_3	$\{1,4\}$	$\{2,4\}$	$\{3,4\}$	$\{4\}$		

Insgesamt erhalten wir also die folgenden $S(4,3) = 6$ *Partitionen*

k	2	3
P	$\{\{2\},\{3\},\{1,4\}\}$	$\{\{1,2\},\{3\},\{4\}\}$
	$\{\{1\},\{3\},\{2,4\}\}$	$\{\{1,3\},\{2\},\{4\}\}$
	$\{\{1\},\{2\},\{3,4\}\}$	$\{\{2,3\},\{1\},\{4\}\}$

Beispiel 1.9.12 *Mit Satz 1.9.10 können wir durch rekursives Anwenden der Formel beliebige Stirlingzahlen berechnen, z.B. erhalten wir (entsprechend dem vorherigen Beispiel)*

$$\begin{aligned} S(4,3) &= \binom{3}{2} \cdot S(2,2) + \binom{3}{3} \cdot S(3,2) \\ &= 3 \cdot S(2,2) + 1 \cdot S(3,2) \end{aligned}$$

ebenso

$$\begin{aligned} S(3,2) &= \binom{2}{1} \cdot S(1,1) + \binom{2}{2} \cdot S(2,1) \\ &= 2 \cdot S(1,1) + 1 \cdot S(2,1) \\ S(2,1) &= \binom{1}{0} \cdot S(0,0) + \binom{1}{1} \cdot S(1,0) \\ &= 1 \cdot S(0,0) + 1 \cdot S(1,0) \end{aligned}$$

und $S(2,2) = S(1,1) = S(0,0) = 1$ *(was aber auch direkt aus der Definition klar ist). Somit ist (mit Satz 1.9.9)*

$$\begin{aligned} S(2,1) &= 1 + 0 = 1 \\ S(3,2) &= 2 + 1 = 3 \\ S(4,3) &= 3 + 3 = 6 \end{aligned}$$

In dem MAPLE*-Paket* `combinat` *ist die Berechnung der Stirlingzahlen implementiert in der Funktion* `stirling2`*:*

```
with(combinat);
stirling2(0,0);
1
stirling2(4,3);
6
```

Aus Satz 1.9.10 erhalten wir auch eine Rekursionsformel für die Bellschen Zahlen (zum Beweis siehe Übung 1.27):

Corollar 1.9.13 *Für die Bellschen Zahlen* B_n *gilt* $B_0 = 1$ *und*

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

für alle $n \geq 0$.

Beispiel 1.9.14 *Es gilt*

$$\begin{aligned} B_1 &= B_0 = 1 \\ B_2 &= B_0 + B_1 = 2 \\ B_3 &= B_0 + 2B_1 + B_2 = 5 \end{aligned}$$

In MAPLE *können wir die Bellschen Zahlen folgendermaßen berechnen:*

```
with(combinat);
seq(bell(j),j=0..10);
1, 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975
```

Im Folgenden wollen wir noch effizientere Methoden zur Berechnung der Stirlingzahlen entwickeln. Zunächst eine Rekursionsgleichung mit nur 2 Summanden (für den Beweis siehe Übung 1.25):

Satz 1.9.15 *Für die Stirlingzahlen gilt*

$$S(n+1, m+1) = S(n, m) + (m+1) \cdot S(n, m+1)$$

für alle $n, m \geq 0$.

Beispiel 1.9.16 *Wir berechnen damit die Stirlingzahlen*

$$\begin{aligned} S(3,0) &= 0 \\ S(3,1) &= S(2,1) = S(1,1) = 1 \\ S(3,2) &= S(2,1) + 2 \cdot S(2,2) = 1 + 2 \cdot 1 = 3 \\ S(3,3) &= S(2,2) = S(1,1) = 1 \end{aligned}$$

entsprechend den Partitionen in Beispiel 1.9.2.

Bemerkung 1.9.17 *Ausgehend von der Formel*

$$S(3,2) = S(2,1) + 2 \cdot S(2,2)$$

erhalten wir folgenden Algorithmus zum Aufzählen aller Partitionen von $\{1,2,3\}$ *in* 2 *Teilmengen:*

- *Bestimme alle Partitionen von* $\{1,2\}$ *in* 1 *Menge*

$$\{\{1,2\}\}$$

und füge $\{3\}$ *hinzu:*

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

- *Bestimme alle Partitionen von* $\{1,2\}$ *in* 2 *Mengen*

$$\{\{1\},\{2\}\}$$

und füge 3 *auf alle möglichen Weisen zu einem der Partitionselemente hinzu:*

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

Insgesamt erhalten wir:

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

In Verallgemeinerung davon liefert der kombinatorische Beweis der Rekursionsgleichung in Satz 1.9.15 einen rekursiven Algorithmus zur Bestimmung aller Partitionen einer endlichen Menge N in m Teilmengen. Die Rekursion endet in einem der Fälle von Satz 1.9.9. Zur Implementierung siehe Aufgabe 1.26.

Abschließend beweisen wir noch eine geschlossene Formel für die Stirlingzahlen, indem wir Partitionen mit surjektiven Abbildungen in Beziehung setzen. Dazu leiten wir zunächst eine Formel für die Anzahl der geordneten Partitionen her:

Nach unserer Definition gibt die Stirlingzahl $S(n,m)$ die Anzahl der Möglichkeiten an, aus n unterscheidbaren Geschenke, m Päckchen zu packen. Wir können aber auch fragen, wieviele Möglichkeiten es gibt, n unterscheidbare Geschenke auf m Personen zu verteilen. Dazu müssen wir $\{P_1,...,P_m\}$ nicht als Menge sondern als Liste auffassen:

Definition 1.9.18 *Eine* ***geordnete Partition*** *einer Menge N ist eine Liste $P = (P_1, ..., P_m)$ von Teilmengen $P_i \subset N$, sodass $\{P_1, ..., P_m\}$ eine Partition von N ist.*

Beispiel 1.9.19 *Für $N = \{a, b, c\}$ und $m = 2$ gibt es 3 Partitionen*

$$\{\{a, b\}, \{c\}\} \qquad \{\{a, c\}, \{b\}\} \qquad \{\{b, c\}, \{a\}\}$$

also Verteilungen der Geschenke a, b, c auf 2 Päckchen.

Dagegegen existieren 6 geordnete Partitionen

$$\begin{array}{lll} (\{a, b\}, \{c\}) & (\{a, c\}, \{b\}) & (\{b, c\}, \{a\}) \\ (\{c\}, \{a, b\}) & (\{b\}, \{a, c\}) & (\{a\}, \{b, c\}) \end{array}$$

d.h. Verteilungen der Geschenke a, b, c auf 2 Personen.

Bemerkung 1.9.20 *Aus jeder Partition $\{P_1, ..., P_m\}$ kann man genau $m!$ verschiedene geordnete Partitionen bilden, nämlich*

$$(P_{f(1)}, ..., P_{f(m)})$$

mit $f \in S_m$.

Satz 1.9.21 *Es gibt*

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n$$

geordnete Partitionen $(P_1, ..., P_m)$ einer n-elementigen Menge in m nichtleere Teilmengen.

Beweis. Sei $|N| = n$. Jede surjektive Abbildung $f : N \to \{1, ..., m\}$ definiert eine geordnete Partition $(P_1, ..., P_m)$ von N in die Mengen

$$P_i := f^{-1}(\{i\}) = \{g \in N \mid f(g) = i\}\,.$$

Die P_i sind disjunkt: Wäre $a \in P_i \cap P_j$ für $i \neq j$, dann $f(a) = i$ und $f(a) = j$, was der Abbildungseigenschaft widerspricht.

Umgekehrt definiert jede geordnete Partition $(P_1, ..., P_m)$ eine surjektive Abbildung $f : N \to \{1, ..., m\}$ durch $f(g) = i$ für $g \in P_i$.

Weiter sind diese Zuweisungen zueinander invers, d.h. geben eine Bijektion zwischen der Menge der surjektiven Abbildungen und der Menge der geordneten Partitionen.

Nach Satz 1.8.1 ist die Anzahl der surjektiven Abbildungen $N \to \{1, ..., m\}$ gleich

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n.$$

■

Mit Bemerkung 1.9.20 erhalten wir als Corollar zu Satz 1.9.21 die gesuchte geschlossene Formel für die Stirlingzahlen:

Corollar 1.9.22 *Es gilt*

$$S(n,m) = \frac{1}{m!} \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n$$

für alle $n, m \in \mathbb{N}_0$.

Beispiel 1.9.23 *Wir illustrieren den Beweis von Satz 1.9.21 an einem Beispiel: Die Partition*

$$\{\{a,b\},\{c\}\}$$

von $M = \{a,b,c\}$ *in* $n = 2$ *Teilmengen entspricht den geordneten Partitionen*

$$(\{a,b\},\{c\}) \qquad (\{c\},\{a,b\})$$

und diese den surjektiven Abbildungen

$$\begin{aligned} \{a,b,c\} &\to \{1,2\} \\ a &\mapsto 1 \\ b &\mapsto 1 \\ c &\mapsto 2 \end{aligned}$$

und

$$\begin{aligned} \{a,b,c\} &\to \{1,2\} \\ a &\mapsto 2 \\ b &\mapsto 2 \\ c &\mapsto 1 \end{aligned}$$

1.10 Partitionen von Zahlen

Im letzten Abschnitt haben wir Partitionen und geordnete Partitionen einer n-elementigen Menge N abgezählt. Nach der Mengendefinition sind die Elemente von N unterscheidbar. Beispielsweise könnte N eine Menge von verschiedenen Geschenken sein, die wir auf Päckchen oder Leute verteilen wollen. Oft hat man aber auch keine Idee, welche Geschenke man kaufen soll und verschenkt Geld. Wieviele Möglichkeiten gibt es also, n Euromünzen auf m Päckchen oder Leute zu verteilen? Bei diesem kombinatorischen Problem macht es keinen Sinn die einzelnen Euromünzen zu unterscheiden. Mathematisch übersetzt sich die Frage (im Päckchenfall) wie folgt:

Definition 1.10.1 *Eine **(Zahl)partition** von $n \in \mathbb{N}_0$ ist eine Darstellung von n als Summe positiver ganzer Zahlen. Dabei sehen wir zwei Gleichungen*

$$n = p_1 + ... + p_m$$

als äquivalent an, wenn sie durch das Kommutativgesetz auseinander hervorgehen.

Wir bezeichnen mit $P(n, m)$ die Anzahl aller Partitionen von n in m Zahlen. Die Anzahl aller Partitionen von n ist

$$P(n) = \sum_{m=0}^{n} P(n, m).$$

Beispiel 1.10.2 *Die Gleichungen*

$$4 = 1 + 3$$

und

$$4 = 3 + 1$$

repräsentieren dieselbe Partition von 4.

Beispiel 1.10.3 *Die Partitionen von $n = 4$ sind*

m	*Partitionen*
1	$4 = 4$
2	$4 = 2 + 2$
	$4 = 3 + 1$
3	$4 = 2 + 1 + 1$
4	$4 = 1 + 1 + 1 + 1$

Somit ist

m	0	1	2	3	4
$P(4,m)$	0	1	2	1	1

also

$$P(4) = 5.$$

Bemerkung 1.10.4 *Analog zum Mengenfall gilt*

$$P(0,0) = 1,$$

denn die leere Summe gibt 0*. Ebenso ist*

$$P(n,0) = 0 \text{ für } n > 0$$
$$P(0,m) = 0 \text{ für } m > 0.$$

Eine Berechnung von $P(n,m)$ aus $S(n,m)$ ist nicht ohne Weiteres möglich. Wir wissen nur:

Bemerkung 1.10.5 *Es gilt*

$$S(n,m) \geq P(n,m),$$

denn jede Mengenpartition

$$N = P_1 \cup ... \cup P_m$$

gibt eine Zahlpartition

$$|N| = |P_1| + ... + |P_m|.$$

Allerdings können wir wie im Mengenfall eine Rekursionsgleichung für die $P(n,m)$ angeben. Dazu bemerken wir zunächst:

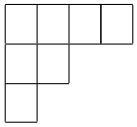
Bemerkung 1.10.6 *In einer Zahlpartition* $n = p_1 + ... + p_m$ *kann man annehmen, dass die* p_i *absteigend sortiert sind. Somit entspricht eine Zahlpartition einer Liste* $(p_1, ..., p_m)$ *mit*

$$n = p_1 + ... + p_m$$

und

$$n \geq p_1 \geq ... \geq p_m > 0.$$

Diese Liste können wir als ***Young-Diagramm*** *der Form*



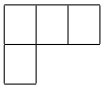
schreiben, wobei in der i-ten Zeile linksbündig p_i Kästchen stehen.

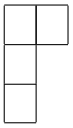
Es ist also $P(n,m)$ die Zahl der Young-Diagramme mit n Kästchen und m Zeilen.

Beispiel 1.10.7 *Die Partitionen von 4 als Young-Diagramm sind*

$4 = 4$ 

$4 = 2 + 2$ 

$4 = 3 + 1$ 

$4 = 2 + 1 + 1$ 

$4 = 1 + 1 + 1 + 1$ 

Satz 1.10.8 *Für $n < m$ ist*

$$P(n,m) = 0$$

und für alle $n \geq m \geq 0$ gilt

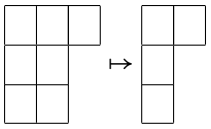
$$P(n+1,m+1) = P(n-m,m+1) + P(n,m).$$

Beweis. Die erste Aussage ist klar. Sei $Y(n,m)$ die Menge der Young-Diagramme mit n Kästchen und m Zeilen. Wir konstruieren eine bijektive Abbildung

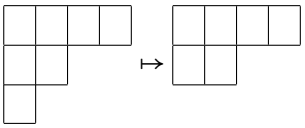
$$\begin{array}{rccc} f: & Y(n+1,m+1) & \to & Y(n-m,m+1) \cup Y(n,m) \\ & P & \mapsto & f(P) \end{array}$$

durch folgende Abbildungsvorschrift: Sei P ein beliebiges Young-Diagramm mit p_i Kästchen in Zeile i. Es gibt zwei Möglichkeiten:

1) Sind alle $p_i \geq 2$, so erhalten wir durch Streichen der ersten Spalte in P ein Young-Diagramm $f(P) \in Y(n-m, m+1)$, d.h. mit $m+1$ Zeilen und $n-m$ Kästchen (aus dem sich das urspüngliche Diagramm durch Hinzufügen der Spalte wieder rekonstruieren lässt):



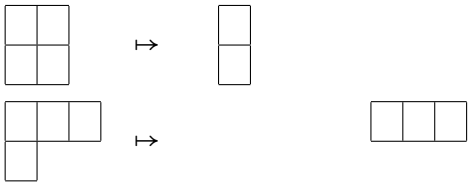
2) Ist $p_{m+1} = 1$, so erhalten wir durch Streichen der letzten Zeile in P ein Young-Diagramm $f(P) \in Y(n, m)$, d.h. mit m Zeilen und n Kästchen (aus dem sich das urspüngliche Diagramm durch Hinzufügen des Kästchens wieder rekonstruieren lässt):



Da $P(n,m) = |Y(n,m)|$ folgt die Behauptung. ■

Beispiel 1.10.9 *Mit dem Satz erhalten wir*

$$P(4,2) \quad = \quad P(2,2) \quad + \quad P(3,1)$$



mit der entsprechenden Korrespondenz von Young-Diagrammen. Ebenso bekommen wir

$$P(2,2) = P(3,1) = 1$$

(was aber auch direkt aus der Definition klar ist) und somit

$$P(4,2) = 1 + 1 = 2.$$

Bemerkung 1.10.10 *Der Beweis des Satzes gibt einen rekursiven Algorithmus zur Bestimmung aller Partitionen von n in m positive Summanden. Siehe dazu Aufgabe 1.31.*

Bemerkung 1.10.11 *Auf einer Gruppe G ist durch $g_1 \sim g_2$ wenn $\exists h \in G$ mit $g_1 h = h g_2$ eine Äquivalenzrelation gegeben. Die Elemente g_1 und g_2 heißen dann **konjugiert**. Man kann zeigen, dass für $G = S_n$ die Äquivalenzklassen (**Konjugationsklassen**) in Bijektion mit den Partitionen von n stehen. Siehe dazu auch die Übungsaufgaben 1.33 und 1.34.*

Wie im Mengenfall bleibt noch die Frage nach der Anzahl der geordneten Partitionen von $n \in \mathbb{N}$. Für Zahlen ist die Antwort wesentlich einfacher. Allerdings besteht keine einfache Beziehung zu $P(n,m)$, denn Permutation der Summanden kann dieselbe geordnete Partition liefern (z.B. für $4 = 2+2$).

Definition 1.10.12 *Eine **geordnete (Zahl)partition** von $n \in \mathbb{N}$ ist eine Liste $P = (p_1, ..., p_m) \in \mathbb{N}^m$ sodass*

$$n = p_1 + ... + p_m.$$

Es ist also $p_i \geq 1$. Im Fall $n = 0$ haben wir wieder wie im ungeordneten Fall die leere Summe. In Übung 1.32 zeigen wir:

Satz 1.10.13 *Für $n, m \in \mathbb{N}$ gibt es genau*

$$\binom{n-1}{m-1}$$

geordnete Partitionen von n in m Zahlen.

Daraus folgt mit Übung 1.9:

Corollar 1.10.14 *Eine Zahl $n \in \mathbb{N}$ hat genau 2^{n-1} geordnete Partitionen.*

Beispiel 1.10.15 *Für $n = 4$ haben wir*

Partitionen	*geordnete Partitionen*	m	$\binom{n-1}{m-1}$
$4 = 4$	(4)	1	1
$4 = 3+1$	$(3,1)$, $(1,3)$	2	3
$4 = 2+2$	$(2,2)$		
$4 = 1+1+2$	$(1,1,2)$, $(1,2,1)$, $(2,1,1)$	3	3
$4 = 1+1+1+1$	$(1,1,1,1)$	4	1

es gibt also insgesamt

$$2^3 = 8$$

geordnete Partitionen von 4.

Was passiert, wenn wir auch $p_i = 0$ zulassen, d.h. wir verteilen $n \in$ auf m Personen, wobei manche auch leer ausgehen dürfen?

Satz 1.10.16 *Für $n, m \in \mathbb{N}$ gibt es genau*

$$\binom{n+m-1}{m-1}$$

Listen $(p_1, ..., p_m) \in \mathbb{N}_0^m$ mit

$$n = p_1 + ... + p_m.$$

Beweis. Jede Summe

$$n = p_1 + ... + p_m$$

mit $p_i \geq 0$ gibt eine Summe

$$n + m = (p_1 + 1) + ... + (p_m + 1)$$

und umgekehrt. Wir haben also eine bijektive Abbildung von geordneten Partitionen

$$\{\text{Partitionen von } n \text{ in } m \text{ mit } 0\} \to \{\text{Partitionen von } n+m \text{ in } m\}$$

Somit folgt die Behauptung aus Satz 1.10.13. ■

Diese Listen nennen wir geordnete Zahlpartitionen von n in m mit 0.

Beispiel 1.10.17 *Wir illustrieren den Beweis für $n = 4$ und $m = 2$:*

geordnete Partitionen von n in m mit 0	*geordnete Partitionen von $n+m$ in m*
$4 = 4 + 0$	$6 = 5 + 1$
$4 = 3 + 1$	$6 = 4 + 2$
$4 = 2 + 2$	$6 = 3 + 3$
$4 = 1 + 3$	$6 = 2 + 4$
$4 = 0 + 4$	$6 = 1 + 5$

Es gibt also

$$\binom{4+2-1}{2-1} = \binom{5}{1} = 5$$

Darstellungen von 4 als geordnete Summe von 2 nichtnegativen Zahlen.

1.11 Multimengen

In vielen Anwendungen wollen wir in einer Menge mehrfache Elemente zulassen. Beispielsweise würden wir die (ungeordnete) Zahlpartition

$$4 = 2 + 1 + 1$$

gerne als eine Menge auffassen, in der 1 zweimal und 2 einmal vorkommt. Der Mengenbegriff erlaubt allerdings keine mehrfachen Elemente, da alle Elemente einer Menge nach Definition unterscheidbar sind. Dies ist auch richtig so, denn wir können solche Multimengen problemlos mit dem herkömmlichen Mengenbegriff modellieren:

Definition 1.11.1 *Eine* ***Multimenge*** $\mathcal{M}$ *ist eine Abbildung* $a : M \to \mathbb{N}_0$. *Man sagt, dass* $x \in M$ *ein* $a(x)$*-faches Element von* $\mathcal{M}$ *ist.*

Für $|M| < \infty$ *ist die Anzahl der Elemente von* $\mathcal{M}$ *definiert als*

$$|\mathcal{M}| = \sum_{x \in M} a(x).$$

Notation 1.11.2 *Ist* $M = \{x_1, ..., x_m\}$, *dann schreiben wir*

$$\mathcal{M} = \{|\underbrace{x_1, ..., x_1}_{a(x_1)}, ..., \underbrace{x_m, ..., x_m}_{a(x_m)}|\}$$

Beispiel 1.11.3 *Die Multimenge* $\{x, y, z\} \to \mathbb{N}_0$, $x \mapsto 2$, $y \mapsto 1$, $z \mapsto 3$ *hat also die Kurzschreibweise*

$$\mathcal{M} = \{|x, x, y, z, z, z|\}.$$

Dabei können wir die Elemente beliebig sortieren, z.B. ist $\{x, y, z\} = \{y, x, z\}$, *also auch*

$$\mathcal{M} = \{|y, x, x, z, z, z|\}.$$

Jede Menge M *kann man auf natürliche Weise als Multimenge mit* $a(m) = 1$ *für alle* $m \in M$ *auffassen.*

Multimengen verhalten sich also genau wie gewöhnliche Mengen, nur dürfen Elemente auch mehrfach vorkommen.

Beispiel 1.11.4 *Multimengen treten bei der Primfaktorisierung von ganzen Zahlen auf. Beispielsweise können wir die Faktorisierung*

$$84 = 2^2 \cdot 3 \cdot 7$$

darstellen als die Multimenge

$$\{|2, 2, 3, 7|\}.$$

Beispiel 1.11.5 *Ebenso kann man natürlich auch für andere Ringe vorgehen, in denen es eine sinnvolle Primfaktorisierung gibt, z.B. für den Polynomring $K[X]$: Die Faktorisierung*

$$f = X^3 - 6X^2 + 9X = X \cdot (X-3)^2$$

lässt sich darstellen als die Multimenge

$$\{|X,\ X-3,\ X-3|\}.$$

Entsprechend bilden auch die Nullstellen von f keine Menge, sondern eine Multimenge

$$\{|0,\ 3,\ 3|\}$$

denn 3 ist ein 2-fache Nullstelle von f.

Die Kombinatorik von Multimengen können wir mit Hilfe des Satzes über geordnete Zahlpartitionen mit 0 beschreiben:

Satz 1.11.6 *Für $|M| = m$ gibt es*

$$\binom{n+m-1}{m-1}$$

Multimengen mit n Elementen aus M.

Beweis. Jede Liste $(p_1, ..., p_m) \in \mathbb{N}_0^m$ mit $p_1 + ... + p_m = n$ entspricht einer Multimenge

$$\mathcal{M} = \{|\ \underbrace{x_1, ..., x_1}_{p_1}, ..., \underbrace{x_m, ..., x_m}_{p_m}|\}$$

mit $n = |\mathcal{M}|$ Elementen und umgekehrt. Somit folgt die Behauptung aus Satz 1.10.16. ■

Beispiel 1.11.7 *Wir illustrieren den Beweis für* $n = 4$ *und* $M = \{x, y\}$*:*

$n = p_1 + p_2$	$\mathcal{M}$
$4 = 4 + 0$	$\{\|x, x, x, x\|\}$
$4 = 3 + 1$	$\{\|x, x, x, y\|\}$
$4 = 2 + 2$	$\{\|x, x, y, y\|\}$
$4 = 1 + 3$	$\{\|x, y, y, y\|\}$
$4 = 0 + 4$	$\{\|y, y, y, y\|\}$

1.12 Systematik im kombinatorischen Zoo

Viele der bisher behandelten praktischen kombinatorischen Fragestellungen lassen sich in das Zählen von Abbildungen oder Äquivalenzklassen von Abbildungen übersetzen. Damit kann man (einem Teil des) umfangreichen Zoos von Abzählproblemen eine Systematik geben. Es gibt $16 = 4 \cdot 4$ naheliegende Möglichkeiten: Wir können beliebige, injektive, surjektive oder bijektive Abbildungen $N \to M$ zwischen endlichen Mengen M und N zählen. Weiter können wir das Zählproblem bis auf Permutation von N oder/und von M betrachten. Im Wesentlichen haben wir schon alle diese Möglichkeiten kennengelernt (z.B. unterscheidbare oder ununterscheidbare Geschenke verteilt auf Päckchen oder Leute).

In Definition 1.5.3 wurde schon die Notation M^N für die Menge aller Abbildungen $f : N \to M$ einführt.

Definition 1.12.1 *Gegeben Mengen* M *und* N*, schreiben wir*

$$\begin{aligned} \mathrm{Inj}(M^N) &= \{f \in M^N \mid f \text{ injektiv}\} \\ \mathrm{Surj}(M^N) &= \{f \in M^N \mid f \text{ surjektiv}\} \\ \mathrm{Bij}(M^N) &= \{f \in M^N \mid f \text{ bijektiv}\} \end{aligned}$$

für die Menge der injektiven, surjektiven bzw. bijektiven Abbildungen.

Als leichte Übung zeigt man:

Proposition 1.12.2 *Auf* M^N *sind durch*

$$\begin{aligned} f \simeq g &\Leftrightarrow \exists \tau \in S(M) \text{ mit } \tau \circ f = g \\ f \approx g &\Leftrightarrow \exists \mu \in S(N) \text{ mit } f \circ \mu = g \\ f \approx g &\Leftrightarrow \exists \mu \in S(N) \text{ und } \tau \in S(M) \text{ mit } \tau \circ f \circ \mu = g \end{aligned}$$

Äquivalenzrelationen gegeben.

Diese Äquivalenzrelationen kann man auf $\mathrm{Inj}(M^N)$, $\mathrm{Surj}(M^N)$ und $\mathrm{Bij}(M^N)$ einschränken.

Beispiel 1.12.3 *Die Abbildungen $f : N \to M$ von $N = \{1,2,3\}$ nach $M = \{a,b\}$ sind gegeben durch:*

$$\begin{pmatrix} 1 & 2 & 3 \\ a & a & a \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ a & a & b \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ a & b & a \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ b & a & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 & 3 \\ b & b & a \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ b & a & b \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ a & b & b \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ b & b & b \end{pmatrix}$$

Es ist also

$$\left|M^N\right| = 2^3 = 8.$$

Die Abbildungen entsprechen den Worten

$$aaa, aab, ..., bbb$$

oder den Möglichkeiten für 3-mal Ziehen aus $\{a,b\}$ mit Zurücklegen mit Beachtung der Reihenfolge.

Beispiel 1.12.4 *Bezüglich $\approx$ gibt es 4 Klassen*

$$\begin{array}{ccccc} & & \begin{pmatrix} 1 & 2 & 3 \\ a & a & a \end{pmatrix} & & \\ \begin{pmatrix} 1 & 2 & 3 \\ a & a & b \end{pmatrix} & \approx & \begin{pmatrix} 1 & 2 & 3 \\ a & b & a \end{pmatrix} & \approx & \begin{pmatrix} 1 & 2 & 3 \\ b & a & a \end{pmatrix} \\ \begin{pmatrix} 1 & 2 & 3 \\ a & b & b \end{pmatrix} & \approx & \begin{pmatrix} 1 & 2 & 3 \\ b & a & b \end{pmatrix} & \approx & \begin{pmatrix} 1 & 2 & 3 \\ b & b & a \end{pmatrix} \\ & & \begin{pmatrix} 1 & 2 & 3 \\ b & b & b \end{pmatrix} & & \end{array}$$

denn es ist

$$\begin{pmatrix} 1 & 2 & 3 \\ a & a & b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ a & b & a \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

und analog für die anderen Äquivalenzen. Somit gilt

$$\left|M^N / \approx\right| = 4.$$

Die Klassen entsprechen den Multimengen

$$\{|a,a,a|\},\{|a,a,b|\},\{|a,b,b|\},\{|b,b,b|\},$$

äquivalent den geordneten Zahlpartitionen mit 0

$$\begin{aligned}3&=3+0\\3&=2+1\\3&=1+2\\3&=0+3.\end{aligned}$$

Wir verteilen also 3 € *auf* 2 *Personen, wobei Personen auch leer ausgehen dürfen.*

Beispiel 1.12.5 *Bezüglich* $\simeq$ *haben wir* 4 *Äquivalenzklassen*

$$\begin{pmatrix}1&2&3\\a&a&a\end{pmatrix}\simeq\begin{pmatrix}1&2&3\\b&b&b\end{pmatrix}$$
$$\begin{pmatrix}1&2&3\\a&a&b\end{pmatrix}\simeq\begin{pmatrix}1&2&3\\b&b&a\end{pmatrix}$$
$$\begin{pmatrix}1&2&3\\a&b&a\end{pmatrix}\simeq\begin{pmatrix}1&2&3\\b&a&b\end{pmatrix}$$
$$\begin{pmatrix}1&2&3\\b&a&a\end{pmatrix}\simeq\begin{pmatrix}1&2&3\\a&b&b\end{pmatrix}$$

denn es ist

$$\begin{pmatrix}1&2&3\\b&b&b\end{pmatrix}=\begin{pmatrix}a&b\\b&a\end{pmatrix}\circ\begin{pmatrix}1&2&3\\a&a&a\end{pmatrix}$$
$$\begin{pmatrix}1&2&3\\a&a&b\end{pmatrix}=\begin{pmatrix}a&b\\b&a\end{pmatrix}\circ\begin{pmatrix}1&2&3\\b&b&a\end{pmatrix}$$

und analog für die anderen Äquivalenzen. Die Klassen stehen in Bijektion zu den Partitionen von $\{1,2,3\}$ *in maximal* 2 *Teilmengen*

$$\begin{aligned}\{1,2,3\}&=\{1,2,3\}\\\{1,2,3\}&=\{1,2\}\cup\{3\}\\\{1,2,3\}&=\{1,3\}\cup\{2\}\\\{1,2,3\}&=\{2,3\}\cup\{1\}.\end{aligned}$$

Die Anzahl ist also

$$\left|M^N/\simeq\right| = S(3,1) + S(3,2) = 1 + 3 = 4.$$

Beispiel 1.12.6 *Schließlich ist*

$$\begin{array}{ccccc}
 & & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & a & a \end{array}\right) & & \\
 & & \simeq & & \\
\left(\begin{array}{ccc} 1 & 2 & 3 \\ a & a & b \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & a \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & a \end{array}\right) \\
\simeq & & & & \simeq \\
\left(\begin{array}{ccc} 1 & 2 & 3 \\ b & b & a \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & b \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & b \end{array}\right)
\end{array}$$

also

$$\left|M^N/\approx\right| = 2.$$

Die Klassen entsprechen den ungeordneten Zahlpartitionen von 3 *mit* 0 *in* 2 *Summanden*

$$3 = 3 + 0$$
$$3 = 2 + 1.$$

Beispiel 1.12.7 *Von allen* 8 *Abbildungen sind folgende Abbildungen* $f : N \to M$ *von* $N = \{1,2,3\}$ *nach* $M = \{a,b\}$ *surjektiv:*

$$\begin{array}{ccc}
\left(\begin{array}{ccc} 1 & 2 & 3 \\ a & a & b \end{array}\right) & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & a \end{array}\right) & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & a \end{array}\right) \\
\left(\begin{array}{ccc} 1 & 2 & 3 \\ b & b & a \end{array}\right) & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & b \end{array}\right) & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & b \end{array}\right)
\end{array}$$

es ist also

$$\left|\mathrm{Surj}(M^N)\right| = 6.$$

Die Abbildungen entsprechen den Möglichkeiten 3 *Geschenke auf* 2 *Personen zu verteilen (wobei niemand leer ausgeht), d.h. den geordneten Partitionen von* N *in* 2 *Teile.*

Beispiel 1.12.8 *Wir haben zwei Klassen*

$$\begin{array}{ccccc}
\left(\begin{array}{ccc} 1 & 2 & 3 \\ a & a & b \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & a \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & a \end{array}\right) \\
\left(\begin{array}{ccc} 1 & 2 & 3 \\ a & b & b \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & a & b \end{array}\right) & \simeq & \left(\begin{array}{ccc} 1 & 2 & 3 \\ b & b & a \end{array}\right)
\end{array}$$

also

$$\left|\mathrm{Surj}(M^N)/\simeq\right| = 2.$$

Die beiden Äquivalenzklassen entsprechen den beiden geordneten Zahlpartitionen

$$3 = 2 + 1$$
$$3 = 1 + 2$$

d.h. wir verteilen $3\,€$ *auf zwei Personen (Person* a *bekommt* $2\,€$ *und Person* b *bekommt* $1\,€$*, und umgekehrt, wobei niemand leer ausgeht).*

Beispiel 1.12.9 *Wir haben* 3 *Äquivalenzklassen*

$$\begin{pmatrix} 1 & 2 & 3 \\ a & a & b \end{pmatrix} \simeq \begin{pmatrix} 1 & 2 & 3 \\ b & b & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 & 3 \\ a & b & a \end{pmatrix} \simeq \begin{pmatrix} 1 & 2 & 3 \\ b & a & b \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 & 3 \\ a & b & b \end{pmatrix} \simeq \begin{pmatrix} 1 & 2 & 3 \\ b & a & a \end{pmatrix}$$

Somit gilt

$$\left|\mathrm{Surj}(M^N)/\simeq\right| = 3.$$

Die drei Äquivalenzklassen entsprechen den (ungeordneten) Mengenpartitionen

$$\{1,2,3\} = \{1,2\} \cup \{3\}$$
$$\{1,2,3\} = \{1,3\} \cup \{2\}$$
$$\{1,2,3\} = \{2,3\} \cup \{1\},$$

d.h. wir verteilen 3 *Geschenke auf* 2 *Päckchen.*

Beispiel 1.12.10 *Modulo* $\approx$ *sind alle* 6 *Abbildungen äquivalent*

$$\begin{array}{ccccc} \begin{pmatrix} 1 & 2 & 3 \\ a & a & b \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 & 3 \\ a & b & a \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 & 3 \\ b & a & a \end{pmatrix} \\ \simeq & & \simeq & & \simeq \\ \begin{pmatrix} 1 & 2 & 3 \\ b & b & a \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 & 3 \\ b & a & b \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 & 3 \\ a & b & b \end{pmatrix} \end{array}$$

es ist also

$$\left|\mathrm{Surj}(M^N)/\approx\right| = 1.$$

Die einzige Äquivalenzklasse entspricht der (ungeordneten) Zahlpartition

$$3 = 2 + 1,$$

d.h. wir verteilen $3 \in$ *auf zwei Päckchen.*

Für die Betrachtung injektiver Abbildungen müssen wir unser Beispiel modifizieren, da es nach dem Schubfachprinzip keine injektive Abbildung $N = \{1,2,3\} \to M = \{a,b\}$ gibt.

Beispiel 1.12.11 *Die injektiven Abbildungen*

$$N = \{1,2\} \to M = \{a,b,c\}$$

sind

$$\begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ b & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ a & c \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ c & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ b & c \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ c & b \end{pmatrix}$$

entsprechend dem 2-maligen Ziehen aus $\{a,b,c\}$ *ohne Zurücklegen unter Beachtung der Reihenfolge. Es ist*

$$\left|\mathrm{Inj}(M^N)\right| = 3 \cdot 2 = 6.$$

Beispiel 1.12.12 *Modulo* $\approx$ *haben wir*

$$\begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix} \approx \begin{pmatrix} 1 & 2 \\ b & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ a & c \end{pmatrix} \approx \begin{pmatrix} 1 & 2 \\ c & a \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 \\ b & c \end{pmatrix} \approx \begin{pmatrix} 1 & 2 \\ c & b \end{pmatrix}$$

entsprechend dem 2-maligen Ziehen aus $\{a,b,c\}$ *ohne Zurücklegen ohne Beachtung der Reihenfolge, d.h. Lotto. Es ist also*

$$\left|\mathrm{Inj}(M^N)/\approx\right| = \binom{3}{2} = 3.$$

Beispiel 1.12.13 *Modulo* $\simeq$ *und damit auch modulo* $\approx$ *sind alle Abbildungen äquivalent*

$$\begin{array}{ccc}
\begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 \\ b & a \end{pmatrix} \\
\simeq & & \simeq \\
\begin{pmatrix} 1 & 2 \\ a & c \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 \\ c & a \end{pmatrix} \\
\simeq & & \simeq \\
\begin{pmatrix} 1 & 2 \\ b & c \end{pmatrix} & \simeq & \begin{pmatrix} 1 & 2 \\ c & b \end{pmatrix}
\end{array}$$

denn es gilt

$$\begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix} = \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix} \circ \begin{pmatrix} 1 & 2 \\ b & c \end{pmatrix}$$

und analog für die anderen Äquivalenzen. Dies ist auch allgemein so, also

$$\left|\mathrm{Inj}(M^N)/\simeq\right| = \left|\mathrm{Inj}(M^N)/\approx\right| = 1.$$

falls $\mathrm{Inj}(M^N) \neq \varnothing$.

Wir fassen diese Ideen allgemein zusammen:

Satz 1.12.14 *Seien* N *und* M *Mengen,* $n = |N|$ *und* $m = |M|$. *Die Mengen* A *von beliebigen, injektiven, surjektiven bzw. bijektiven Abbildungen* $N \to M$ *und deren Mengen von Äquivalenz-*

klassen $A/\eqsim$, $A/\simeq$ und $A/\approx$ lassen sich wie folgt interpretieren:

A	A	$A/\eqsim$	$A/\simeq$	$A/\approx$
M^N	*Worte oder ziehe n aus m mit Zurücklegen mit Reihenfolge*	*Multimengen oder geordnete Zahlpart. von n in m mit $p_i \geq 0$*	*Part. von N in maximal m Mengen*	*Zahlpart. von n in m mit $p_i \geq 0$*
$\mathrm{Inj}(M^N)$	*Ziehe n aus m ohne Zurücklegen mit Reihenfolge*	*Lotto, d.h. ziehe n aus m ohne Zurücklegen ohne Reihenfolge*	*1 Element falls $n \leq m$ sonst $\varnothing$*	*1 Element falls $n \leq m$ sonst $\varnothing$*
$\mathrm{Surj}(M^N)$	*geordnete Part. von N in m*	*geordnete Zahlpart. von n in m mit $p_i \geq 1$*	*Part. von N in m*	*Zahlpart. von n in m mit $p_i \geq 1$*
$\mathrm{Bij}(M^N)$	*Permutationen falls $n = m$ sonst $\varnothing$*	*1 Element falls $n = m$ sonst $\varnothing$*	*1 Element falls $n = m$ sonst $\varnothing$*	*1 Element falls $n = m$ sonst $\varnothing$*

Damit gelten die folgenden Formeln für ihre Anzahl von Elementen:

A	$\lvert A\rvert$	$\lvert A/\eqsim\rvert$	$\lvert A/\simeq\rvert$	$\lvert A/\approx\rvert$
M^N	m^n	$\binom{n+m-1}{n}$	$\sum_{k=0}^{m} S(n,k)$	$P(n+m,m)$
$\mathrm{Inj}(M^N)$	$\prod_{i=0}^{n-1}(m-i)$	$\binom{m}{n}$	1 *für $n \leq m$* 0 *sonst*	1 *für $n \leq m$* 0 *sonst*
$\mathrm{Surj}(M^N)$	$m! \cdot S(n,m)$	$\binom{n-1}{m-1}$	$S(n,m)$	$P(n,m)$
$\mathrm{Bij}(M^N)$	$m!$ *für $n = m$* 0 *sonst*	1 *für $n = m$* 0 *sonst*	1 *für $n = m$* 0 *sonst*	1 *für $n = m$* 0 *sonst*

Beweis. Ohne Einschränkung ist $N = \{1, ..., n\}$ und $M = \{1, ..., m\}$. Wir behandeln die 16 Fälle spaltenweise:

1) Ohne Äquivalenzrelation:

   (a) Abbildungen: Interpretation und Anzahl folgt aus Bemerkung 1.6.4 und Satz 1.6.6.

(b) Injektive Abbildungen: Der Beweis von Satz 1.7.1 gibt die Interpretation und die Formel.

(c) Surjektive Abbildungen: Satz 1.8.1 und Satz 1.9.21.

(d) Bijektive Abbildungen: Corollar 1.7.3.

2) Modulo $\approx$ (Permutation von N):

(a) Abbildungen: Durch eine Abbildung $f : N \to M$ erhalten wir eine disjunkte Vereinigung

$$N = \bigcup_{m \in M} f^{-1}(\{m\})$$

wobei $|f^{-1}(m)| \geq 0$. Durch Permutation von N können wir annehmen, dass

$$\begin{aligned} f^{-1}(\{1\}) &= \{1, 2, ..., p_1\} \\ f^{-1}(\{2\}) &= \{p_1 + 1, ..., p_2\} \\ &\vdots \\ f^{-1}(\{m\}) &= \{p_{m-1} + 1, p_{m-1} + 2, ..., p_m\} \end{aligned}$$

Somit gibt die Klasse von f modulo $\approx$ eine eindeutige geordnete Summe

$$n = p_1 + ... + p_m$$

mit $p_i \geq 0$. Nach Satz 1.11.6 gibt es $\binom{n+m-1}{m-1}$ solche Tupel $(p_1, ..., p_m)$.

(b) Injektive Abbildungen: Modulo $\approx$ ist jede injektive Abbildung durch ihr Bild festgelegt. Um dieses auszuwählen haben wir $\binom{m}{n}$ Möglichkeiten (Definition 1.2.1).

(c) Surjektive Abbildungen: Wie $2(a)$, jedoch ist für surjektives f jedes $p_i \geq 1$. Somit erhalten wir eine geordnete Zahlpartition. Nach Satz 1.10.13 ist die Anzahl solcher Partitionen $\binom{n-1}{m-1}$.

(d) Bijektive Abbildungen: Durch Permutation von N können wir erreichen, dass $f(i) = i \ \forall i$. Alle Abbildungen f liegen also in derselben Äquivalenzklasse.

3) Modulo $\simeq$ (Permutation von M):

(a) Abbildungen: Jedes $f : N \to M$ liefert eine (ungeordnete) Partition

$$N = \bigcup_{m \in f(N)} f^{-1}(\{m\})$$

in $k := |f(N)|$ Teilmengen, wobei f und g dieselbe Partition liefern genau dann wenn $f \simeq g$. Die Elemente von $M^N/\simeq$ entsprechen also genau den Partitionen von N in k Teilmengen für $k = 0, ..., m$. Nach Definition 1.9.6 gibt es für festes k genau $S(n, k)$ solche Partitionen, insgesamt also

$$\sum_{k=0}^{m} S(n, k).$$

(b) Injektive Abbildungen: Nach dem Schubfachprinzip gibt es eine injektive Abbildung $f : N \to M$ nur wenn $n \leq m$. Durch Permutation von M können wir erreichen, dass $f(i) = i$ für alle $i \in N$. Alle solchen Abbildungen liegen also in derselben Äquivalenzklasse.

(c) Surjektive Abbildungen: Corollar 1.9.22.

(d) Bijektive Abbildungen: Folgt aus $3(c)$, da jede bijektive Abbildung auch injektiv ist.

4) Modulo $\approx$ (Permutation von N und M):

(c) Surjektive Abbildungen: Wie $2(c)$, nur erhalten wir durch zusätzliche Permutation von M eine ungeordnete Zahlpartition. Nach Satz 1.10.1 gibt es $P(n, m)$ solche Partitionen.

(a) Abbildungen: Folgt aus $4(c)$, da jede Gleichung

$$n = p_1 + ... + p_m$$

mit $p_i \geq 0$ einer Gleichung

$$n + m = (p_1 + 1) + ... + (p_m + 1)$$

entspricht.

(b) Injektive Abbildungen: Folgt aus $3(b)$, da

$$f \simeq g \Rightarrow f \approx g.$$

(d) Bijektive Abbildungen: Folgt aus $3(d)$, da

$$f \simeq g \Rightarrow f \approx g.$$

■

1.13 Übungsaufgaben

Übung 1.1 *Verwenden Sie den Induktionsbeweis der Formel*

$$\left|2^M\right| = 2^{|M|}$$

für endliche Mengen M, um alle Teilmengen von $M = \{1, 2, 3, 4\}$ aufzuzählen.

Übung 1.2 *1) Zeigen Sie für endliche Mengen M und N, dass*

$$|M \cup N| = |M| + |N| - |M \cap N|$$

und

$$|M \times N| = |M| \cdot |N|$$

2) Gegeben drei Mengen M, N und L, stellen Sie eine Formel für $|M \cup N \cup L|$ auf, und beweisen Sie diese.

Übung 1.3 *Geben Sie je ein Beispiel für eine Abbildung $\mathbb{N} \to \mathbb{N}$ an, die*

1) injektiv aber nicht surjektiv ist.

2) surjektiv aber nicht injektiv ist.

Übung 1.4 *Seien M, N endliche Mengen mit $|M| = |N|$ und $f : M \to N$ eine Abbildung. Zeigen Sie, dass folgende Aussagen äquivalent sind:*

1) f ist bijektiv,

2) f ist injektiv,

3) f ist surjektiv.

Übung 1.5 *Zeigen Sie: Die Komposition von Abbildungen ist assoziativ, das heißt für Abbildungen*

$$M \xrightarrow{f} N \xrightarrow{g} L \xrightarrow{h} K$$

gilt

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Sei $M = \{0,1\}^n$ die Menge aller n-Bit Binärzahlen und $0 \leq k \leq n$. Wieviele Elemente von M enthalten genau k-mal die 1?

Übung 1.6 *Seien $M, N, L \neq \varnothing$ Mengen und $f : M \to N$ und $h : N \to L$ Abbildungen. Zeigen Sie:*

1) Sind f und h injektiv, dann ist auch $h \circ f$ injektiv.

2) Sind f und h surjektiv, dann ist auch $h \circ f$ surjektiv.

3) f ist injektiv genau dann, wenn es eine Abbildung $g : N \to M$ gibt mit $g \circ f = \mathrm{id}_M$.

4) f ist surjektiv genau dann, wenn es eine Abbildung $g : N \to M$ gibt mit $f \circ g = \mathrm{id}_N$.

Übung 1.7 *Seien M, N endliche Mengen und $f : M \to N$ eine Abbildung.*

1) Ist f injektiv, dann gilt $|M| \leq |N|$.

2) Ist f surjektiv, dann gilt $|M| \geq |N|$.

3) Ist f bijektiv, dann gilt $|M| = |N|$.

Übung 1.8 *Zeigen Sie, dass sich n^5 als Linearkombination der Binomialkoeffizienten $\binom{n}{0}, \ldots, \binom{n}{5}$ schreiben lässt, d.h. finden sie $c_j \in \mathbb{Q}$ mit*

$$n^5 = \sum_{j=0}^{5} c_j \binom{n}{j}$$

für alle $n \in \mathbb{N}_0$. Folgern Sie, dass $n^5 - n$ für $n \in \mathbb{N}_0$ stets durch 30 teilbar ist.

Übung 1.9 *Zeigen Sie für alle $n \in \mathbb{N}_0$:*

1) $\sum_{j=0}^{n} \binom{n}{j} = 2^n$

2) $\sum_{j=0}^{n} \binom{n}{j}^2 = \binom{2n}{n}$

Übung 1.10 *Implementieren Sie die Berechnung der Binomialkoeffizienten $\binom{n}{k}$ für $n, k \in \mathbb{N}_0$ mittels der Rekursionsformel*

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

Übung 1.11 *Sie K ein Körper. Folgern Sie aus dem Binomialsatz, dass für alle $x, y \in K$ und $n \in \mathbb{N}_0$ gilt*

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

Übung 1.12 *In einem amerikanischen Stadtplan mit $n+1$ Avenues und $m+1$ Streets (siehe Abbildung 1.5) wollen wir von Punkt A nach Punkt B gehen. Wieviele kürzeste Wege gibt es?*

Beweisen Sie die Formel mit vollständiger Induktion nach $n+m$.

Übung 1.13 *Sei K ein Körper und $c \in K$. Zeigen Sie, dass für alle Polynome $p, q \in K[X]$ gilt*

$$(p \cdot q)(c) = p(c) \cdot q(c) \qquad (p+q)(c) = p(c) + q(c).$$

Übung 1.14 *Sei K ein Körper. Zeigen Sie, dass die Menge der Polynome $K[X]$ zusammen mit der in Definition und Satz 1.2.16 definierten Addition und Multiplikation ein kommutativer Ring mit 1 ist.*

Übung 1.15 *Implementieren Sie Addition und Multiplikation für die dicht besetzte Darstellung von Polynomen $f = a_0 + a_1 X^1 + \ldots + a_n X^n \in \mathbb{Q}[x]$ als Liste $(a_0, \ldots, a_n)$ mit $a_n \neq 0$.*

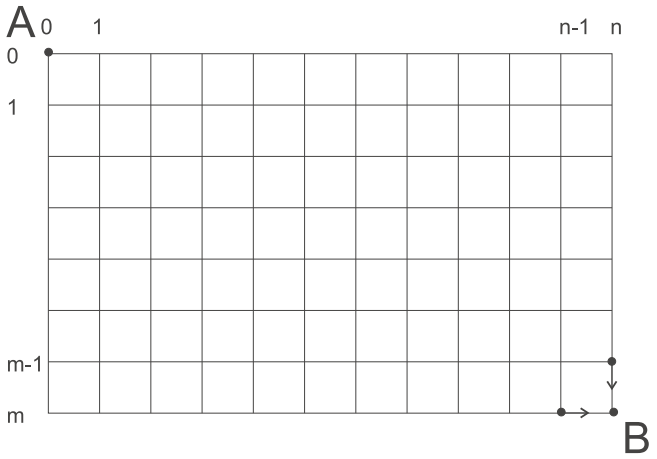


Abbildung 1.5: Wieviele kürzeste Wege gibt es von A nach B.

Übung 1.16 *1) Bestimmen Sie mit Hilfe der Siebformel die Anzahl der ganzen Zahlen* $1 \leq n \leq 1000000$*, die durch* $2, 3, 5$ *oder* 7 *teilbar sind.*

2) Schreiben Sie ein Programm, das für gegebenes N *mittels Division mit Rest die Anzahl aller durch* $2, 3, 5$ *oder* 7 *teilbaren Zahlen* $1 \leq n \leq N$ *bestimmt. Überprüfen Sie Ihr Ergebnis aus (1).*

Hinweis: Sie können die MAPLE*-Funktion* `irem` *verwenden.*

Übung 1.17 *1) Bestimmen Sie mit Hilfe der Siebformel die Anzahl der geraden Zahlen* $1 \leq n \leq 100000$*, die durch* $3, 5, 7$ *oder* 11 *teilbar sind.*

2) Schreiben Sie ein Programm, das für gegebenes N *mittels Division mit Rest die Anzahl aller durch* $3, 5, 7$ *oder* 11 *teilbaren geraden Zahlen* $1 \leq n \leq N$ *bestimmt. Überprüfen Sie Ihr Ergebnis aus (1).*

Übung 1.18 *1) Ein zerstreuter Professor hat* 4 *verschiedene Briefe geschrieben, zugeklebt, aber nicht adressiert. Nun schreibt er zufällig die* 4 *Adressaten auf die Umschläge. Wie groß ist die Wahrscheinlichkeit, dass keiner der Empfänger den für ihn bestimmten Brief bekommt?*

2) Bestimmen Sie die Anzahl aller fixpunktfreien Permutationen einer n-elementigen Menge, d.h. die Anzahl der bijektiven Abbildungen $f : \{1, ..., n\} \to \{1, ..., n\}$ mit

$$f(x) \neq x \quad \text{für alle } x \in \{1, ..., n\}.$$

Hinweis: Siebformel.

Übung 1.19 *1) Der Eintrittspreis für ein Kino sei 10 €. Die Kinokasse wurde gerade geleert und es warten noch 6 Leute, 2 davon haben genau einen 20 € Schein und 4 genau einen 10 € Schein. Wieviele Möglichkeiten gibt es, eine Warteschlange zu bilden, sodass der Kassierer stets genügend Wechselgeld hat?*

2) In einem Stadtplan mit $n+1$ Avenues und $m+1$ Streets (siehe Abbildung 1.6) wollen wir von Punkt A nach Punkt B gehen. In dem Gebiet unterhalb der Winkelhalbierenden treiben Straßengangs ihr Unwesen (Punkte auf der Winkelhalbierenden sind also noch sicher). Zeigen Sie, dass es für $n \geq m$ genau

$$\binom{n+m}{n} - \binom{n+m}{n+1}$$

sichere kürzeste Wege von A nach B gibt.

Übung 1.20 *1) Schreiben Sie ein rekursives Programm, das alle kürzesten Wege von A nach B in einem Stadtplan mit $n+1$ Avenues und $m+1$ Streets aufzählt (siehe Abbildung*

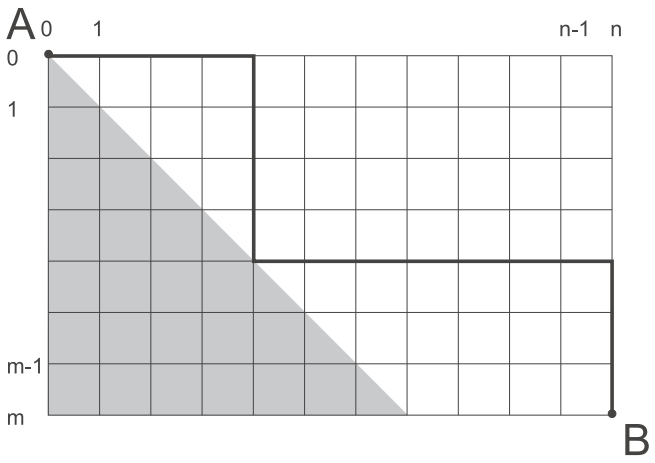
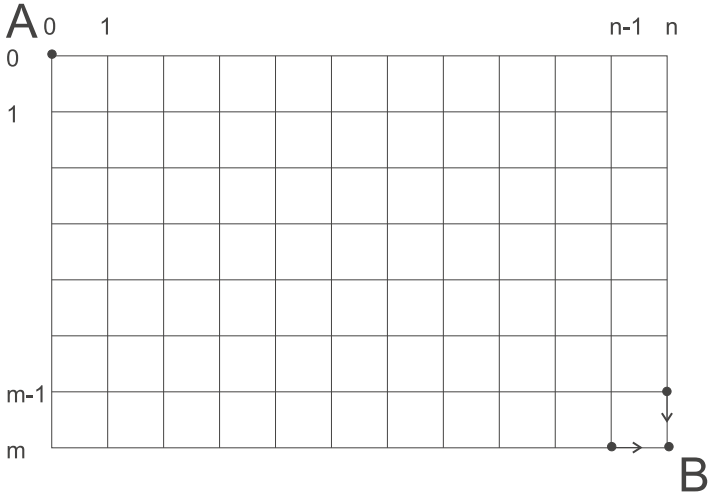


Abbildung 1.6: Kürzeste Wege oberhalb der Winkelhalbierenden.

1.5). Codieren Sie Wege als Listen von Binärziffern.



2) Modifizieren Sie Ihr Programm so, dass es nur Wege aufzählt, die das Gebiet der Straßengangs unterhalb der Winkelhalbierenden vermeiden (siehe Abbildung 1.6).

Übung 1.21 *Für $n \in \mathbb{N}$ sei*

$$\varphi(n) = |\{r \in \mathbb{Z} \mid 1 \leq r \leq n,\ \mathrm{ggT}(r,n) = 1\}|$$

die Anzahl der zu n teilerfremden Zahlen. Sei weiter

$$T(n) = \{p \in \mathbb{N} \mid p \text{ prim und } p \text{ teilt } n\}$$

die Menge der Primteiler von n.

1) Zeigen Sie mit Hilfe der Siebformel, dass für alle n gilt

$$\varphi(n) = n \prod_{p \in T(n)} (1 - \frac{1}{p})$$

2) Erstellen Sie einen Plot von $\varphi : \mathbb{N} \to \mathbb{Z}$, $n \mapsto \varphi(n)$ für $n = 1, ..., 2000$.

Bemerkung: Die Eulersche Phi-Funktion φ spielt eine wichtige Rolle im RSA Public-Key-Kryptosystem.

Übung 1.22 *Welche Elemente der S_4 lassen sich als Symmetrien (Drehungen oder Spiegelungen) des Quadrats (Abbildung 1.7) interpretieren?*

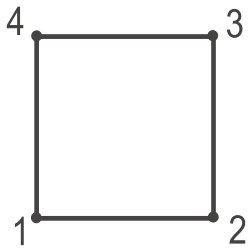


Abbildung 1.7: Quadrat mit Nummerierung der Ecken.

Übung 1.23 *Bestimmen Sie alle Elemente der S_5, die sich geometrisch als Symmetrien (Drehungen oder Spiegelungen) des regelmäßigen Fünfecks (Abbildung 1.8) interpretieren lassen.*

Übung 1.24 *Bei einem Würfelspiel wird der Würfel n-mal geworfen und man gewinnt, wenn dabei alle Zahlen $1, \ldots, 6$ mindestens einmal auftreten.*

1) Wie groß ist die Gewinnwahrscheinlichkeit für $n = 7$?

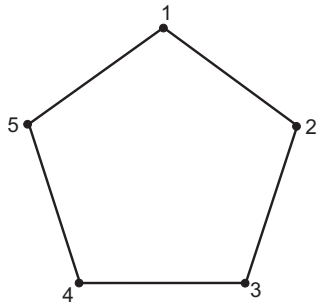


Abbildung 1.8: Regelmäßiges Fünfeck mit Nummerierung der Ecken.

2) Wie groß muss n mindestens gewählt werden, damit die Gewinnwahrscheinlichkeit positiv ist? Welchen Wert nimmt sie dann an?

3) Wie groß darf die Bank n maximal wählen, damit sie eine höhere Gewinnwahrscheinlichkeit als der Spieler hat? Welchen Wert hat die Gewinnwahrscheinlichkeit dann?

4) Überprüfen Sie Ihre Ergebnis aus (1) *anhand einer Stichprobe von* 100000 *Durchläufen des Spiels.*

Hinweis: Sie dürfen dazu den Computer verwenden. Die MAPLE*-Funktion* `rand(n)()` *liefert eine Zufallszahl in* $\{0, \ldots, n-1\}$.

Übung 1.25 *1) Ein zerstreuter Professor will* 5 *verschiedene Geschenke auf* 3 *Päckchen verteilen. Die Päckchen sehen von außen alle gleich aus. Nachdem er alle Möglichkeiten durchprobiert und aufgeschrieben hat, stellt er fest, dass er eines der Geschenke vergessen hat (er hat also nur* 4 *Geschenke auf* 3 *Päckchen verteilt). Wie kann er seinen Fehler korrigieren, ohne nochmals komplett von vorne anzufangen?*

2) Zeigen Sie, dass für die Stirlingzahlen gilt

$$S(n+1, m+1) = S(n, m) + (m+1) \cdot S(n, m+1)$$

für alle $n, m \geq 0$.

3) Bestimmen Sie $S(5,3)$.

Übung 1.26 *Implementieren Sie ein rekursives Verfahren zur Bestimmung aller Partitionen einer n-elementigen Menge in m Teile.*

Hinweis: Verwenden Sie den kombinatorischen Beweis der Formel aus Aufgabe 1.25.2.

Übung 1.27 *1) Bestimmen Sie alle Äquivalenzrelationen auf der Menge* $M = \{1,2,3,4\}$.

2) Zeigen Sie, dass für die Bellschen Zahlen B_n *gilt* $B_0 = 1$ *und*

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

für alle $n \geq 0$.

3) Berechnen Sie B_4.

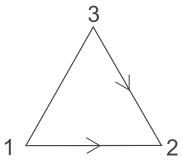
Übung 1.28 *1) Bestimmen Sie alle reflexiven Relationen* $R \subset M \times M$ *auf* $M = \{1,2\}$.

2) Zeigen Sie, dass es auf einer n-elementigen Menge M genau $2^{n(n-1)}$ *reflexive Relationen gibt.*

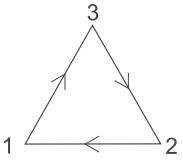
Übung 1.29 *Sei M eine Menge mit n Elementen. Wieviele Totalordnungen gibt es auf M?*

Übung 1.30 *1) In einem Spiel zeichnet man in einem Dreieck auf jeder Kante zufällig einen Pfeil im oder gegen den Uhrzeigersinn oder keinen Pfeil (durch Würfeln mit einem dreiseitigen Würfel). Der Spieler verliert, wenn die Figur mindestens zwei Pfeile enthält und alle Pfeile in dieselbe Richtung zeigen, z.B.*

gewonnen:



verloren:



Wie hoch ist die Gewinnwahrscheinlichkeit?

2) Bestimmen Sie alle Halbordnungen auf $\{1,2,3\}$. Welche sind Totalordnungen?

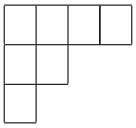
Übung 1.31 *1) Entwickeln Sie einen rekursiven Algorithmus, der für $n,m \in \mathbb{N}_0$ alle Zahlpartitionen von n in m positive Summanden bestimmt.*

Hinweis: Verwenden Sie den Beweis von Satz 1.10.8.

2) Berechnen Sie damit alle Partitionen von 7 in höchstens 3 Summanden.

3) Implementieren Sie Ihren Algorithmus.

Hinweis: Jede Zahlparitition $(p_1,...,p_m)$ können wir als Young-Diagramm der Form



schreiben, wobei in der i-ten Zeile linksbündig p_i Kästchen stehen.

Übung 1.32 *Zeigen Sie, dass es für $n,m \in \mathbb{N}$ genau*

$$\binom{n-1}{m-1}$$

geordnete Zahlpartitionen von n in m positive Summanden gibt.

Übung 1.33 *Durch Nummerieren der Ecken können wir die Symmetriegruppe des Tetraeders (Abbildung 1.9) mit der S_4 identifizieren.*

1) Für festes $f \in S_4$ seien zwei Ecken a und b äquivalent, wenn man durch mehrfaches Anwenden von f die Ecke a auf die Ecke b abbilden kann. Zeigen Sie, dass dadurch eine Äquivalenzrelation $\sim$ auf $\{1,...,4\}$ definiert ist.

2) Bestimmen Sie für jedes $f \in S_4$ die Partition von $\{1,...,4\}$ in Äquivalenzklassen und die entsprechende Zahlpartition $p(f)$ von 4.

Übung 1.34 *Zwei Elemente $f,g \in S_4$ seien äquivalent wenn $p(f) = p(g)$. Bestimmen Sie die Äquivalenzklassen und geben Sie für jede Klasse eine geometrische Interpretation.*

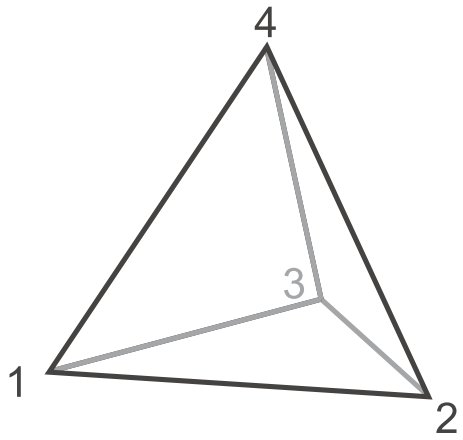


Abbildung 1.9: Tetraeder mit Nummerierung der Ecken

2

Grundlagen der Stochastik

2.1 Übersicht

Die Stochastik oder auch Wahrscheinlichkeitstheorie ist die mathematische Sprache zur Quantifizierung von zufälligen Prozessen. Es geht also darum, Eigenschaften des Resultats eines gegebenen, zufälligen datenerzeugenden Prozesses zu beschreiben. Das kann etwa der Wurf einer Münze, das Lotto-Glücksspiel, oder ein anderer physikalischer Prozess sein. Ihren Beginn hat die Stochastik mit der Arbeit von Mathematikern wie Pascal oder Fermat im 17. Jahrhundert genommen, also viel später als die Algebra im Altertum, aber etwa zur gleichen Zeit wie die Analysis in der Arbeit von Newton und anderen.

In der Informatik ist die Stochastik zentral vertreten bei der Analyse von **randomisierten Algorithmen**. Solche Algorithmen bezeichnet man deshalb auch als **stochastische** oder **probabilistische** Algorithmen. Sogenannte **Las-Vegas-Algorithmen** enthalten eine randomisierte Komponente, liefern aber immer ein korrektes Ergebnis (man kann zulassen, dass sie irgendwann aufgeben). Ein Beispiel sind Sortieralgorithmen wie diverse Varianten von Quicksort. Bei solchen Algorithmen bestimmt man dann nicht die Laufzeit des Algorithmus, sondern nur eine erwartete Laufzeit. Dazu führen wir den Begriff des **Erwartungswerts** ein. Der Erwartungswert ist die mit den jeweiligen Wahrscheinlichkeiten gewichtete Summe der möglichen Ergebnisse.

Beispiel 2.1.1 *Beim Würfeln treten* $1, \ldots, 6$ *jeweils mit der Wahr-*

scheinlichkeit $\frac{1}{6}$ *auf. Der Erwartungswert des Würfelexperiments ist dann*

$$\frac{1}{6}(6+5+4+3+2+1) = \frac{7}{2}.$$

Bei einer zufälligen Auswahl müssen nicht (wie etwa beim Münzwurf oder Lotto) alle Möglichkeiten gleich wahrscheinlich sein. Dies modelliert man mit dem Konzept einer Wahrscheinlichkeitsverteilung. Zum Beispiel erfüllt die Körpergröße einer Menge von Menschen (im Grenzwert) die Verteilung einer Gaußkurve (siehe Abbildung 12). Hier sehen wir, dass man im Allgemeinen nicht nur mit einer diskreten Verteilung arbeiten kann (wie z.B. beim Münzwurf Kopf oder Zahl je mit Wahrscheinlichkeit $\frac{1}{2}$), sondern, dass man auch kontinuierliche Verteilungen braucht. Klar ist, dass hier dann auch Methoden der Analysis verwendet werden müssen.

Ein anderer Typ von randomisierten Algorithmen sind die sogenannten **Monte-Carlo-Algorithmen**. Hier ist es nicht garantiert, dass das Ergebnis korrekt ist. Man will dann neben dem Erwartungswert für die Rechenzeit, auch die Fehlerwahrscheinlichkeit für den Rückgabewert bestimmen oder zumindest abschätzen. Ein Beispiel sind Integrationsverfahren. Bei Entscheidungsproblemen (also Rückgabewert wahr oder falsch) gibt es zwei Fälle:

1) Algorithmen mit einem zweiseitigen Fehler dürfen sowohl false Positives als auch false Negatives berechnen. Ist die Fehlerwahrscheinlicheit nicht kleiner als $\frac{1}{2}$ kann man den Algorithmus verwerfen, denn ein Münzwurf ist genauso gut. Ein klassisches Beispiel eines Monte-Carlo-Algorithmus mit zweiseitigem Fehler ist die Verifikation einer Identität mit einem Fingerabdruck.

2) Bei Algorithmen mit einem einseitigen Fehler haben nur genau eine der beiden Fehlermöglichkeiten. Beispiele sind Primzahltests, oder das Testen von Gleichheit von zwei Polynomausdrücken etwa $f = (x_1 - x_2)^2$ und $g = x_1^2 - 2x_1x_2 + x_2^2$ durch Einsetzen: Wenn für eine gegebene Anzahl von Stützstellen x die Werte $f(x) = g(x)$ übereinstimmen, dann gehen wir davon aus, dass $f = g$. Wenn die Stützstellen nicht ausreichen, um f und g aus den Funktionswerten

zu interpolieren, dann können f und g zufällig dieselben Werte annehmen, obwohl sie nicht gleich sind. Mit beliebig vielen Stützstellen der Form $x = (a, 0)$ und $x = (0, a)$, $a \in \mathbb{R}$ würden wir z.B. $f = (x_1 - x_2)^2$ und $g = x_1^2 + 2x_1x_2 + x_2^2$ für gleich halten, denn sie nehmen auf dem Koordinatenkreuz dieselben Werte an.

Die Beziehung zwischen dem Erwartungswert und den praktischen Anwendungen wird durch das **Gesetz der großen Zahlen** hergestellt: Dies besagt, dass die Wahrscheinlichkeit einer großen Abweichung des (arithmetischen) Mittelwerts von dem Erwartungswert bei einer mehrfachen Durchführung eines Zufallsprozesses gegen 0 geht. Das gibt uns z.B. eine Möglichkeit um experimentell Erwartungswerte approximativ zu finden. Präzise lautet die Formulierung: Ist

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - E)$$

mit den zufälligen Werten X_i mit Erwartungswert E, und bezeichnet

$$P(|\overline{X}_n| \geq \varepsilon)$$

die Wahrscheinlichkeit, dass $|\overline{X}_n| \geq \varepsilon$ ist, dann gilt für jedes $\varepsilon > 0$, dass

$$\lim_{n\to\infty} P(|\overline{X}_n| \geq \varepsilon) = 0.$$

Beispiel 2.1.2 *Würfeln wir* 10*-mal und erhalten die Sequenz von Ergebnissen*

$$6, 3, 5, 3, 1, 1, 3, 4, 2, 1$$

dann nimmt $\overline{X}_{10}$ *den Wert*

$$\begin{aligned}\overline{X}_{10} &= \frac{1}{10}((6 - \frac{7}{2}) + (3 - \frac{7}{2}) + \ldots + (1 - \frac{7}{2})) \\ &= -\frac{6}{10}\end{aligned}$$

an. In MAPLE *können wir diese Rechnung z.B. durchführen mit*

```
N:=10:
X:=[seq(rand(1..6)(),j=1..N)];
X := [6, 3, 5, 3, 1, 1, 3, 4, 2, 1]
```

```
simplify(1/N*sum(X[i]-7/2,i=1..N));
-3/5
```

Machen wir N größer, wird sich das Ergebnis mit hoher Wahrscheinlichkeit 0 annähern. Wichtig ist hier zu verstehen, dass dies nur mit hoher Wahrscheinlichkeit passiert: Wir könnten Pech haben und in unserem Experiment z.B. N-mal die 1 würfeln.

In dem Programm rufen wir mit `rand` *einen* ***Pseudozufallszahlengenerator*** *auf. Hier werden im Computer Zahlen erzeugt, die sich so zufällig wie möglich verhalten (typischerweise sogar reproduzierbar, wenn man einen sogannten* ***random seed*** *festlegt). Zu verstehen wie solche Generatoren funktionieren ist auch eine spannende Fragestellung.*

Das Gesetz der großen Zahlen spielt zum Beispiel eine wichtige Rolle in der Konzeption von Versicherungen. Auch wenn man nicht weiss, wer genau vom Schaden getroffen wird, kann man bei immer größeren Versicherungsgemeinschaften immer sicherer abschätzen, wie häufig ein Schaden auftritt. Auch bei Messungen z.B. in den Natur- oder Ingenieurwissenschaften (etwa bei Beschleunigerexperimenten in der Hochenergiephysik oder auch in der Astronomie) ist das wichtig: Nichtsystematische Messfehler können durch Wiederholung herausgemittelt werden. Wichtig ist hier zu verstehen, was Stochastik nicht leisten kann. Das Gesetz der großen Zahlen ist kein Gesetz des Ausgleichs der Wahrscheinlichkeiten:

Beispiel 2.1.3 *Betrachten wir einen Münzwurf mit Ergebnis 0 und 1 (Kopf und Zahl). Der Erwartungswert ist dann $\frac{1}{2}$. Nach einer Sequenz von*

$$0, 0, 0, 1, 0$$

haben wir die folgenden Häufigkeiten

Kopf	*Zahl*
$\frac{4}{5}$	$\frac{1}{5}$

Das bedeutet nicht, dass nun Kopf Nachholbedarf hätte, was der eine oder andere Glücksspieler denkt. Setzen wir die Sequenz fort, dann erhalten wir vielleicht (mit MAPLE *erzeugt)*

$$0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0$$

und damit die folgenden Häufigkeiten

Kopf	*Zahl*
$\frac{13}{20}$	$\frac{7}{20}$

Wir sehen, dass im Vergleich zu

$$\frac{1}{5} = 0.2$$

die Zahl

$$\frac{7}{20} = 0.35$$

wesentlich näher an dem Erwartungswert 0.5 *liegt. Nichtsdestotrotz ist der Vorsprung von Kopf sogar von* 3 *auf* 6 *Würfe angewachsen. Abweichungen werden also nicht ausgeglichen, sondern werden nur mit der wachsenden Zahl von Experimenten weniger wichtig.*

2.2 Anwendungen

2.2.1 Sortieren

Ein Beispiel für einen Las-Vegas-Algorithmus ist ein randomisiertes Quicksort-Verfahren. Die Grundidee von Quicksort ist einfach: Wir wollen die Elemente einer Menge $M = \{x_1, \ldots, x_n\} \subset \mathbb{Z}$ von ganzen Zahlen sortieren.[1] Wir suchen ein Element $x_i \in M$, sodass die Mengen $M_1 = \{x \in M \mid x < x_i\}$ und $M_2 = \{x \in M \mid x > x_i\}$ in etwa gleich viele Elemente enthalten und wenden das Verfahren induktiv auf diese Mengen an.

Beispiel 2.2.1 *Die Menge*

$$M = \{100, 3, 7, 2, 11, 1, 33\}$$

teilen wir bei $x_3 = 7$ *auf in*

$$M_1 = \{3, 2, 1\} \text{ und } M_2 = \{100, 11, 33\}.$$

[1] Alternativ könnte man auch eine Multimenge betrachten. Als Datenstruktur wird M in der Informatik realisiert als die Liste $(x_1, \ldots, x_r)$.

Iterativ erhalten wir die sortierte Darstellung der Menge

$$M = \{1, 2, 3, 7, 11, 33, 100\}.$$

Zur Aufteilung einer n-elementigen Menge benötigen wir $n-1$ Vergleiche in konstanter Laufzeit. Das wesentliche Problem ist die Bestimmung von x_i. Man könnte nun x_i einfach als das letzte Element x_n nehmen oder zufällig ein Element auswählen, um den Preis, dass M_1 und M_2 nicht gleich mächtig sind. Bei einer zufälligen Auswahl von x_i spricht man vom **randomisierten Quicksort**-Algorithmus. Was können wir über die Laufzeit sagen? Ganz ohne Stochastik erhalten wir die folgende Abschätzung:

Proposition 2.2.2 *Die worst-case Laufzeit des randomisierten Quicksort-Algorithmus auf einer n-elementigen Menge ist in $O(n^2)$.*

Beweis. Wir zeigen mit Induktion, dass es ein $c > 0$ gibt, sodass für die Laufzeit $T(n)$ des Quicksort-Algorithmus gilt $T(n) \leq cn^2$. Wir haben

$$T(n) \leq \max_{0 \leq a \leq n-1} (T(a) + T(n-a-1)) + n - 1$$

wobei der Summand $n-1$ die Vergleiche mit dem zufällig gewählten x_i zählt. Somit ist nach Induktionsvoraussetzung

$$\begin{aligned} T(n) &\leq \max_{0 \leq a \leq n-1} (ca^2 + c(n-a-1)^2) + n - 1 \\ &= c \max_{0 \leq a \leq n-1} (a^2 + (n-a-1)^2) + n - 1. \end{aligned}$$

Die Parabel

$$f(a) = a^2 + (n-a-1)^2$$

hat ihr Minimum bei a mit

$$0 = f'(a) = 4a - 2n + 2$$

d.h. bei

$$a = \frac{n-1}{2}.$$

Wegen

$$0 \leq \frac{n-1}{2} \leq n-1$$

ist für $0 \leq a \leq n-1$ also

$$f(a) \leq f(0) = f(n-1) = n^2 - 2n + 1.$$

Wir haben also

$$\begin{aligned} T(n) &\leq c(n^2 - 2n + 1) + n - 1 \\ &= cn^2 - c(2n-1) + n - 1 \\ &\leq cn^2 \end{aligned}$$

wenn wir c groß genug wählen. ■

Das ist nicht so fantastisch, denn die Laufzeit des folgenden trivialen Sortieralgorithmus ist die gleiche: Im **Selectionsort**-Algorithmus suchen wir in der gegebenen Liste das kleinste Element und tauschen dieses auf den ersten Platz. Dann fahren wir induktiv mit der restlichen Liste fort.

Beispiel 2.2.3 *Für die Menge*

$$M = \{100, 3, 7, 2, 11, 1, 33\}$$

gehen wir wie folgt vor

100	3	7	2	11	1	33
1	3	7	2	11	100	33
1	2	7	3	11	100	33
1	2	3	7	11	100	33
1	2	3	7	11	100	33
1	2	3	7	11	100	33
1	2	3	7	11	33	100

wobei das auf den ersten Platz zu tauschende Element blau und die schon sortierten Elemente rot markiert sind.

Proposition 2.2.4 *Der Selectionsort-Algorithmus hat Laufzeit* $O(n^2)$.

Beweis. Der Algorithmus benötigt

$$(n-1)+(n-2)+\ldots+1=\sum_{i=1}^{n-1} i=\binom{n}{2}=\frac{n^2-n}{2}$$

Vergleiche. ■

Wir können aber mit Hilfe von Stochastik die erwartete Laufzeit des randomisierten Quicksort-Algorithmus bestimmen und erhalten:

Satz 2.2.5 *Der Erwartungswert für die Laufzeit des randomisierten Quicksort-Algorithmus ist in* $O(n\log(n))$.

Dies werden wir mit den in der Stochastik entwickelten Methoden beweisen. Siehe auch Aufgabe 2.1.

Tatsächlich ist der randomisierte Quicksort-Algorithmus der heute bevorzugte Sortieralgorithmus, da er in der Praxis schneller als andere $O(n\log(n))$-Algorithmen ist und dabei sehr einfach zu implementieren. Mergesort hat z.B. auch eine $O(n\log(n))$ Laufzeit, eine bessere worst-case Komplexität, aber eine größere Konstante in der Landaunotation, sodass er in der Praxis langsamer ist.

Bemerkung 2.2.6 *Es gibt auch einen Sortieralgorithmus mit richtig schlechter Laufzeit in* $O(n\cdot n!)$*, der* ***Stupidsort****-Algorithmus. Auch dies ist ein Las-Vegas-Algorithmus, benötigt also zur Laufzeitanalyse nicht-triviale Stochastik.*

2.2.2 Primzahltests

Eine wichtige Klasse von Beispielen von Monte-Carlo-Algorithmen sind Primzahltests. Der Fermat Primzahltest basiert auf dem kleinen Satz von Fermat:

Satz 2.2.7 (Kleiner Satz von Fermat) *Ist* p *eine Primzahl und* $a\in\mathbb{Z}$ *mit* $p\nmid a$*, dann ist*

$$a^{p-1}\equiv 1 \bmod p.$$

Beweis. Es gilt $\left|(\mathbb{Z}/p)^\times\right| = p-1$ und die Ordnung jedes Elements einer Gruppe teilt die Gruppenordnung ■

Wir erinnern: Ein Element $\overline{a} \in \mathbb{Z}/n$ ist invertierbar genau dann, wenn es ein $b \in \mathbb{Z}$ gibt mit

$$\overline{a} \cdot \overline{b} = \overline{1},$$

das heißt, wenn es $b, k \in \mathbb{Z}$ gibt mit

$$a \cdot b + k \cdot n = 1.$$

Solche b und k erhalten wir mit dem erweiterten Euklidischen Algorithmus, falls

$$\mathrm{ggT}\,(a, n) = 1.$$

Haben wir umgekehrt eine solche Darstellung der 1, dann müssen natürlich a und n teilerfremd sein (denn jeder gemeinsame Teiler teilt auch 1). Somit können wir die Elemente der Einheitengruppe von $\mathbb{Z}/n$ beschreiben:

Satz 2.2.8 *Für $n \in \mathbb{N}$ ist*

$$(\mathbb{Z}/n)^\times = \left\{\overline{a} \in \mathbb{Z}/n \mid \mathrm{ggT}\,(a, n) = 1\right\}.$$

Beispiel 2.2.9 *Die Restklasse $\overline{8} \in \mathbb{Z}/15$ hat ein Inverses, d.h. $\overline{8} \in (\mathbb{Z}/15)^\times$, denn*

$$\mathrm{ggT}\,(8, 3 \cdot 5) = 1.$$

Mit dem erweiterten Euklidischen Algorithmus erhalten wir eine Darstellung des größten gemeinsamen Teilers

$$1 = (2) \cdot 8 + (-1) \cdot 15$$

also ist

$$\overline{8}^{-1} = \overline{2}.$$

Aus dem kleinen Satz von Fermat erhalten wir Algorithmus 2.1 um zu testen, ob eine gegebene Zahl $n \in \mathbb{N}$, $n \geq 2$ prim ist.

Falls n prim ist, bricht dieses Verfahren nicht ab, man kann also nur durch mehrfaches Durchlaufen der Schleife (mit verschiedenen a) die Wahrscheinlichkeit erhöhen, dass wir n korrekterweise für prim halten. Es gibt auch Zahlen, bei denen der

Algorithmus 2.1 Fermat Primzahltest

Wir wollen testen, ob $n \in \mathbb{N}$, $n \geq 2$ eine Primzahl ist.

1) Zunächst wählen wir ein $a \in \mathbb{Z}$, $1 < a < n$ und bestimmen $\mathrm{ggT}(a,n)$ mit dem Euklidischen Algorithmus. Falls $\mathrm{ggT}(a,n) \neq 1$, war n nicht prim.

2) Ist $\mathrm{ggT}(a,n) = 1$ (und damit $\overline{a} \in (\mathbb{Z}/n)^\times$ nach Satz 2.2.8), dann testen wir, ob

$$a^{n-1} \equiv 1 \bmod n$$

*Gilt dies nicht, dann kann n nach dem kleinen Satz von Fermat 2.2.7 auch nicht prim gewesen sein. Man bezeichnet dann a (oder seine Klasse modulo n) als **Fermat-Zeugen** für die Zerlegbarkeit von n. Anderenfalls können wir keine Aussage machen und gehen zurück zu (1).*

Test in (2) für kein a mit $\mathrm{ggT}(a,n) = 1$ erkennt, dass sie nicht prim sind, die sogenannten **Carmichael-Zahlen**. Diese erkennt aber Schritt (1) für geeignetes a (was aber natürlich sehr ineffizient ist). Man kann zeigen, dass es unendlich viele Carmichael-Zahlen gibt, der Beweis ist aber nicht einfach und wurde erst 1994 geführt.

Definition 2.2.10 *Eine Zahl n heißt **Fermatsche Pseudoprimzahl** zur Basis a, wenn n nicht prim ist, aber dennoch $a^{n-1} \equiv 1 \bmod n$ gilt.*

Beispiel 2.2.11 *Die Rechnung*

$$2^8 \equiv 4 \bmod 9$$

beweist, dass 9 nicht prim ist.

Dagegen gilt

$$2^{340} \equiv 1 \bmod 341,$$

aber unglücklicherweise ist

$$341 = 11 \cdot 31$$

nicht prim, also 341 *eine Fermatsche Pseudoprimzahl zur Basis* $a = 2$. *Testen wir nochmals zur Basis* $a = 3$ *erhalten wir*

$$3^{340} \equiv 56 \bmod 341$$

und haben damit gezeigt, dass 341 *keine Primzahl ist (überprüfen Sie die Rechnungen mit* MAPLE*).*
Man beachte:
Dies konnten wir erkennen, ohne einen Teiler zu finden.

Glücklicherweise sind Carmichaelzahlen unter den ganzen Zahlen nicht so häufig und lassen sich gut berechnen. Man kann zeigen:

Satz 2.2.12 *Eine zusammengesetze Zahl* $n \in \mathbb{N}$ *ist eine Carmichael-Zahl ist genau dann, wenn für alle Primteiler* p *von* n *gilt, dass*

$$p^2 \nmid n$$

und

$$(p-1) \mid (n-1).$$

Wenn wir die Carmichaelzahlen aussortieren (z.B. da wir diese in einer Liste aufgezählt haben), dann können wir unter zufälliger Wahl von a eine Wahrscheinlichkeit angeben, dass wir n fälschlicherweise für prim halten:

Lemma 2.2.13 *Sei* $n \in \mathbb{N}$ *nicht prim und keine Carmichael-Zahl, dann sind mindestens die Hälfte aller* $\overline{a} \in (\mathbb{Z}/n)^\times$ *Fermat-Zeugen für die Zerlegbarkeit von* n.

Beweis. Die Menge

$$\begin{aligned} A &= \{\overline{a} \in (\mathbb{Z}/n)^\times \mid n \text{ ist Fermatsche Pseudoprimzahl zur Basis } \overline{a}\} \\ &= \{\overline{a} \in (\mathbb{Z}/n)^\times \mid \overline{a}^{n-1} \equiv 1 \bmod n\} \end{aligned}$$

ist offenbar eine Untergruppe von $(\mathbb{Z}/n)^\times$. Für n keine Carmichael-Zahl ist A eine echte Untergruppe (da es einen Zeugen für die Zerlegbarkeit geben muss) und hat somit Index

$$\frac{\left|(\mathbb{Z}/n)^\times\right|}{|A|} \geq 2.$$

■

Nach m Durchläufen mit zufällig gewähltem a ist also die Fehlerwahrscheinlichkeit $\leq \frac{1}{2^m}$.

Einen weiteren Mont-Carlo-Algorithmus mit Bezug zur linearen Algebra untersuchen wir in Übungsaufgabe 2.2.

2.3 Diskrete Wahrscheinlichkeitsverteilungen

2.3.1 Wahrscheinlichkeitsfunktionen

Zunächst werden wir Zufallsexperimente mit einer endlichen oder abzählbar unendlichen Ergebnismenge betrachten. In diesem Fall lassen sich Zufallsexperimente sehr einfach beschreiben. Ein **Zufallsexperiment** wird beschrieben durch eine Wahrscheinlichkeitsfunktion, die jedem möglichen Ergebnis des Experiments eine Wahrscheinlichkeit zuordnet.

Definition 2.3.1 *Sei Ω eine abzählbare Menge (d.h. endlich oder abzählbar unendlich). Eine **Wahrscheinlichkeitsfunktion** auf dem **Ergebnisraum** Ω ist eine Abbildung*

$$m : \Omega \to \mathbb{R}_{\geq 0}$$

*die jedem $\omega \in \Omega$ eine **Wahrscheinlichkeit** $m(\omega)$ zuordnet und für die gilt*

$$\sum_{\omega \in \Omega} m(\omega) = 1.$$

*Den Ergebnisraum Ω zusammen mit m bezeichnen wir auch als **diskreten Wahrscheinlichkeitsraum**.*

Insbesondere fordern wir, dass die Summe $\sum_{\omega \in \Omega} m(\omega)$ konvergiert. Man beachte, dass wir die Summe in der Definition stets als endliche Summe oder als Reihe

$$\sum_{n=1}^{\infty} m(\omega_n)$$

auswerten können (indem wir die Elemente einer abzählbar unendlichen Menge Ω durchnummerieren). Bei der Summe $\sum_{\omega \in \Omega} m(\omega)$

kommt es nicht auf die Reihenfolge an, da alle Summanden ≥ 0 und damit Konvergenz und absolute Konvergenz äquivalent sind.

Wir können unseren Begriff des diskreten Wahrscheinlichkeitsraums noch etwas verallgemeinern: Ist Ω eine beliebige Menge und ist $m(\omega) \neq 0$ nur für abzählbar viele $\omega \in \Omega$, dann funktioniert unsere Definition auch.

Wir werden sehen, dass man den Begriff des Wahrscheinlichkeitsraums im nicht-diskreten Fall (d.h. für Ω nicht-abzählbar) noch etwas verallgemeinern muss.

Schon im Fall Ω endlich gibt es aber viele interessante Anwendungsfälle.

Beispiel 2.3.2 *Für den Wurf eines Würfels ist der Ergebnisraum*

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

*und (falls der Würfel **fair**, d.h. unmanipuliert ist) gilt*

$$m(\omega) = \frac{1}{6}$$

für alle $\omega \in \Omega$.

Siehe auch Aufgabe 2.7 für den abzählbar unendlichen Fall.

Beispiel 2.3.3 *Wir werfen eine Münze bis zum ersten mal Kopf kommt. Das Ergebnis des Experiments sei die Anzahl n der Würfe bis erstmals Kopf erscheint. Der Ergebnisraum ist $\Omega = \mathbb{N} \cup \{\infty\}$, wobei ∞ für das Ergebnis steht, dass immer nur Zahl kommt. Die Wahrscheinlichkeit, dass Kopf im n-ten Wurf zum ersten mal kommt ist*

$$m(n) = \left(\frac{1}{2}\right)^n.$$

Wegen

$$\sum_{n=1}^{\infty} m(n) = \frac{1}{1 - \frac{1}{2}} - 1 = 1$$

müssen wir $m(\infty) = 0$ setzen damit m eine Wahrscheinlichkeitsfunktion wird.

2.3.2 Ereignisse

Oft will man nicht nur die Wahrscheinlichkeit eines einzelnen Ergebnisses wissen, sondern ist daran interessiert mit welcher Wahrscheinlichkeit das Ergebnis in einer gegeben Menge von Ergebnissen liegt.

Definition 2.3.4 *Sei $m : \Omega \to \mathbb{R}_{\geq 0}$ eine Wahrscheinlichkeitsfunktion. Jede Teilmenge $M \subset \Omega$ bezeichnen wir als **Ereignis** und ordnen M die Wahrscheinlichkeit*

$$P(M) = \sum_{\omega \in M} m(\omega)$$

zu.

Beispiel 2.3.5 *Für den Wurf eines Würfels hat das Ereignis*

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

dass eine ungerade Zahl gewürfelt wird, die Wahrscheinlichkeit

$$P(M) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Für ein weitere Beispiele siehe die Übungsaufgaben 2.4 und 2.7.

Bemerkung 2.3.6 *Sei $m : \Omega \to \mathbb{R}_{\geq 0}$ eine Wahrscheinlichkeitsfunktion. Offenbar gilt*

$$P(\Omega) = 1,$$

und

$$P(M) \geq 0$$

für alle Ereignisse $M \subset \Omega$.

Weiter gilt für die Inklusionsbeziehung zwischen Ereignissen

$$M \subset N \subset \Omega \Longrightarrow P(M) \leq P(N)$$

Für das 1-elementige Ereignis $M = \{\omega\}$ mit $\omega \in \Omega$ gilt

$$P(\{\omega\}) = m(\omega).$$

Beispiel 2.3.7 *Für den zweimaligen Wurf einer Münze mit Ergebnis* 0 *und* 1 *(Kopf und Zahl) ist der Ergebnisraum*

$$\Omega = \{0,1\}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$$

und

$$m(\omega) = \frac{1}{4}$$

für alle $\omega \in \Omega$. *Für das Ereignis*

$$M = \{(0,0),(0,1),(1,0)\}$$

dass wenigstens 1*-mal Kopf gewürfelt wird ist die Wahrscheinlichkeit*

$$P(M) = \frac{3}{4}.$$

Beispiel 2.3.8 *Wahrscheinlichkeitsfunktionen können auch implizit gegeben sein. Stehe* B *dafür, dass es morgen bedeckt ist,* R *dafür, dass es regnet, und* S *dafür, dass es sonnig ist. Regen und Sonne sollen gleichwahrscheinlich sein, aber bedecktes Wetter soll* 3*-mal so häufig auftreten wie Sonne. Sei also* $\Omega = \{B,R,S\}$ *und*

$$\begin{aligned} m(B) &= 3 \cdot m(R) = 3 \cdot m(S) \\ 1 &= m(B) + m(R) + m(S) \end{aligned}$$

also

$$1 = m(B) + m(R) + m(S) = 3 \cdot m(R) + m(R) + m(R)$$

also

$$m(R) = \frac{1}{5}$$

und somit

$$\begin{aligned} m(S) &= \frac{1}{5} \\ m(B) &= \frac{3}{5} \end{aligned}$$

Um die Wahrscheinlichkeitsfunktion explizit zu lösen müssen im Fall von linearen Relationen ein lineare Gleichungssystem mit dem Gaußalgorithmus lösen. Eine wohldefinierte Funktion m *erhalten wir, falls das Gleichungssystem eindeutig lösbar ist.*

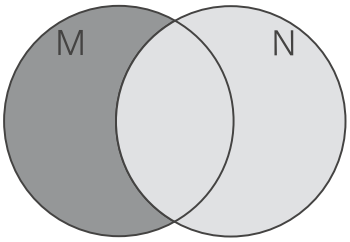


Abbildung 2.1: Komplement von zwei Mengen

Bemerkung 2.3.9 *Entsprechend den Standard-Mengenkonstruktionen lassen sich aus Ereignissen neue Ereignisse konstruieren. Sind* $M, N \subset \Omega$ *Ereignisse dann ist*

$$M \backslash N = \{m \in M \mid m \notin N\}$$

das ***Komplement*** *(oder die* ***Differenz****) von* N *in* M*, als Venn-Diagramm siehe Abbildung 2.1. entsprechen der Bedingung, dass Ereignis* M *eintritt, aber* N *nicht.*

Da wir in Zusammenhang mit Ereignissen nur Mengen $N \subset \Omega$ *betrachten, können wir auch von dem* ***Komplement der Teilmenge*** N *von* Ω *sprechen und meinen hier*

$$\overline{N} = \Omega \backslash N,$$

siehe Abbildung 2.2. Dies entspricht der Bedingung, dass das Er-

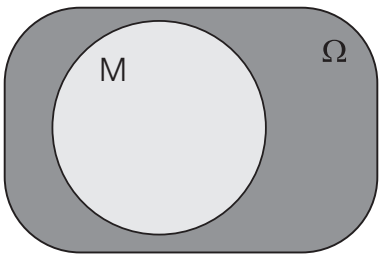


Abbildung 2.2: Komplement

eignis M *nicht eintritt.*

Weiter ist

$$M \cup N = \{m \mid m \in M \text{ oder } m \in N\}$$

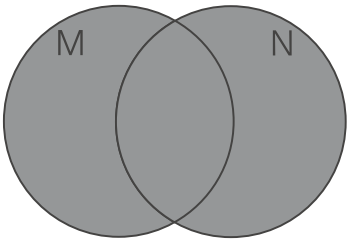


Abbildung 2.3: Vereinigung

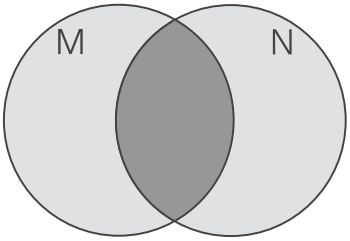


Abbildung 2.4: Durchschnitt

die ***Vereinigung*** *von* M *und* N*, siehe Abbildung 2.3, ensprechend der Bedingung, dass Ereignis* M *oder* N *eintritt.*

Ebenso entspricht der ***Durchschnitt*** *von* M *und* N

$$M \cap N = \{m \mid m \in M \text{ und } m \in N\}$$

dem Ereignis, dass sowohl M *als auch* N *eintreten, siehe Abbildung 2.4.*

Bemerkung 2.3.10 *Ist* M *ein Ereignis, dann gilt*

$$P(\overline{M}) = 1 - P(M).$$

Beispiel 2.3.11 *Wir beschreiben mit*

$$\Omega = \{B, R, S\}^2$$

das Wetter der kommenden zwei Tage (wobei B, R, S *wieder für bedeckt, regen und sonnig stehen) und nehmen der Einfachheit halber an, dass*

$$m(\omega) = \frac{1}{9}$$

für alle ω. Ist M das Ereignis, dass es morgen regnet, also

$$\begin{aligned} M &= \{(R, w_2) \mid w_2 \in \Omega\} \\ &= \{(R,R),(R,S),(R,B)\} \end{aligned}$$

dann ist

$$\overline{M} = \Omega \backslash M = \{(S,R),(S,S),(S,B),(B,R),(B,S),(B,B)\}$$

das komplementäre Ereignis, dass es morgen nicht regnet. Es gilt somit

$$P(M) = \frac{3}{9}$$

also

$$P(\overline{M}) = 1 - P(M) = \frac{6}{9}.$$

Für zwei Mengen $M, N \subset \Omega$ gilt die Siebformel

$$P(M \cup N) = P(M) + P(N) - P(M \cap N).$$

Einerseits tritt in

$$P(M \cup N) = \sum_{\omega \in M \cup N} m(\omega)$$

der Summand $m(\omega)$ für $\omega \in M \cup N$ genau einmal auf, und andererseits tritt der Summand $m(\omega)$ in

$$P(M) + P(N) = \sum_{\omega \in M} m(\omega) + \sum_{\omega \in N} m(\omega)$$

ebenso genau einmal auf, ausser im Fall $\omega \in M \cap N$, in dem er zweimal summiert wird. Allgemeiner haben wir:

Satz 2.3.12 (Siebformel) *Ist $m : \Omega \to \mathbb{R}_{\geq 0}$ eine Wahrscheinlichkeitsfunktion und sind $M_1, \ldots, M_n \subset \Omega$ Ereignisse, dann gilt*

$$P(M_1 \cup \ldots \cup M_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{|T|=k} P(M_T)$$

mit

$$M_T = \bigcap_{i \in T} M_i$$

für $T \subset \{1, ..., n\}$.

Der Beweis folgt dem Beweis der Siebformel für die Mächtigkeit von Mengen und ist Übungsaufgabe 2.10.

Beispiel 2.3.13 *Wir beschreiben wieder mit*

$$\Omega = \{B, R, S\}^2$$

das Wetter der kommenden zwei Tage und nehmen an, dass

$$m(\omega) = \frac{1}{9}$$

für alle ω. Ist M das Ereignis, dass es morgen regnet, also

$$\begin{aligned} M &= \{(R, \omega_2) \mid \omega_2 \in \Omega\} \\ &= \{(R,R), (R,S), (R,B)\} \end{aligned}$$

und N das Ereignis, dass es übermorgen sonnig ist

$$N = \{(B,S), (R,S), (S,S)\}$$

dann können wir die Wahrscheinlichkeit, dass es morgen regnet oder übermorgen sonnig ist berechnen als

$$\begin{aligned} P(M \cup N) &= P(M) + P(N) - P(M \cap N) \\ &= \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9} \end{aligned}$$

mit

$$M \cap N = \{(R,S)\}.$$

Explizit ist $M \cup N$ die rot markierte Teilmenge von

$$\begin{aligned} \Omega = \{ \quad & (R,R), \quad (R,S), \quad (R,B), \\ & (S,R), \quad (S,S), \quad (S,B), \\ & (B,R), \quad (B,S), \quad (B,B) \quad \}. \end{aligned}$$

2.3.3 Hintereinanderausführen von Experimenten

Eine Möglichkeit aus Wahrscheinlichkeitsexperimenten neue Experimente zu konstruieren ist das Hintereinanderausführen von

Zufallsexperimenten. Ein Beispiel könnte etwa die Hintereinanderausführung von randomisierten Algorithmen sein, etwa von mehreren Primzahltests. Dabei kann das Ergebnis des vorausgegangen Algorithmus entscheiden, welcher Algorithmus als nächstes ausgeführt wird. Wir haben also ein Programm, das iterativ in Unterprogramme verzweigt. Die logische oder zeitliche Abfolge beschreibt man am leichtesten mit Hilfe eines Baumdiagramms.

Definition 2.3.14 *Ein **Wahrscheinlichkeitsbaum** ist ein endlicher gerichteter Graph, der die Form eines Baumes hat und in dem alle Kanten, die aus einem Vertex herausgehen, zu den Ergebnissen desselben Zufallsexperiments korrepondieren.*

Der Wurzel des Baum geben wir die Wahrscheinlichkeit 1*. Die Wahrscheinlichkeiten der weiteren Vertices berechnen sich induktiv: Korrepondiert eine Kante im Graphen zum Ergebnis ω mit Wahrscheinlichkeit $m(\omega)$ und hat der Ausgangsknoten die Wahrscheinlichkeit p, dann hat der Endknoten der Kante die Wahrscheinlichkeit $p \cdot m(\omega)$.*

Bemerkung 2.3.15 *Offenbar gilt dann: Die Wahrscheinlichkeit der Blätter des Baums (d.h. der Knoten, aus denen keine Kanten herausgehen) addieren sich zu* 1 *(zeigen Sie dies als Übung). Insbesondere beschreibt ein Wahrscheinlichkeitsbaum wieder ein Zufallsexperiment.*

Beispiel 2.3.16 *In Abbildung 2.5 wird abhängig von dem Ergebnis des Experiments mit Ergebnisraum $\{\omega_1, \omega_2\}$ das Experiment mit Ergebnisraum $\{\omega_1', \omega_2'\}$ bzw. das mit Ergebnisraum $\{\omega_1'', \omega_2'', \omega_3''\}$ ausgeführt. Die zugehörigen Wahrscheinlichkeitsfunktionen bezeichnen wir mit m, m' und m''. Die Wahrscheinlichkeit für das Ergebnis (ω_2, ω_3'') ist dann z.B.*

$$m(\omega_2) \cdot m''(\omega_3'').$$

Beispiel 2.3.17 *Mit Wahrscheinlichkeit* 60% *ist es heute sonnig, anderenfalls regnet es. Falls es sonnig ist, dann wird es mit* 50% *Wahrscheinlichkeit windig, falls es regnet, ist es mit* 80% *windig. Das Baumdiagramm in Abbildung 2.6 beschreibt dieses Experiment, wobei die Ergebnisse in schwarz, deren Wahrscheinlichkeiten in rot und die Wahrscheinlichkeiten der Vertices (ensprechend der Ergebnisfolgen von der Wurzel bis zu dem Vertex)*

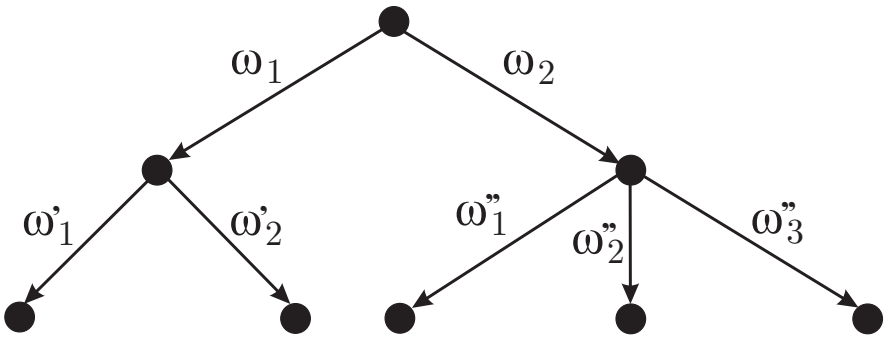


Abbildung 2.5: Wahrscheinlichkeitsbaum

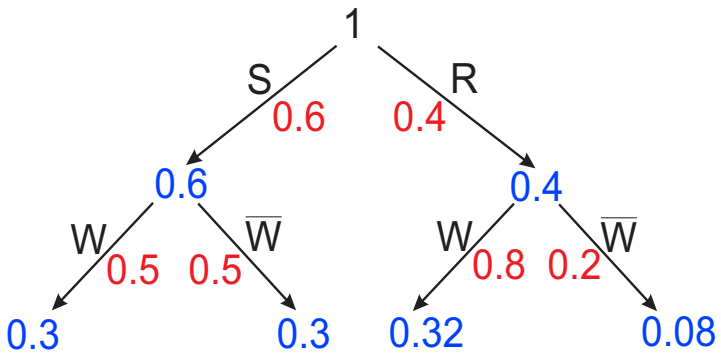


Abbildung 2.6: Wahrscheinlichkeitsbaum

in blau notiert sind. Die Wahrscheinlichkeit für die möglichen Ergebnisfolgen sind dann

ω	(S,W)	$(S,\overline{W})$	(R,W)	$(R,\overline{W})$
	$\frac{6}{10}\cdot\frac{5}{10}$	$\frac{6}{10}\cdot\frac{5}{10}$	$\frac{4}{10}\cdot\frac{8}{10}$	$\frac{4}{10}\cdot\frac{2}{10}$

also

ω	(S,W)	$(S,\overline{W})$	(R,W)	$(R,\overline{W})$
	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{32}{100}$	$\frac{8}{100}$

Hier können wir dann z.B. die Wahrscheinlichkeit für Wind ablesen als

$$\frac{3}{10}+\frac{32}{100}=\frac{62}{100}$$

entsprechend den Ergebnissequenzen (S,W) und (R,W). Die

Wahrscheinlichkeit, dass es windstill ist, ist komplentär dazu

$$\frac{3}{10} + \frac{8}{100} = \frac{38}{100}$$

entsprechend den Ergebnissequenzen $(S, \overline{W})$ *und* $(R, \overline{W})$.

Als Spezialfall eines Wahrscheinlichkeitsbaums haben wir:

Bemerkung 2.3.18 *Die unabhängige Ausführung von Zufallsexperimenten mit Ergebnisräumen* $\Omega_1, \ldots, \Omega_r$ *mit Wahrscheinlichkeitsfunktionen* $m_1, \ldots, m_r$ *wird beschrieben durch das kartesische Produkt*

$$\Omega = \Omega_1 \times \ldots \times \Omega_r$$

mit der Wahrscheinlichkeitsfunktion

$$m(\omega_1, \ldots, \omega_r) = m(\omega_1) \cdot \ldots \cdot m(\omega_r).$$

Beispiel 2.3.19 *Einmal Würfeln und einmal Münzwurf hat den Ergebnisraum*

$$\{1, 2, 3, 4, 5, 6\} \times \{0, 1\}$$

(wobei 0 *für Kopf und* 1 *für Zahl steht) und*

$$m(\omega_1, \omega_2) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Dieses kartesische Produkt können wir wie in Abbildung 2.7 gezeigt auf zwei Weisen mit Hilfe eines Baums aufzählen (indem wir erst die Würfel- oder erst die Münzwurf-Komponente wählen). Als Programm können wir uns das also als zwei verschachtelte For-Schleifen vorstellen, wobei die äußere Schleife entweder durch die Ergebnisse des Würfeln oder des Münzwurfs iteriert.

Weitere Anwendungen von Wahrscheinlichkeitsbäumen werden wir in den Übungsaufgaben 2.12 und 2.13 sehen.

2.4 Wahrscheinlichkeiten und Chancen

Häufig wird in der Praxis, z.B. bei Glücksspielen, eine Wahrscheinlichkeit durch eine Gewinnchance angegeben.

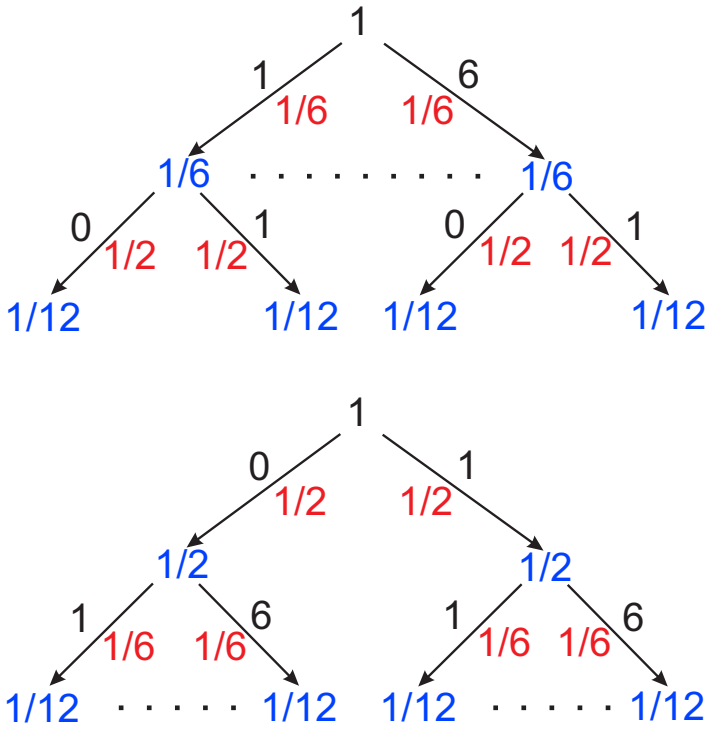


Abbildung 2.7: Kartesisches Produkt als Wahrscheinlichkeitsbaum

Definition 2.4.1 *Wir sagen, die* ***Chance*** *für das Ereignis M ist $r : s$, falls für die Wahrscheinlichkeit $P(M)$ für das Ereignis M gilt*

$$\frac{P(M)}{P(\overline{M})} = \frac{r}{s}.$$

Gegeben $P(M)$ ist also

$$\frac{r}{s} = \frac{P(M)}{1 - P(M)}.$$

Umgekehrt, gegeben die Chance $r : s$ für M, erhalten wir mit

$$P(M) = \frac{r}{s}(1 - P(M)) = \frac{r}{s} - \frac{r}{s}P(M)$$

dass

$$P(M) = \frac{\frac{r}{s}}{\frac{r}{s} + 1} = \frac{r}{r + s}.$$

Beispiel 2.4.2 *Bei einem Glücksspiel ist die Gewinnchance*

$$1:5$$

es gilt damit für die Gewinnwahrscheinlichkeit p, dass

$$p = \frac{1}{5}(1-p)$$

also ist

$$p = \frac{1}{6}.$$

Wir könnten das Spiel also durch Würfeln realisieren, wobei wir gewinnen falls eine 6 gewürfelt wird.

Siehe auch Aufgabe 2.8.

2.5 Zufallsvariablen

Oft will man bei einem Zufallsexperiment nur eine aus den Ergebnissen abgeleitete Größe untersuchen. Dazu betrachtet man eine Abbildung, die einem Ergebnis die abgeleitete Größe zuordnet, und hier die Wahrscheinlichkeit mit der diese Größe einen bestimmten Wert annimmt.

Definition 2.5.1 *Auf einem diskreten Wahrscheinlichkeitsraum mit Ergebnisraum Ω und Wahrscheinlichkeitsfunktion*

$$m : \Omega \to \mathbb{R}_{\geq 0}$$

*ist eine **Zufallsvariable** X eine Abbildung*

$$X : \Omega \to N$$

in eine Menge N, die als Bildraum von X bezeichnet wird.

*Die **Verteilung** der Zufallsvariable X im Bildraum N ist gegeben durch die Funktion*

$$\begin{array}{rccc} m_X: & N & \to & \mathbb{R}_{\geq 0} \\ & n & \mapsto & P(X^{-1}(\{n\})) \end{array}$$

wir bilden also das Urbild von n unter der Abbildung X und von diesem Ereignis die Wahrscheinlichkeit.

Notation 2.5.2 *Man verwendet auch die Notation*

$$P(X = n) := m_X(n) = P(X^{-1}(\{n\})).$$

Bemerkung 2.5.3 *1) Nach der gilt*

$$P(X = n) = \sum_{\substack{\omega \in \Omega \\ X(\omega) = n}} m(\omega).$$

2) Es gilt

$$\sum_{n \in N} P(X = n) = 1.$$

Beweis.

1) Ist klar nach der Definition des Urbilds

$$X^{-1}(\{n\}) = \{\omega \in \Omega \mid X(\omega) = n\} \subset \Omega$$

und der Wahrscheinlichkeit von Ereignissen.

2) Gilt, da

$$\sum_{n \in N} P(X = n) = \sum_{n \in N} \sum_{\substack{\omega \in \Omega \\ X(\omega) = n}} m(\omega) = \sum_{\omega \in \Omega} m(\omega) = 1.$$

wobei wir verwenden, dass mit der Abbildungseigenschaft von X die Urbilder verschiedener n disjunkt sind. Man beachte, dass $P(X = n)$ nur für abzählbar viele n ungleich 0 ist und die Summe nicht von der Summationsreihenfolge abhängt, da die Reihe $\sum_{\omega \in \Omega} m(\omega)$ nach Voraussetzung absolut konvergiert.

■

Beispiel 2.5.4 *Durch die identische Abbildung erhalten wir stets eine Zufallsvariable.*

Beispiel 2.5.5 *Für den 2-maligen Wurf mit einem fairen Würfel ist*

$$\Omega = \{1, \ldots, 6\}^2$$

und

$$m(\omega) = \frac{1}{6^2} = \frac{1}{36}$$

für alle $\omega = (a,b) \in \Omega$. *Die Zufallsvariable*

$$\begin{array}{rccc} X: & \Omega & \to & N = \{2,\dots,12\} \\ & (a,b) & \mapsto & a+b \end{array}$$

bildet auf die Augensumme ab. Um die Verteilung auf N *zu berechnen, bestimmen wir die Anzahl der geordneten Partitionen von* $n \in N$ *in zwei Summanden* ≥ 0. *Beispielsweise ist*

$$X^{-1}(\{4\}) = \{(1,3),(2,2),(3,1)\}$$

also die Wahrscheinlichkeit Augensumme 4 *zu würfeln gleich*

$$P(X^{-1}(\{4\})) = \frac{3}{36} = \frac{1}{12}.$$

Nach Satz 1.9.21 gibt es

$$\binom{n-1}{2-1} = n-1$$

solche Partitionen. Für $n > 7$ *ist der Satz aber nicht auf unser Problem anwendbar, da wir auch Partitionen wie* $8 = 1+7$ *zählen würden, die beim Würfeln nicht auftreten können. Allerdings haben wir eine Bijektion*

$$\begin{array}{c} \Omega \to \Omega \\ (a,b) \mapsto (7-a, 7-b) \end{array}$$

Diese bildet Ergebnisse mit Augensumme n *auf Ergebnisse mit Augensumme* $14-n$ *ab, etwa*

$$(6,5) \mapsto (1,2).$$

Statt Partitionen von $n > 7$ *können wir also auch Partitionen von* $14-n < 7$ *zählen. Das folgende Diagramm listet alle Elemente*

von Ω auf

	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
2	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
3	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
4	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
6	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$
7	8	9	10	11	12	

wobei die Augensumme der jeweiligen Diagonalen in rot notiert ist. Wir erhalten also auf N die Verteilung

n	2	3	4	5	6	7	8	9	10	11	12
$\lvert X^{-1}(\{n\})\rvert$	1	2	3	4	5	6	5	4	3	2	1
$P(X=n)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.6 Erwartungswerte

Wenn wir einen Las-Vegas-Algorithmus wie den randomisierten Quicksort-Algorithmus betrachten, dann stellen sich zwei Fragen: Was ist die worst-case Laufzeit, und was ist die mittlere Laufzeit? Die erste der beiden Fragen haben wir schon in Proposition 2.2.2 beantwortet. Die zweite Frage ist die in der Praxis wichtigere: Wir werden eine randomisierte Version von Quicksort typischerweise dann einsetzen, wenn viele Vergleiche notwendig sind, also für große Mengen. In diesem Fall können wir erwarten, dass sich unglückliche und glückliche Wahlen der Pivotelemente über die Gesamtlaufzeit herausmitteln. Wir werden die Frage nach der mittleren Laufzeit beantworten, indem wir den Begriff des Erwartungswerts einführen.

2.6.1 Mittelwert und Erwartungswert

Beispiel 2.6.1 *In Beispiel 2.1.2 hatten wir 10-mal gewürfelt und die Sequenz von Ergebnissen*

$$6,3,5,3,1,1,3,4,2,1$$

erhalten. Der Mittelwert der Ergebnisse ist

$$\begin{aligned}\frac{29}{10} &= \frac{1}{10}(6+3+5+3+1+1+3+4+2+1)\\ &= 1\cdot\frac{3}{10}+2\cdot\frac{1}{10}+3\cdot\frac{3}{10}+4\cdot\frac{1}{10}+5\cdot\frac{1}{10}+6\cdot\frac{1}{10}.\end{aligned}$$

wobei wir in der zweiten Gleichung nach den Würfelergebnissen sortiert haben. Die ***relative Häufigkeit*** *eines Ergebnisses gibt an, welcher Anteil der Experimente zu diesem Ergebnis geführt hat. In unserem Experiment haben wir die Zahlen* $1,\dots,6$ *mit den folgenden relativen Häufigkeiten erhalten*

	1	2	3	4	5	6
relative Häufigkeit	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

die in der obígen Summe als Faktoren vor den Ergebnissen $1,\dots,6$ *stehen. Da jedes Ergebnis* $1,\dots,6$ *beim Würfeln gleich wahrscheinlich ist, würden wir in einem idealen Experiment erwarten, dass bei* 10 *Würfen jede Zahl mit relativer Häufigkeit*

$$\frac{10\cdot\frac{1}{6}}{10}=\frac{1}{6}$$

auftritt. Ersetzen wir im Mittelwert die gemessenen relativen Häufigkeiten mit den theoretischen relativen Häufigkeiten, also mit den Wahrscheinlichkeiten der jeweiligen Ergebnisse, dann erhalten wir den Erwartungswert

$$1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{7}{2}.$$

Bei einer tatsächlichen Durchführung des Experiments kann natürlich alles passieren, wir könnten etwa immer eine 6 *würfeln, was zu relativen Häufigkeiten*

	1	2	3	4	5	6
relative Häufigkeit	0	0	0	0	0	1

führen würde. Im Gesetz der großen Zahlen werden wir zeigen, dass die Wahrscheinlichkeit für ein solches seltsames Ergebnis mit der Anzahl der Würfe tatsächlich gegen 0.

Zunächst müssen wir aber erst einmal formal den Begriff des Erwartungswerts einführen. Eine wesentlich Beobachtung ist, dass wir bei der Bildung des Mittelwerts oder Erwartungswerts Ergebnisse mit ihren relativen Häufigkeiten bzw. Wahrscheinlichkeiten multiplizieren und die Resultate dann addieren müssen. Dies ist gewährleistet, wenn die Werte der betrachteten Zufallsvariable in einem $\mathbb{R}$-Vektorraum liegen.[2]

Definition 2.6.2 *Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$,*

$$X : \Omega \to N$$

*eine Zufallsvariable mit N ein $\mathbb{R}$-Vektorraum. Der **Erwartungswert** von X ist dann*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot m(\omega),$$

falls diese Summe existiert und eindeutig ist. Anderenfalls sagen wir, dass X keinen Erwartungswert hat.

Bemerkung 2.6.3 *Wir verlangen hier, dass für jede Abzählung von Ω die entsprechende Reihe (bezüglich einem sinnvollen Konvergenzbegriff in N) konvergiert und denselben Grenzwert liefert.*

Wir werden uns nur für die folgenden unproblematischen Fälle interessieren:

- *Ist Ω endlich, dann ist auch die Summe endlich.*
- *Ist $N \subset \mathbb{R}$ eine Teilmenge der reellen Zahlen und*

$$\sum_{\omega \in \Omega} X(\omega) \cdot m(\omega)$$

[2] Man beachte: Gemessene relative Häufigkeiten sind stets in $\mathbb{Q}$, zur Bildung eines Mittelwerts reicht also eine $\mathbb{Q}$-Vektorraumstruktur. Für Wahrscheinlichkeiten lassen wir Werte in $\mathbb{R}$ zu. Eine irrationale Wahrscheinlichkeit können wir also in jedem Fall nur näherungsweise als relative Häufigkeit erhalten.

Man beachte auch: Tatsächlich bilden wir bei Mittelwert und Erwartungswert nur sogenannte Konvexkombinationen von Elementen $n_i \in N$ also Summen $\sum_i n_i \cdot m_i$ wobei $\sum_i m_i = 1$ und alle $m_i \geq 0$. Es reicht also zu fordern, dass N konvex ist, d.h. alle Konvexkombinationen von Elementen von N wieder in N liegen.

absolut konvergent, dann ist der Grenzwert unabhängig von der Summationsreihenfolge.[3]

In anderen Situationen als diesen ist es sinnvoll, ein axiomatisches Konzept von Wahrscheinlichkeitsräumen zu verwenden (siehe Abschnitt 5.1 im Anhang) und dann in der jeweiligen Situation zu zeigen, wie dieses Konzept realisiert werden kann.

Bemerkung 2.6.4 *Ist Ω selbst ein $\mathbb{R}$-Vektorraum, dann können wir insbesondere für die Zufallsvariabele die identische Abbildung $X = \mathrm{id}$ nehmen und sprechen dann von dem Erwartungswert*

$$E(m) := E(\mathrm{id}) = \sum_{\omega \in \Omega} \omega \cdot m(\omega)$$

des durch die Wahrscheinlichkeitsfunktion m beschriebenen Zufallsexperiments.

Bemerkung 2.6.5 *Den Erwartungswert in Definition 2.6.2 können wir auch über eine Summe im Bildraum von X bestimmen, denn nach Bemerkung 2.5.3 gilt*

$$P(X = n) = \sum_{\substack{\omega \in \Omega \\ X(\omega) = n}} m(\omega)$$

und nach Voraussetzung dürfen wir bei Existenz des Erwartungswerts von X umsortieren, also ist

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega) \cdot m(\omega) \\ &= \sum_{n \in N} n \cdot P(X = n) \end{aligned}$$

indem wir alle Summanden mit $X(\omega) = n$ für festes n zusammenfassen und n ausklammern. Man beachte, dass $P(X = n)$ nur für abzählbar viele n ungleich 0 ist.

[3] Wir erinnern uns: Falls eine Reihe konvergiert, aber nicht absolut konvergiert, dann kann man durch Umordnen jeden Grenzwert erreichen kann. Ein schwächerer Konvergenzbegriff als absolute Konvergenz macht im Fall $N = \mathbb{R}$ also keinen Sinn.

Beispiel 2.6.6 *Sei X die Anzahl von Kopf bei einem 4-maligen Münzwurf (wobei wir 0 für Kopf und 1 für Zahl schreiben). Dann haben wir die ensprechenden Ergebnisse und Wahrscheinlichkeiten*

n	0	1	2	3	4
ω *mit* $X(\omega) = n$	1111	1110	1100	0001	0000
		1101	1001	0010	
		1011	0011	0100	
		0111	1010	1000	
			0101		
			0110		
$P(X = n)$	$\frac{1}{2^4}$	$\frac{4}{2^4}$	$\frac{6}{2^4}$	$\frac{4}{2^4}$	$\frac{1}{2^4}$

Der Erwartungswert ist also

$$\begin{aligned}
E(X) &= \sum_{\omega\in\Omega} X(\omega)\cdot m(\omega)\\
&= 0\cdot\frac{1}{2^4}\\
&+ 1\cdot\frac{1}{2^4} + 1\cdot\frac{1}{2^4} + 1\cdot\frac{1}{2^4} + 1\cdot\frac{1}{2^4}\\
&+ 2\cdot\frac{1}{2^4} + 2\cdot\frac{1}{2^4} + 2\cdot\frac{1}{2^4} + 2\cdot\frac{1}{2^4} + 2\cdot\frac{1}{2^4} + 2\cdot\frac{1}{2^4}\\
&+ 3\cdot\frac{1}{2^4} + 3\cdot\frac{1}{2^4} + 3\cdot\frac{1}{2^4} + 3\cdot\frac{1}{2^4}\\
&+ 4\cdot\frac{1}{2^4}\\
&= 2
\end{aligned}$$

oder als Summe im Bildraum nach Bemerkung 2.6.5

$$\begin{aligned}
E(X) &= \sum_{n\in N} n\cdot P(X = n)\\
&= 0\cdot\frac{1}{2^4} + 1\cdot\frac{4}{2^4} + 2\cdot\frac{6}{2^4} + 3\cdot\frac{4}{2^4} + 4\cdot\frac{1}{2^4} = 2.
\end{aligned}$$

Wenn man im Ergebnisraum summiert, muss man also oft mehr Summanden betrachten, da mehrere Ergebnisse ω zum selben Wert $X(\omega)$ führen können.

Beispiel 2.6.7 *Wir werfen eine Münze bis zum ersten mal Kopf kommt. Nach Beispiel 2.3.3 ist der Ergebnisraum* $\Omega = \mathbb{N} \cup \{\infty\}$ *mit Wahrscheinlichkeitsfunktion*

$$m(n) = \left(\frac{1}{2}\right)^n$$

und $m(\infty) = 0$. *Der Erwartungswert des Experiments ist also*

$$E(m) = \sum_{n=1}^{\infty} n\frac{1}{2^n}.$$

Ableiten der geometrischen Reihe gibt

$$\sum_{n=1}^{\infty} nx^{n-1} = \left(\sum_{n=0}^{\infty} x^n\right)' = \frac{1}{(1-x)^2}$$

also

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}.$$

Damit erhalten wir

$$E(m) = \sum_{n=1}^{\infty} n\frac{1}{2^n} = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} = 2.$$

Beispiel 2.6.8 *Erwartungswerte müssen nicht existieren: Falls bei dem Münzwurfexperiment aus Beispiel 2.6.7 beim n-ten Wurf zum ersten Mal Kopf kommt, dann gewinnen wir* 2^n €. *Den erwarteten Gewinn können wir also mit der Zufallsvariable* $X(n) = 2^n$ *ausdrücken als den Erwartunswert*

$$E(X) = \sum_{n=1}^{\infty} 2^n\frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty.$$

Die Reihe konvergiert nicht, der Erwartungswert existiert also nicht. Tatsächlich divergiert die Reihe bestimmt gegen unendlich, falls man dieses Spiel lange genug spielt, wird man also beliebig reich. Siehe dazu auch Übungsaufgabe 2.11, wo wir berücksichtigen, dass es auf der Welt nur maximal etwa 2^{47} € *Geldmittel als Gewinn existieren.*

Beispiel 2.6.9 *Können wir den Mittelwert der Zahlensequenz*

$$6, 3, 5, 3, 1, 1, 3, 4, 2, 1$$

als Erwartungswert eines Zufallsexperiments erhalten? Dazu betrachten wir das Zufallsexperiment, das aus den obigen Zahlen zufällig eine auswählt und die Zufallsvariable, die diesen Wert ausgibt, d.h. wir setzen

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

und

$$m(i) = \frac{1}{10}$$

für alle $i \in \Omega$ und definieren die Zufallsvariable $X : \Omega \to \{1, \ldots, 6\}$ durch

i	1	2	3	4	5	6	7	8	9	10
$X(i)$	6	3	5	3	1	1	3	4	2	1

Als Programm in MAPLE *können wir das Experiment z.B. wie folgt durchführen:*

```
L:=[6,3,5,3,1,1,3,4,2,1];
i:=rand(1..10)();
```

9

```
L[i];
```

2

Der Erwartungswert von X ist dann genau der Mittelwert:

$$E(X) = \frac{1}{10}(6 + 3 + 5 + 3 + 1 + 1 + 3 + 4 + 2 + 1) = \frac{29}{10}.$$

Beispiel 2.6.10 *Wir bestimmen den Erwartungswert für die Augensumme bei zweimaligem Würfeln. Wie in Beispiel 2.5.5 ist also*

$$\Omega = \{1, \ldots, 6\}^2$$

mit

$$m(\omega) = \frac{1}{6^2} = \frac{1}{36}$$

und wir müssen den Erwartungswert der Zufallsvariable

$$\begin{array}{rccc} X: & \Omega & \to & N = \{2, \ldots, 12\} \\ & (a, b) & \mapsto & a + b \end{array}$$

bestimmen. Entsprechend der Tabelle in Beispiel 2.5.5 ist

$$\begin{aligned} E(X) &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} \\ &\quad + 7 \cdot \frac{6}{36} \\ &\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\ &= 7 \end{aligned}$$

Für weitere Beispiele siehe die Übungsaufgaben 2.12, 2.13 und 2.14.

2.6.2 Linearität von Erwartungswerten

Beispiel 2.6.11 *Für zweimaliges Würfeln betrachten sei*

$$\begin{array}{rccc} X: & \Omega & \to & N = \{2, \ldots, 12\} \\ & (a,b) & \mapsto & a+b \end{array}$$

wie in Beispiel 2.6.10 die Zufallsvariable, die die Augensumme berechnet. In Beispiel hatten wir den Erwartungswert $E(X)$ aus der Verteilung von X berechnet (diese Verteilung hatten wir in Beispiel 2.5.5 bestimmt). Wie wir im folgenden sehen werden, lässt sich diese Rechnung signifikant vereinfachen. Dazu zerlegt man X in eine Summe von Zufallsvariablen

$$\begin{array}{rccc} X_1: & \Omega & \to & N = \{2, \ldots, 12\} \\ & (a,b) & \mapsto & a \end{array}$$

und

$$\begin{array}{rccc} X_2: & \Omega & \to & N = \{2, \ldots, 12\} \\ & (a,b) & \mapsto & b \end{array}$$

die die Augenzahl beim ersten und zweiten Würfeln repräsentieren. Es gilt dann

$$X = X_1 + X_2$$

in dem Sinne, dass

$$X(\omega) = X_1(\omega) + X_2(\omega).$$

Die Erwartungswerte von X_1 und X_2, also den Erwartungswert für die Augenzahl bei einmaligem Würfeln hatten wir schon in Beispiel 2.6.1 als

$$E(X_1) = E(X_2) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

bestimmt. Nach Beispiel 2.6.10 ist

$$E(X) = 7,$$

wir könnten uns also die Frage stellen, ob vielleicht

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Im folgenden Satz zeigen wir, dass diese tatsächlich immer gilt. Sei im folgenden Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$.

Satz 2.6.12 *Seien $X_1 : \Omega \to N$ und $X_2 : \Omega \to N$ Zufallsvariablen, für die jeweils der Erwartungswert existiert. Die Summe $X_1 + X_2$ der Zufallsvariablen ist definiert als die Zufallsvariable*

$$\begin{array}{rccc} X_1 + X_2 : & \Omega & \to & N \\ & \omega & \mapsto & X_1(\omega) + X_2(\omega) \end{array}$$

Dann existiert auch der Erwartungswert von $X_1 + X_2$ und es gilt

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Beweis. Es ist mit Bemerkung 2.6.5

$$\begin{aligned} E(X_1 + X_2) &= \sum_{\omega \in \Omega} (X_1 + X_2)(\omega) \cdot m(\omega) \\ &= \sum_{\omega \in \Omega} (X_1(\omega) + X_2(\omega)) \cdot m(\omega) \\ &= \sum_{\omega \in \Omega} (X_1(\omega) \cdot m(\omega) + X_2(\omega) \cdot m(\omega)) \\ &\overset{(*)}{=} \sum_{\omega \in \Omega} X_1(\omega) \cdot m(\omega) + \sum_{\omega \in \Omega} X_2(\omega) \cdot m(\omega) \\ &= E(X_1) + E(X_2). \end{aligned}$$

Dabei folgt $(*)$ mit den Rechenregeln für konvergente Reihen (wie wir sie zumindest in dem für uns interessanten Fall von Reihen mit reellen Summanden in der Analysis kennengelernt haben). ■

Bemerkung 2.6.13 *Mit Induktion gilt für Zufallsvariablen* $X_i : \Omega \to N$*, dass*

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n).$$

Bemerkung 2.6.14 *Ist* $c \in \mathbb{R}$ *und* $X : \Omega \to N$ *eine Zufallsvariable, dann erhalten wir eine Zufallsvariable*

$$\begin{array}{rccc} c \cdot X : & \Omega & \to & N \\ & \omega & \mapsto & c \cdot X(\omega) \end{array}$$

Falls $E(X)$ *existiert, dann auch* $E(c \cdot X)$*, und es gilt*

$$E(c \cdot X) = c \cdot E(X).$$

Beweis. Es gilt

$$\begin{aligned} E(c \cdot X) &= \sum_{\omega \in \Omega} c \cdot X(\omega) \cdot m(\omega) \\ &\overset{(*)}{=} c \cdot \sum_{\omega \in \Omega} X(\omega) \cdot m(\omega) \\ &= c \cdot E(X). \end{aligned}$$

wobei $(*)$ mit den Rechenregeln für konvergente Reihen folgt. ■

Beispiel 2.6.15 *Wir berechnen den Erwartungswert der Anzahl von Zahl bei einem* 4*-maligen Münzwurf auf einfachere Weise als in Beispiel 2.6.6 (wobei wir* 0 *für Kopf und* 1 *für Zahl schreiben): Dazu betrachten wir* 4 *Zufallsvariablen* $X_1, \ldots, X_4 : \Omega = \{0,1\}^4 \to N = \{0,1\}$ *mit*

$$X_i(\omega_1, \ldots, \omega_4) = \omega_i$$

und die Wahrscheinlichkeitsfunktion

$$m(0) = m(1) = \frac{1}{2}.$$

Die Variable X_i *gibt also die Anzahl von Zahl in einem einzelnen Münzwurf an und hat den Erwartungswert*

$$E(X_i) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

Die Variable $X_1 + \ldots + X_4$ beschreibt also die Anzahl von Zahl bei einem 4-maligen Münzwurf und für den Erwartungswert gilt nach Satz 2.6.12

$$\begin{aligned} E(X_1 + \ldots + X_4) &= E(X_1) + \ldots + E(X_4) \\ &= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 2. \end{aligned}$$

Beispiel 2.6.16 *Die Zufallsvariable X beschreibe die erwartete Anzahl an Rekordsommern innerhalb von n Jahren. Mit den Zufallsvariablen*

$$X_1, \ldots, X_n : \{0,1\}^n \to N = \{0,1\}$$

definiert durch

$$X_i(\omega) = \begin{cases} 1 & \text{falls das } i\text{-te Jahr ein Rekordjahr war} \\ 0 & \text{sonst} \end{cases}$$

können wir X schreiben als

$$X = \sum_{i=1}^{n} X_i.$$

Wenn wir davon ausgehen, dass der Rekordsommer über eine Zahl von i Jahren gleichwahrscheinlich in jedem der Jahre auftritt, d.h.

i	1	2	3	$\ldots$
$P(X_i = 1)$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\ldots$

also das erste Jahr ist trivialerweise das Rekordjahr, das zweite Jahr ist mit Wahrscheinlichkeit $\frac{1}{2}$ heißer als das vorhergehende und so weiter, dann haben wir

$$E(X_i) = 1 \cdot \frac{1}{i} + 0 \cdot (1 - \frac{1}{i}) = \frac{1}{i}.$$

Somit ist

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{i}$$

die ***n-te harmonische Zahl***

$$H_n := \sum_{i=1}^{n} \frac{1}{i}.$$

Die erwarteten Anzahlen von Rekordsommern innerhalb von 5, 10 *bzw.* 100 *Jahren sind also*

$$\begin{array}{lcll} H_5 & = & \frac{137}{60} & \approx 2.28 \\ H_{10} & = & \frac{7381}{2520} & \approx 2.93 \\ H_{100} & & & \approx 5.19. \end{array}$$

Siehe auch die Übungsaufgaben 2.14 und 2.15.

2.7 Erwartete Abweichung vom Erwartungswert: Varianz und Standardabweichung

Die Varianz ist ein Maß für die Streuung der Ergebnisse um den Erwartungswert. Sie misst, welche Abweichung der Ergebnisse einer Zufallsvariable $X : \Omega \to N$ vom Erwartungswert im Mittel zu erwarten ist. Die Varianz ist also wieder ein Erwartungswert, allerdings für eine Zufallsvariable, die die Abweichung von X vom Erwartungswert $E(X)$ liefert. Man könnte nun denken, dass $X - E(X)$ eine hierfür sinnvolle Zufallsvaraible sein könnte, jedoch ist nach Satz 2.6.12

$$E(X - E(X)) = E(X) - E(E(X)) = 0,$$

denn die konstante Zufallsvariable $Y = E(X)$ hat wieder den Erwartungswert

$$E(Y) = 1 \cdot E(X) = E(X)$$

Beschränken wir uns wieder auf den Fall $N = \mathbb{R}$, dann ist das Quadrat $(X - E(X))^2$ ein sinnvolles Maß für die Abweichung, denn diese Größe ist das Quadrat des Euklidischen Abstands.

Wir bemerken: Das Quadrat ist leichter zu handhaben als der Betrag $|X - E(X)|$ der Abweichung, denn die Betragsfunktion ist nicht differenzierbar, was z.B. zu Problemen führt, wenn man Extremwerte bestimmen möchte.

Definition 2.7.1 *Sei* Ω *ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion* $m : \Omega \to \mathbb{R}_{\geq 0}$ *und* $X : \Omega \to \mathbb{R}$

eine Zufallsvariable, für die der Erwartungswert existiert. Die ***Varianz*** *von X ist*

$$V(X) := E((X - E(X))^2)$$

falls dieser Erwartungswert existiert. Anderenfalls sagen wir, dass die Varianz nicht existiert. Die ***Standardabweichung*** *von X ist*

$$\sigma(X) = \sqrt{V(X)}.$$

Man beachte: Würden wir die Wurzel vor Berechnung des Erwartungswerts ziehen, dann erhielten wir den Betrag des Abstands, denn $\sqrt{x^2} = |x|$.

Bemerkung 2.7.2 *Die Standardabweichung führt man ein, da in der Praxis die Werte einer Zufallsvariable X oft mit einer Einheit kommen (etwa cm) und die Varianz als Quadrat eine davon abweichende Einheit hat (in unserem Beispiel cm^2). Die Standardabweichung hat wieder dieselbe Einheit wie die Werte von X.*

Bemerkung 2.7.3 *Es gilt*

$$\begin{aligned} V(X) &= E((X - E(X))^2) \\ &= \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 \cdot m(\omega) \\ &= \sum_{n \in N} (n - E(X))^2 \cdot P(X = n) \end{aligned}$$

wobei nach Voraussetzung die erste Summe absolut konvergent ist, ebenso die zweite Summe (man beachte: in dieser Summe sind nur abzählbar viele Summanden $\neq 0$).

Bemerkung 2.7.4 *Zur Konvergenz der Varianz: Konvergiert die Summe*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot m(\omega)$$

zur Berechnung des Erwartungswerts absolut, dann muss

$$\sum_{\omega \in \Omega} (X(\omega) - E(X))^2 \cdot m(\omega)$$

nicht notwendig konvergieren, siehe Aufgabe 2.16.

Beispiel 2.7.5 *Wir werfen einen Würfel und X beschreibe die Augenzahl, also $\Omega = \{1, \dots, 6\}$ und $X(\omega) = \omega$. Der Erwartungswert war*

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Zur Bestimmung der Varianz berechnen wir die Verteilung der Zufallsvariable $X - E(X)$ und daraus die von $(X - E(X))^2$:

ω	1	2	3	4	5	6
$\omega - E(X)$	$-\frac{5}{2}$	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{5}{2}$
$(\omega - E(X))^2$	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$

Jetzt lässt sich die Varianz leicht ablesen als

$$V(X) = \left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) \cdot \frac{1}{6} = \frac{35}{12} = 2.916666\dots$$

und die Standardabweichung als

$$\sigma(X) = 1.707825\dots$$

Satz 2.7.6 *Die Varianz von X lässt sich aus den Erwartungswerten von X^2 und X berechnen als*

$$V(X) = E(X^2) - E(X)^2.$$

Beweis. Mit Bemerkung 2.6.14 und Satz 2.6.12 ist

$$\begin{aligned} V(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2 \cdot X \cdot E(X) + E(X)^2) \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

■

Bemerkung 2.7.7 *Die Erwartungswerte $E(X^k)$ für $k \in \mathbb{N}$ bezeichnet man auch als die Momente von X. Die Existenz von $E(X^{k+1})$ impliziert die Existenz von $E(X^k)$. Zum Beweis siehe Aufgabe 2.19. Existiert $E(X^2)$ dann also auch $V(X)$ und $E(X)$.*

Beispiel 2.7.8 *Die Zufallsvariable X beschreibe wieder die Augenzahl beim Wurf eines Würfels. Dann hat X^2 die Verteilung*

ω	1	2	3	4	5	6
$X(\omega)^2$	1	4	9	16	25	36

also ist

$$E(X^2) = (1+4+9+16+25+36)\cdot\frac{1}{6} = \frac{91}{6}$$

und damit

$$V(X) = E(X^2) - E(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Satz 2.7.9 *Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die die Varianz existiert und $c \in \mathbb{R}$, dann gilt*

$$V(c\cdot X) = c^2 \cdot V(X)$$

und

$$V(X + c) = V(X).$$

Für den Beweis siehe Übungsaufgabe 2.18.

2.8 Anwendung: Erwartete Laufzeit des randomisierten Quicksort

Im Folgenden wollen wir beweisen, dass die erwartete Laufzeit des randomisierten Quicksort-Algorithmus auf einer n-elementigen Menge in $O(n\log(n))$ liegt. Wir codieren den Ablauf des Algorithmus in einen Baum von Pivotelementen. Sei Ω die Menge aller dieser Bäume. Die Zufallsvariable

$$X : \Omega \to \mathbb{N}_0$$

berechnet die Anzahl $X(\omega)$ der Vergleiche des Algorithmus für einen festgelegten Ablauf ω.

Lemma 2.8.1 *Der Erwartungswert für die Anzahl der Vergleiche im randomisierten Quicksort-Algorithmus auf einer n-elementigen Menge ist*

$$E(X) = \sum_{1\le i<j\le n} \frac{2}{j-i+1}$$

Beweis. In sortierter Form seien die Elemente der Menge gegeben als

$$y_1 < \ldots < y_n.$$

Wir definieren Zufallsvariablen

$$X_{ij} : \Omega \to \{0, 1\}$$

mit

$$X_{ij}(\omega) = \begin{cases} 1 & \text{falls im Lauf des Algorithmus } y_i \text{ mit } y_j \text{ verglichen wird} \\ 0 & \text{sonst} \end{cases}$$

Im Quicksortalgorithmus wird in jedem Schritt die zu sortierende Menge in zwei Teilmengen und das Pivotelement aufgeteilt. Zwei verschiedene Elemente $y_i \neq y_j$ werden also nie zweimal verglichen. Somit ist

$$X = \sum_{1 \leq i < j \leq n} X_{ij}$$

und damit nach Satz 2.6.12

$$E(X) = \sum_{1 \leq i < j \leq n} E(X_{ij}).$$

Wir bestimmen nun die Wahrscheinlichkeitsverteilung von X_{ij} für $i < j$:

- Liegt ein Pivotelement zwischen y_i und y_j, dann werden diese Elemente bei der Aufteilung nicht verglichen. Sie werden in zwei verschiedene Teilmengen sortiert und deshalb auch im Folgenden nicht mehr verglichen.
- Ist ein Pivotelement gleich y_i oder y_j, dann werden die beiden Elemente verglichen.
- Ist ein Pivotelement kleiner als y_i oder größer als y_j, dann werden die beiden Elemente bei der Aufteilung nicht verglichen. Sie werden unsortiert in dieselbe Teilmenge sortiert, müssen also durch Wahl eines weiteren Pivotelements verglichen werden.

Wir können uns dies als ein Spiel vorstellen, bei dem wir zufällig Elemente in

$$y_1, \ldots, y_n$$

auswählen. Gewinn bedeutet, dass ein Vergleich zwischen y_i und y_j stattfindet, Verlust, dass kein Vergleich stattfindet. Das Spiel endet (vom Standpunkt von y_i und y_j), sobald ein Element im Bereich $y_i, \ldots, y_j$ gewählt wird. Wir gewinnen, falls y_i oder y_j gewählt werden, und wir verlieren falls ein Element im Bereich $y_{i+1}, \ldots, y_{j-1}$ gewählt wird. Wird ein Element außerhalb des Bereichs $y_i, \ldots, y_j$ gewählt, fahren wir mit dem Spiel fort. Somit ist die Gewinnwahrscheinlichkeit gleich

$$\frac{2}{j-i+1}$$

oder anders ausgedrückt

$$P(X_{ij} = 1) = \frac{2}{j-i+1}.$$

Daraus folgt

$$\begin{aligned} E(X_{ij}) &= 1 \cdot \frac{2}{j-i+1} + 0 \cdot \left(1 - \frac{2}{j-i+1}\right) \\ &= \frac{2}{j-i+1}. \end{aligned}$$

■

Bemerkung 2.8.2 *Aus dem Beweis des Lemmas sehen wir: Die Wahrscheinlichkeit, dass y_i mit y_{i+a} verglichen wird ist*

$$\frac{2}{a+1}$$

Die folgende Tabelle gibt diese Wahrscheinlichkeit für verschiedene a an:

a	1	2	3	4	5	...
Wahrscheinlicheit Vergleich y_i mit y_{i+a}	1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{3}$	...

Bemerkung 2.8.3 *Wir können also die Laufzeit des randomisierten Quicksort-Algorithmus mit der i-ten harmonischen Zahl H_i auch schreiben als*

$$E(X) = \sum_{j=1}^{n} 2(H_j - 1).$$

Beweis. Klar nach Bemerkung 2.8.2. Alternativ können wir auch nach Lemma 2.8.1 schreiben

$$\begin{aligned} E(X) &= \sum_{j=1}^{n} \sum_{i=1}^{j-1} \frac{2}{j-i+1} \\ &= \sum_{j=1}^{n} 2(H_j - 1). \end{aligned}$$

■

Bemerkung 2.8.4 *Es ist keine geschlossene Formel für H_n bekannt.*

Beispiel 2.8.5 *Die erwartete Laufzeit des randomisierten Quicksort ist also*

n	1	2	3	4	5	10	20	30	40
H_n	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$				
$E(X)$	0	1	$\frac{8}{3} \approx 3$	$\frac{29}{6} \approx 5$	$\frac{37}{5} \approx 7$	≈ 24	≈ 71	≈ 128	≈ 190

Den Ausdruck in Lemma 2.8.1 werden wir jetzt noch abschätzen und in eine Laufzeitklasse einordnen. Dazu zeigen wir:

Lemma 2.8.6 *Für die n-te harmonische Zahl*

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

gilt

$$\ln(n) + \frac{1}{n} \leq H_n \leq \ln(n) + 1.$$

Beweis. Wir wenden dazu das nachfolgende Lemma 2.8.7 für die monoton fallende Funktion

$$f(i) = \frac{1}{i}$$

an. ■

Lemma 2.8.7 *Ist $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ eine monoton fallende Funktion,*

$$S_n = \sum_{i=1}^{n} f(i)$$

und

$$I_n = \int_1^n f(x)dx$$

dann gilt

$$I_n + f(n) \leq S_n \leq I_n + f(1).$$

Ist f monoton steigend, dann gilt analog

$$I_n + f(1) \leq S_n \leq I_n + f(n).$$

Beweis. Wie Abbildung 2.8 zeigt, ist $S_n - f(1)$ eine Untersumme von I_n. Wie Abbildung 2.9 zeigt, ist $S_n - f(n)$ eine Obersumme

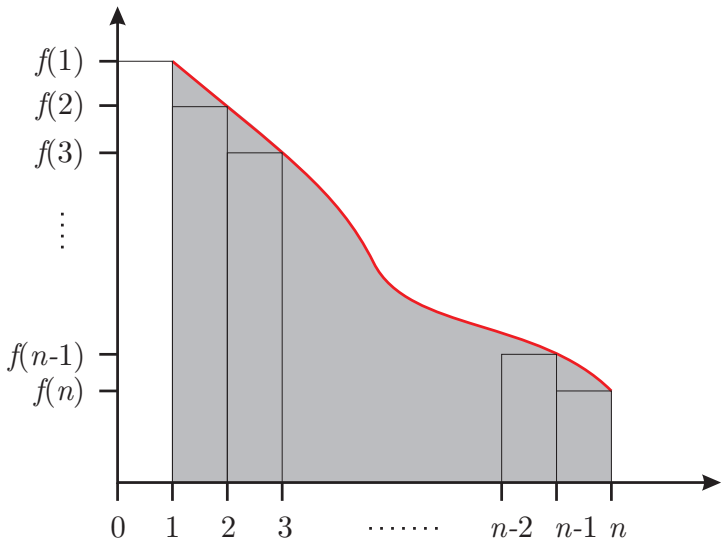


Abbildung 2.8: Untersumme von $f(x)$ auf $[1, n]$

von

$$\int_0^{n-1} f(x+1)dx.$$

Mit der Substitutionsregel gilt aber

$$\int_0^{n-1} f(x+1)dx = \int_1^n f(x)dx = I_n.$$

Wir erhalten also

$$S_n - f(1) \leq I_n$$
$$I_n \leq S_n - f(n)$$

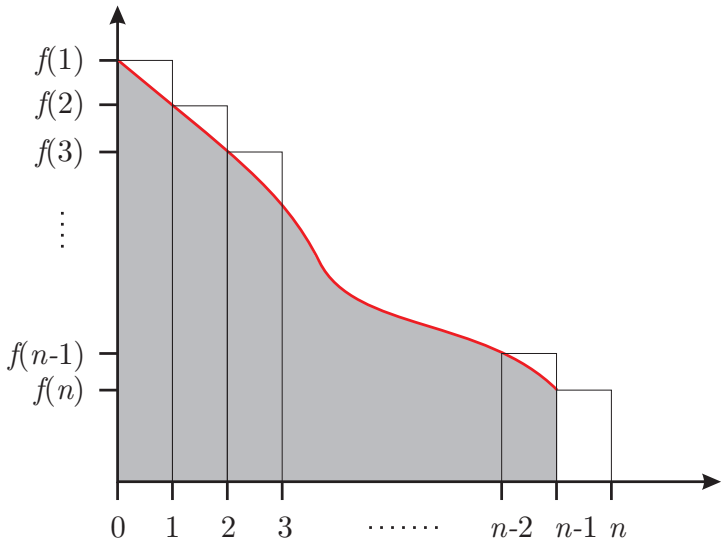


Abbildung 2.9: Obersumme von $f(x+1)$ auf $[0, n-1]$

d.h.

$$I_n + f(n) \leq S_n \leq I_n + f(1).$$

Die Ungleichungen für eine monoton steigende Funktion beweist man analog. ■

Satz 2.8.8 *Der Erwartungswert für die Laufzeit des randomisierten Quicksort-Algorithmus auf einer n-elementigen Menge ist in* $O(n \log(n))$.

Beweis. Wir haben nach Bemerkung 2.8.3 und Lemma 2.8.6, dass

$$\begin{aligned} E(X) &= \sum_{j=1}^{n} 2(H_j - 1) \\ &\leq 2 \cdot n \cdot (H_n - 1) \\ &\leq 2 \cdot n \cdot \ln(n). \end{aligned}$$

■

Ohne Beweis bemerken wir:

Bemerkung 2.8.9 *Die Zufallsvariable X beschreibe die Anzahl der Vergleiche in einem Lauf des randomisierten Quicksortalgorithmus auf einer n-elementigen Menge. Dann gilt für die Varianz von X, dass*

$$V(X) = 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n$$

wobei

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

wieder die n-te harmonische Zahl bezeichnet und

$$H_n^{(k)} = \sum_{i=1}^{n} \frac{1}{i^k}$$

für die n-te harmonische Zahl von Ordnung k steht. Standardabweichungen für die Laufzeit für verschiedene n gibt die folgende Tabelle:

$n =$	1	2	3	4	5	10	20	30	40
$E(X) \approx$	0	1	3	5	7	24	71	128	190
$\sigma(X) = \sqrt{V(X)} \approx$	0	0	0.47	0.89	1.4	3.9	10	16	22

2.9 Unabhängigkeit

2.9.1 Übersicht

Es stellt sich natürlich die Frage, ob es eine zu

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

vergleichbare Formel auch für das Produkt von Zufallsvariablen $X_i : \Omega \to N$ gibt, also ob

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2).$$

Wir bemerken zunächst, dass sich die Frage nur stellt, falls das Produkt

$$\begin{array}{rccc} X_1 \cdot X_2 : & \Omega & \to & N \\ & \omega & \mapsto & X(\omega) \cdot Y(\omega) \end{array}$$

Sinn macht, d.h. falls wir das Produkt $X(\omega) \cdot Y(\omega)$ berechnen können. Dazu fordern wir, dass N nicht nur die Struktur eines $\mathbb{R}$-Vektorraums hat, sondern auch die eines Rings, also dass N eine $\mathbb{R}$-Algebra ist. Wie das folgende Beispiel zeigt, ist $E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$ im Allgemeinen nicht richtig:

Beispiel 2.9.1 *Wir betrachten wieder einen Münzwurf mit Ergebnisraum*

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

und Wahrscheinlichkeitsfunktion

$$m(0) = m(1) = \frac{1}{2}.$$

Die Zufallsvariable X_1 definieren durch

$$X_1(0) = 1$$
$$X_1(1) = 0$$

und die Zufallsvariable X_2 durch

$$X_2(0) = 0$$
$$X_2(1) = 1.$$

Dann ist

$$E(X_1) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$
$$E(X_2) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

Dagegen ist $X_1 \cdot X_2 = 0$ also auch

$$E(X_1 \cdot X_2) = 0 \neq \frac{1}{4} = E(X_1) \cdot E(X_2).$$

Das Problem in dem Beispiel ist, dass die Zufallsvariablen X_1 und X_2 nicht unabhängig voneinander sind, tatsächlich gilt

$$X_1 = 1 - X_2.$$

Wie kann man eine solche Situation ausschliessen?

2.9.2 Unabhängigkeit von Zufallsvariablen

Notation 2.9.2 *Für Zufallsvariablen $X_1 : \Omega \to N_1$ und $X_2 : \Omega \to N_2$ bilden wir die Zufallsvariable*

$$(X_1, X_2) : \Omega \to N_1 \times N_2$$

Für die Verteilung dieser Zufallsvariable schreiben wir auch

$$P(X_1 = n_1, X_2 = n_2) := P((X_1, X_2) = (n_1, n_2)).$$

Definition 2.9.3 *Die Zufallsvariablen $X_1 : \Omega \to N_1$ und $X_2 : \Omega \to N_2$ heißen **unabhängig**, wenn*

$$P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1) \cdot P(X_2 = n_2)$$

für alle $(n_1, n_2) \in N_1 \times N_2$.

Beispiel 2.9.4 *Wir werfen dreimal eine Münze, also $\Omega = \{0,1\}^3$, wobei 0 für Kopf und 1 für Zahl steht. Die Zufallsvariablen*

$$X_1 : \begin{array}{ccc} \Omega & \to & \{0,1\} \\ \omega & \mapsto & \begin{cases} 1 & \text{falls die Anzahl von Kopf gerade} \\ 0 & \text{sonst} \end{cases} \end{array}$$

und

$$X_2 : \begin{array}{ccc} \Omega & \to & \{0,1\} \\ \omega & \mapsto & \begin{cases} 1 & \text{falls die ersten beiden Würfe dasselbe} \\ & \text{Ergebnis haben} \\ 0 & \text{sonst} \end{cases} \end{array}$$

haben die Werte

ω	000	001	010	100	011	101	110	111
$X_1(\omega)$	0	1	1	1	0	0	0	1
$X_2(\omega)$	1	1	0	0	0	0	1	1

also die Verteilungen

n	0	1
$P(X_1 = n)$	$\frac{1}{2}$	$\frac{1}{2}$

n	0	1
$P(X_2 = n)$	$\frac{1}{2}$	$\frac{1}{2}$

Die Zufallsvariable (X_1, X_2) hat die Verteilung

(n_1, n_2)	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
ω *mit* $X_1(\omega) = n_1$ *und* $X_2(\omega) = n_2$	011	010	000	001
	101	100	110	111
$P(X_1 = n_1, X_2 = n_2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Es gilt also stets

$$P(X_1 = n_1, X_2 = n_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X_1 = n_1) \cdot P(X_2 = n_2),$$

d.h. X_1 und X_2 sind unabhängig.

Beispiel 2.9.5 *Für den Münzwurf im vorangegangenen Beispiel nimmt die Zufallsvariable*

$$\begin{array}{rccl} X_3: & \Omega & \to & \{0,1,2,3\} \\ & \omega & \mapsto & \textit{Anzahl der Würfe von Zahl} \end{array}$$

die Werte

ω	000	001	010	100	011	101	110	111
$X_3(\omega)$	0	1	1	1	2	2	2	3

an, hat also die Verteilung

n	0	1	2	3
$P(X_3 = n)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Die Variablen X_2 und X_3 sind nicht unabhängig, denn wir haben zum Beispiel

ω *mit* $X_2(\omega) = 1$	000	001			110	111
ω *mit* $X_3(\omega) = 2$			011	101	110	
ω *mit* $X_2(\omega) = 1$ *und* $X_3(\omega) = 2$					110	

also ist die Wahrscheinlichkeit, dass die ersten beiden Würfe das selbe Ergebnis liefern und genau 2-mal Zahl kommt

$$P(X_2 = 1, X_3 = 2) = \frac{1}{8},$$

jedoch das Produkt der Wahrscheinlichkeit, dass die ersten beiden Würfe das selbe Ergebnis liefern und der Wahrscheinlichkeit, dass genau 2-mal Zahl kommt

$$P(X_2 = 1) \cdot P(X_3 = 2) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16} \neq \frac{1}{8}.$$

Bemerkung 2.9.6 *Wir können uns Unabhängigkeit von $X_1 : \Omega \to N_1$ und $X_2 : \Omega \to N_2$ auch so vorstellen: Die Verteilung von (X_1, X_2) verhält sich als ob X_1 und X_2 auf zwei verschiedenen Kopien von Ω definiert sind, und wir die Zufallsvariable*

$$\Omega \times \Omega \to N_1 \times N_2, \ (\omega_1, \omega_2) \mapsto (X_1(\omega_1), X_2(\omega_2))$$

betrachten: Da X_1 nur von ω_1 abhängt und X_2 nur von ω_2, gilt nach Bemerkung 2.3.18 dann, dass

$$P((X_1, X_2) = (n_1, n_2)) = P(X_1 = n_1) \cdot P(X_2 = n_2).$$

2.9.3 Erwartungswert des Produkts von unabhängigen Zufallsvariablen

Für die Untersuchung des Produkts von Zufallsvariablen $X_i : \Omega \to N$ beschränken wir uns im Folgenden auf den Fall $N = \mathbb{R}$, da wir hier die Konvergenzfragen verstehen.

Satz 2.9.7 *Sind $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen für die die Erwartungswerte existieren. Dann gilt*

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2).$$

Beweis. Wir bemerken, dass die Mengen $N_1 = \text{Bild}(X_1)$ und $N_2 = \text{Bild}(X_2)$ abzählbar sind (da Ω abzählbar ist). Wir sortieren

in Gleichung 2.2 die Elemente von Ω nach den Werten von X_1 und X_2 und verwenden in Gleichung 2.5 die Unabhängigkeit:[4]

$$E(X_1 \cdot X_2) = \sum_{\omega\in\Omega} (X_1(\omega)\cdot X_2(\omega))\cdot m(\omega) \tag{2.1}$$

$$= \sum_{n_1\in N_1}\sum_{n_2\in N_2} \sum_{\substack{\omega\in\Omega \\ X_1(\omega)=n_1 \\ X_2(\omega)=n_2}} n_1\cdot n_2\cdot m(\omega) \tag{2.2}$$

$$= \sum_{n_1\in N_1}\sum_{n_2\in N_2} n_1\cdot n_2\cdot \sum_{\substack{\omega\in\Omega \\ X_1(\omega)=n_1 \\ X_2(\omega)=n_2}} m(\omega) \tag{2.3}$$

$$= \sum_{n_1\in N_1}\sum_{n_2\in N_2} n_1\cdot n_2\cdot P(X_1=n_1, X_2=n_2) \tag{2.4}$$

$$= \sum_{n_1\in N_1}\sum_{n_2\in N_2} n_1\cdot n_2\cdot P(X_1=n_1)\cdot P(X_2=n_2) \tag{2.5}$$

$$= \left(\sum_{n_1\in N_1} n_1\cdot P(X_1=n_1)\right)\cdot\left(\sum_{n_2\in N_2} n_2\cdot P(X_2=n_2)\right) \tag{2.6}$$

$$= E(X_1)\cdot E(X_2). \tag{2.7}$$

■

Beispiel 2.9.8 *Für die Zufallsvariablen aus Beispiel 2.9.4 hat $X_1\cdot X_2$ die Verteilung*

n	0	1
(n_1,n_2) *mit* $n_1+n_2=n$	$(0,0),(1,0),(0,1)$	$(1,1)$
$P(X_1\cdot X_2=n)$	$\frac{3}{4}$	$\frac{1}{4}$

und daher ist

$$E(X_1\cdot X_2) = 0\cdot\frac{3}{4}+1\cdot\frac{1}{4}=\frac{1}{4}.$$

[4]Für die Konvergenzfrage müssen diese Rechnung wie üblich rückwärts lesen: Wir haben in der Gleichung 2.6 und Gleichung 2.3 die Rechenregeln für konvergente Reihen verwendet. Die Gleichheit 2.2 gilt, da wir wegen der absoluten Konvergenz der Reihe umordnen dürfen (siehe auch Bemerkung 2.6.5).

Mit

$$E(X_1) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$
$$E(X_2) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

gilt also

$$E(X_1 \cdot X_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = E(X_1) \cdot E(X_2).$$

Wir können an dem Beispiel auch unseren Beweis nachverfolgen:

$$\begin{aligned} E(X_1 \cdot X_2) &= \sum_{n_1=0,1} \sum_{n_2=0,1} n_1 \cdot n_2 \cdot \tfrac{1}{4} \\ &= \sum_{n_1=0,1} \sum_{n_2=0,1} n_1 \cdot n_2 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} \\ &= \sum_{n_1=0,1} n_1 \cdot \tfrac{1}{2} \cdot \sum_{n_2=0,1} n_2 \cdot \tfrac{1}{2} \\ &= E(X_1) \cdot E(X_2). \end{aligned}$$

2.9.4 Varianz von Summe und Produkt unabhängiger Zufallsvariablen

Wir diskutieren noch, wie sich die Varianz der Summe von Zufallsvariablen verhält. Im Allgemeinen kann man hier keine Aussage treffen, für unabhängige Zufallsvariablen haben wir jedoch:

Satz 2.9.9 *Sind $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen für die jeweils die Varianz existiert, dann gilt*

$$V(X_1 + X_2) = V(X_1) + V(X_2).$$

Beweis. Mit Satz 2.7.6, Satz 2.6.12 und Satz 2.9.7 gilt

$$\begin{aligned} V(X_1 + X_2) &= E((X_1 + X_2)^2) - (E(X_1 + X_2))^2 \\ &= E(X_1^2 + 2X_1X_2 + X_2^2) - (E(X_1) + E(X_2))^2 \\ &= E(X_1^2) + 2E(X_1X_2) + E(X_2^2) - (E(X_1) + E(X_2))^2 \\ &= E(X_1^2) + 2E(X_1)E(X_2) + E(X_2^2) - (E(X_1) + E(X_2))^2 \\ &= E(X_1^2) - E(X_1)^2 + E(X_2^2) - E(X_2)^2 \\ &= V(X_1) + V(X_2). \end{aligned}$$

■

Beispiel 2.9.10 *Für zweimal Würfeln, wobei X_i die Augenzahl des i-ten Würfels bezeichne, ist mit Beispiel 2.7.8 die Varianz der Augensumme*

$$V(X_1 + X_2) = V(X_1) + V(X_2) = 2 \cdot \frac{35}{12} = \frac{35}{6}$$

und ebenso die Varianz der Augendifferenz

$$\begin{aligned} V(X_1 - X_2) &= V(X_1) + V(-X_2) \\ &= V(X_1) + (-1)^2 \cdot V(X_2) \\ &= V(X_1 + X_2). \end{aligned}$$

wobei wir Satz 2.7.9 verwendet haben.

Zum Abschluss geben wir noch eine Formel für die Varianz des Produkts zweier unabhängiger Zufallsvariablen. Der Beweis ist Aufgabe 2.21.

Satz 2.9.11 *Sind $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen für die jeweils die Varianz existiert, dann gilt*

$$V(X_1 \cdot X_2) = V(X_1)V(X_2) + V(X_1)E(X_2)^2 + V(X_2)E(X_1)^2.$$

2.10 Korrelation von Zufallsvariablen

2.10.1 Anwendungsbeispiel

In der Praxis werden Wahrscheinlichkeitsverteilungen von Zufallsvariablen statistisch gemessen, d.h. man zählt wie oft bestimmte Werte angenommen werden und verwendet statt Wahrscheinlichkeiten die relativen Häufigkeiten (wir wir schon in Beispiel 2.6.1 gesehen haben ist z.B. der mit den relativen Häufigkeiten berechnete Erwartungswert der Mittelwert der Messwerte).

Nehmen wir an, wir wollen an einer Anzahl von Menschen (etwa Teilnehmern einer Studie) untersuchen, ob zwischen bestimmten Eigenschaften ein Kausalzusammenhang besteht, etwa zwischen den Zufallsvariablen (mit Wert 1 oder 0 je nachdem, ob die jeweilige Eigenschaft zutrifft oder nicht)

X_1 : die Person ist raucht

X_2 : die Person bekommt Lungenkrebs

oder

X_1 : die Person stirbt

X_2 : die Person ist zu Hause.

Zunächst müssen wir uns klarmachen, dass wir durch solche statistischen Messungen keine Kausalbeziehung beweisen können. Wir werden feststellen können, dass es in beiden Fällen einen starken Zusammenhang zwischen X_1 und X_2 gibt. Jedoch gibt es im ersten Beispiel eine Kausalbeziehung, im zweiten Fall sicher nicht. Man muss auch bemerken, dass eine Kausalbeziehung keine Implikation im mathematischen Sinne ist, denn es gibt ja auch Raucher, die keinen Lungenkrebs bekommen.

Wie kann man nun einen Zusammenhang messen? Die einfache Antwort ist die Unabhängigkeit von $X_1 : \Omega \to N_1$ und $X_2 : \Omega \to N_2$ zu untersuchen. Bei einer Studie kennen wir für jede Person $\omega \in \Omega$ die Werte von X_1 und X_2, also die Verteilung von X_1 von X_2 und (X_1, X_2), also alle Wahrscheinlichkeiten

$$\begin{aligned} &P(X_1 = n_1) \\ &P(X_2 = n_2) \\ &P(X_1 = n_1, X_2 = n_2). \end{aligned}$$

Wir könnten also testen ob

$$P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1) \cdot P(X_2 = n_2).$$

Bemerkung 2.10.1 *Die Verteilung von (X_1, X_2) bestimmt schon die Verteilungen von X_1 und X_2, die sogenannten **Randverteilungen**, denn*

$$P(X_1 = n_1) = \sum_{n_2 \in N_2} P(X_1 = n_1, X_2 = n_2).$$

Die Umkehrung ist i.A. nicht richtig, sie gilt genau dann, wenn X_1 und X_2 unabhängig sind.

Beispiel 2.10.2 *Wir schreiben die Wahrscheinlichkeiten* $P(X_1 = n_1, X_2 = n_2)$ *in ein Diagramm, wobei wir* $X_1 = 1$ *für Raucher und* $X_1 = 0$ *für Nichtraucher, und* $X_2 = 1$ *für Lungenkrebs und* $X_2 = 0$ *für keinen Lungenkrebs. Aus einer Studie erhalten wir die folgenden relativen Häufigkeiten*

	1	0	n_2
1	0.038	0.262	
0	0.002	0.698	
n_1			

(siehe auch Übungsaufgabe 2.22). Durch addieren der relativen Häufigkeiten in den Spalten bzw. Zeilen erhalten wir mit Bemerkung 2.10.1 die Verteilungen von X_1 *und* X_2*:*

		1	0	n_2
		0.04	0.96	$P(X_2 = n_1)$
1	0.3	0.038	0.262	
0	0.7	0.002	0.698	
n_1	$P(X_1 = n_1)$			

Beispiel 2.10.3 *In der obigen Verteilung gilt*

$$P(X_1 = 1) \cdot P(X_2 = 1) = 0.04 \cdot 0.3 = 0.012$$
$$\neq 0.038 = P(X_1 = 1, X_2 = 1)$$

Wir können also sagen, dass X_1 *und* X_2 *nicht unabhängig sind.*

In der Praxis ist eine solche true-false-Aussage allerdings nicht ausreichend. Das sieht man schon daran, dass gemessene relative Häufigkeiten nie exakt sind, die Wahrscheinlichkeit, dass für die gemessenen relative Häufigkeiten von unabhängigen Zufallsvariablen

$$P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1) \cdot P(X_2 = n_2)$$

gilt, ist also praktisch gleich Null. In dem obigen Beispiel ist ziemlich deutlich

$$0.012 < 0.038,$$

also Raucher erkranken wesenlich häufiger an Lungenkrebs als man das für unabhängige Zufallsvariablen erwarten würde (nämlich die Wahrscheinlichkeit zu rauchen mal die Wahrscheinlichkeit Lungenkrebs zu bekommen, also $0.04 \cdot 0.3 = 0.012$). Oft ist

das aber nicht so klar. Wie kann man also einen Zusammenhang zwischen zwei Zufallsvariablen quantifizieren? Dazu verallgemeinern wir zunächst die Definition der Varianz.

2.10.2 Covarianz

Definition 2.10.4 *Sind $X_1, X_2 : \Omega \to \mathbb{R}$ zwei Zufallsvariablen, dann ist die **Covarianz** von X_1 und X_2*

$$\mathrm{Cov}(X_1, X_2) = E(\ (X_1 - E(X_1)) \cdot (X_2 - E(X_2))\ ).$$

Bemerkung 2.10.5 *1) Als Spezialfall erhalten wir die Varianz*

$$V(X) = \mathrm{Cov}(X, X).$$

2) Offenbar ist

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1).$$

Analog zu Satz 2.7.6 haben wir (auf einem diskreten Wahrscheinlichkeitsraum Ω):

Satz 2.10.6 *Für zwei Zufallsvariablen $X_1, X_2 : \Omega \to \mathbb{R}$ gilt*

$$\mathrm{Cov}(X_1, X_2) = E(X_1 \cdot X_2) - E(X_1) \cdot E(X_2).$$

Beweis. Folgt wieder mit Bemerkung 2.6.14 und Satz 2.6.12. Der Beweis ist Aufgabe 2.23. ■

Beispiel 2.10.7 *Wie in Beispiel 2.9.4 werfen wir dreimal eine Münze, also $\Omega = \{0, 1\}^3$, wobei 0 für Kopf und 1 für Zahl steht.*

- *Die Zufallsvariablen X_1 sei 1 falls die Anzahl von Kopf gerade ist und 0 sonst.*
- *Die Zufallsvariable X_2 sei 1 falls die ersten beiden Würfe dasselbe Ergebnis haben und 0 sonst.*

Wir haben dann die Ergebnisse und Werte der Zufallsvariablen

ω	000	001	010	100	011	101	110	111
$X_1(\omega)$	0	1	1	1	0	0	0	1
$X_2(\omega)$	1	1	0	0	0	0	1	1
$(X_1 \cdot X_2)(\omega)$	0	1	0	0	0	0	0	1

und damit die Verteilungen

n	0	1
$P(X_1 = n)$	$\frac{1}{2}$	$\frac{1}{2}$

n	0	1
$P(X_2 = n)$	$\frac{1}{2}$	$\frac{1}{2}$

und

n	0	1
$P(X_1 \cdot X_2 = n)$	0	$\frac{1}{4}$

Die Erwartungswerte sind also

$$E(X_1) = E(X_2) = \frac{1}{2}$$
$$E(X_1 \cdot X_2) = \frac{1}{4}$$

also nach Satz 2.10.6 die Covarianz von X_1 und X_2

$$\mathrm{Cov}(X_1, X_2) = \frac{1}{2} \cdot \frac{1}{2} - \frac{1}{4} = 0.$$

Dies gilt allgemein für unabhängige Zufallsvariablen:

Corollar 2.10.8 *Für unabhängige Zufallsvariablen $X_1, X_2 : \Omega \to \mathbb{R}$ ist*

$$\mathrm{Cov}(X_1, X_2) = 0.$$

Beweis. Folgt sofort aus Satz 2.10.6 und Satz 2.9.7. ■

Beispiel 2.10.9 *Bei zweimal Würfeln, wobei X_1 die Augenzahl des ersten Wurfs und X_2 die Augenzahl des zweiten Wurfs angibt, gilt für Augensumme und Augendifferenz*

$$\begin{aligned}\mathrm{Cov}(X_1 + X_2, X_1 - X_2) &= E((X_1 + X_2) \cdot (X_1 - X_2)) \\ &\quad - E(X_1 + X_2) \cdot E(X_1 - X_2) \\ &= E(X_1^2 - X_2^2) = 0\end{aligned}$$

denn

$$E(X_1^2 - X_2^2) = E(X_1^2) - E(X_2^2) = 0$$
$$E(X_1 - X_2) = E(X_1) - E(X_2) = 0$$

da X_1 und X_2 dieselbe Verteilung haben.

Trotzdem sind X_1 und X_2 nicht unabhängig: Wir bestimmen die $\omega \in \Omega = \{1,\dots,6\}^2$ mit $(X_1 + X_2)(\omega) = 3$ und die $\omega \in \Omega$ mit $(X_1.X_2)(\omega) = 0$:

$\omega \in \Omega$ *mit* $(X_1 + X_2)(\omega) = 3$	$(1,2),(2,1)$
$\omega \in \Omega$ *mit* $(X_1 - X_2)(\omega) = 0$	$(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)$

also ist

$$P((X_1 + X_2)(\omega) = 3) = \tfrac{2}{36} = \tfrac{1}{18}$$
$$P((X_1 - X_2)(\omega) = 0) = \tfrac{6}{36} = \tfrac{1}{6}$$

aber sogar

$$P((X_1 + X_2)(\omega) = 3,\ (X_1 - X_2)(\omega) = 0) = 0 \neq \frac{1}{18} \cdot \frac{1}{6}.$$

Siehe dazu auch Aufgabe 2.24.

Bemerkung 2.10.10 *Es ist*

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\,\mathrm{Cov}(X_1, X_2).$$

Insbesondere gilt die Gleichung $V(X_1+X_2) = V(X_1)+V(X_2)$ sogar wenn X_1 und X_2 nicht unabhängig sind (wie in Satz 2.9.9), sondern nur $\mathrm{Cov}(X_1, X_2) = 0$ ist. Tatsächlich zeigt unsere Rechnung, dass die Gleichung eine Charakterisierung von $\mathrm{Cov}(X_1, X_2) = 0$ ist, d.h.

$$\begin{array}{lll} X_1,\ X_2 \text{ unabhängig} & \Rightarrow & \\ \mathrm{Cov}(X_1, X_2) = 0 & \Leftrightarrow & V(X_1 + X_2) = V(X_1) + V(X_2). \end{array}$$

Beweis. Die Aussagen (1) und (2) sind klar, die Aussage (3) erhalten wir aus der Definition der Varianz und Satz 2.6.12, da

$$\begin{aligned}(X_1 + X_2 - E(X_1) - E(X_2))^2 = {} & X_1^2 - E(X_1)^2 + X_2^2 - E(X_2)^2 \\ & + (X_1 - E(X_1)) \cdot (X_2 - E(X_2)).\end{aligned}$$

■

Beispiel 2.10.11 *Wir bestimmen die für die Verteilung*

		1	0	n_2
		0.04	0.96	$P(X_2 = n_1)$
1	0.3	0.038	0.262	
0	0.7	0.002	0.698	
n_1	$P(X_1 = n_1)$			

aus Beispiel 2.10.2 die Covarianz. Die Verteilung von $X_1 \cdot X_2$ ist

n	0	1
$P(X_1 \cdot X_2 = n)$	0.962	0.038

Wir haben also

$$E(X_1) = 0.3$$
$$E(X_2) = 0.04$$
$$E(X_1 \cdot X_2) = 0.038$$

und damit

$$\mathrm{Cov}(X_1, X_2) = 0.038 - 0.3 \cdot 0.04 = 0.026.$$

Nach Corollar 2.10.8 können wir also folgern, dass X_1 und X_2 nicht unabhängig sind. Aber was sagt uns diese Zahl sonst noch?

2.10.3 Korrelation

Das Problem der Covarianz liegt darin, dass unklar ist, was der eigentliche Zahlenwert bedeutet. Wir müssen ihr relativ zu einer anderen Zahl sehen. Man vergleicht die Covarianz mit dem Produkt der Standardabweichungen der Zufallsvariablen:

Definition 2.10.12 *Sind $X_1, X_2 : \Omega \to \mathbb{R}$ Zufallsvariablen mit positiver Varianz, dann ist die **Korrelation** von X_1 und X_2*

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma(X_1) \cdot \sigma(X_2)}.$$

Die Bedingung positive Varianz stellt hier sicher, dass der Nenner nicht 0 ist. Für die Korrelation erhalten wir dann:

Satz 2.10.13 *Für die Korrelation von zwei Zufallsvariablen* $X_1, X_2 : \Omega \to \mathbb{R}$ *mit positiver Varianz gilt*

$$-1 \leq \mathrm{Corr}(X_1, X_2) \leq 1$$

insbesondere existiert die Korrelation.

Für $\mathrm{Corr}(X_1, X_2) > 0$ heißen X_1 und X_2 **korreliert**, falls $\mathrm{Corr}(X_1, X_2) < 0$ dann **antikorreliert**. Ein Wert der Korrelation nahe bei ± 1 zeigt an, dass die Werte von X_1 und X_2 stark gleichförmig verhalten, ein Wert nahe bei -1, dass sich X_1 und X_2 stark gegensätzlich verhalten. Was ein Wert von 0 bedeutet ist etwas diffiziler und wir werden darauf noch zurückkommen (wir haben schon gesehen daraus nicht folgt, dass X_1 und X_2 unabhängig sind, siehe Beispiel 2.10.9).

Beispiel 2.10.14 *Vergleichen wir eine Zufallsvariable mit sich selbst erhalten wir mit Bemerkung 2.10.5*

$$\mathrm{Corr}(X, X) = \frac{V(X)}{\sigma(X) \cdot \sigma(X)} = 1.$$

Beispiel 2.10.15 *Für Beispiel 2.10.2 bestimmen wir die Varianzen*

$$V(X_1) = E(X_1^2) - E(X_1)^2 = 0.3 - 0.3^2 = 0.21$$
$$V(X_2) = E(X_2^2) - E(X_2)^2 = 0.04 - 0.04^2 = 0.0384$$

also sind

$$\sigma(X_1) \approx 0.458$$
$$\sigma(X_2) \approx 0.196$$

und damit

$$\mathrm{Corr}(X_1, X_2) \approx \frac{0.026}{0.458 \cdot 0.196} \approx 0.289.$$

Die Zufallsvariablen X_1 *und* X_2 *sind also korreliert. Die Korrelation ist deutlich größer als* 0 *aber eben auch nicht* 1. *Das liegt daran, dass der Anteil an Raucher in der Gesamtbevölkerung mit* 0.3 *wesentlich größer ist als der Anteil* 0.04 *an Menschen die Lungenkrebs bekommen. Dies wiederum kann daran liegen, dass Raucher an anderen Erkrankungen frühzeitig versterben, bevor sie überhaupt Lungenkrebs bekommen.*

Bemerkung 2.10.16 *Sei X eine Zufallsvariable mit positiver Varianz $a, b \in \mathbb{R}$ mit $a \neq 0$. Dann ist*

$$\mathrm{Corr}(X,\ a \cdot X + b) = \mathrm{sgn}(a) = \begin{cases} 1 & \text{falls } a > 0 \\ -1 & \text{falls } a < 0 \end{cases}$$

wobei $\mathrm{sgn}(a)$ *das Vorzeichen von a bzeichne. Dies folgt direkt aus Satz 2.6.12 und Satz 2.7.9 (der Beweis ist Aufgabe 2.27).*

Beispiel 2.10.17 *Betrachten wir in Beispiel 2.10.2 die Zufallsvariable $1 - X_1$, die den Wert 1 annimmt, falls die Person nicht raucht, dann ist*

$$\mathrm{Corr}(X_1, 1 - X_1) = -1.$$

Allgemeiner gilt:

Bemerkung 2.10.18 *Für Zufallsvariablen X_1 und X_2 und $a, b \in \mathbb{R}$ ist*

$$\mathrm{Cov}(X_1, a \cdot X_2 + b) = a \cdot \mathrm{Cov}(X_1, X_2)$$

und falls X_1 und X_2 positive Varianz haben und $a \neq 0$ ist haben wir

$$\mathrm{Corr}(X_1, a \cdot X_2 + b) = \mathrm{sgn}(a) \cdot \mathrm{Corr}(X_1, X_2).$$

Beispiel 2.10.19 *In Beispiel 2.10.2 ist die Korrelation zwischen der Zufallsvariable $1 - X_1$, die den Wert 1 annimmt, falls die Person nicht raucht und X_2 die den Wert 1 annimmt falls die Person Lungenkrebs bekommt*

$$\mathrm{Corr}(1 - X_1, X_2) = -0.289,$$

die beiden Variablen sind also antikorreliert.

Zum Beweis von Satz 2.10.13 verwenden wir die Cauchy-Schwarz-Ungleichung, deren Beweis wir im nachfolgenden Abschnitt diskutieren werden.

Lemma 2.10.20 (Cauchy-Schwarz für Zufallsvariablen) *Sind $X_1, X_2 : \Omega \to \mathbb{R}$ Zufallsvariablen für die $E(X_i^2)$ existiert, dann gilt*

$$E(X_1 \cdot X_2)^2 \leq E(X_1^2) \cdot E(X_2^2).$$

Nun zum Beweis von Satz 2.10.13:

Beweis. Wir wenden die Cauchy-Schwarz-Ungleichung auf die Zufallsvariablen $X_1 - E(X_1)$ und $X_2 - E(X_2)$ an:

$$\begin{aligned} &E(\ (X_1 - E(X_1)) \cdot (X_2 - E(X_2))\ )^2 \\ &\leq E((X_1 - E(X_1))^2)\ \cdot\ E((X_2 - E(X_2))^2) \end{aligned}$$

also

$$\mathrm{Cov}(X_1, X_2)^2 \leq V(X_1) \cdot V(X_2)$$

und damit

$$|\mathrm{Corr}(X_1, X_2)| \leq 1.$$

■

Wir haben schon in Bemerkung 2.10.16 gesehen, dass ein linearer Zusammenhang von Zufallsvariablen zu Korrelation ± 1 führt, ebenso aber auch in Beispiel 2.10.9, dass für abhängige Zufallsvariablen die Korrelation 0 sein kann. Hier ist ein weiteres seltsames Beispiel einer direkten quadratischen Abhängigkeit, die dennoch zu Korrelation 0 führt.

Beispiel 2.10.21 *Sei* $\Omega = \{-1, 0, 1\}$ *mit* $m(\omega) = \frac{1}{3}$ *und* $X_1(\omega) = \omega$ *und* $X_2 = X_1^2$. *Offensichtlich sind* X_1 *und* X_2 *nicht unabhängig. Formal sehen wir das auch, da z.B.*

$$P(X_1 = -1, X_2 = 1) = \frac{1}{3} \neq \frac{1}{3} \cdot \frac{2}{3} = P(X_1 = -1) \cdot P(X_2 = 1).$$

Weiter ist

$$E(X_1) = 0$$

und

$$E(X_1 \cdot X_2) = E(X_1^3) = 0$$

also

$$\mathrm{Corr}(X_1, X_1^2) = 0.$$

Der Grund hierfür ist, dass die Korrelation nur lineare Abhängigkeiten misst. Dies werden wir im Folgenden beweisen. Ergebnisse mit Wahrscheinlichkeit 0 tragen zu Erwartungswerten und insbesondere zu der Korrelation nicht bei. Deshalb können wir nur auf eine Aussage bis auf solche Ergebnisse treffen:

Definition 2.10.22 *Zwei Zufallsvariablen* $X_1, X_2 : \Omega \to \mathbb{R}$ *heißen* ***stochastisch äquivalent****, geschrieben*

$$X_1 \sim X_2$$

wenn

$$P(X_1 - X_2 = 0) = 1.$$

Dies bedeutet also, dass das Ereignis aller $\omega \in \Omega$ mit $X_1(\omega) \neq X_2(\omega)$ Wahrscheinlichkeit 0 hat.

Lemma 2.10.23 *In der Cauchy-Schwarz-Ungleichung aus Lemma 2.10.20 gilt Gleichheit genau dann, wenn es* $a, b \in \mathbb{R}$ *nicht beide* 0 *gibt mit*

$$a \cdot X_1 + b \cdot X_2 \sim 0.$$

Auch dies werden wir im nachfolgenden Abschnitt beweisen.

Bemerkung 2.10.24 *Für Zufallsvariablen* X_1 *und* X_2 *mit positiver Varianz gilt*

$$|\mathrm{Corr}(X_1, X_2)| = 1$$

genau dann, wenn es es $a, b, c \in \mathbb{R}$ *gibt mit* a, b *nicht beide* 0 *und*

$$a \cdot X_1 + b \cdot X_2 + c \sim 0.$$

Beweis. Aus dem Beweis von Satz 2.10.13 sehen wir, dass

$$|\mathrm{Corr}(X_1, X_2)| = 1$$

genau dann, wenn

$$\mathrm{Cov}(X_1, X_2)^2 = V(X_1) \cdot V(X_2),$$

nach Lemma 2.10.23 also genau dann, wenn es $a, b \in \mathbb{R}$ nicht beide 0 gibt mit

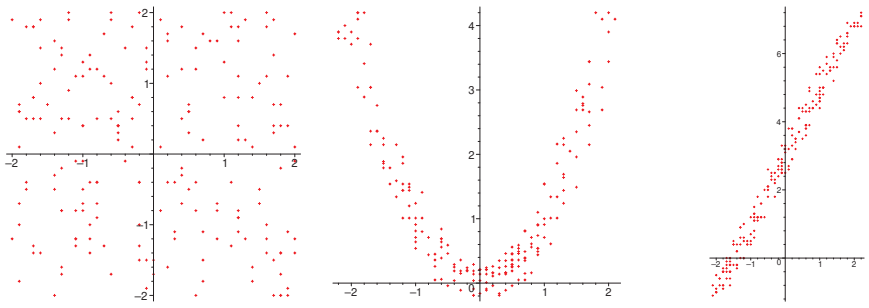
$$a \cdot (X_1 - E(X_1)) + b \cdot (X_2 - E(X_2)) \sim 0.$$

Es gibt also ein $c \in \mathbb{R}$ mit

$$a \cdot X_1 + b \cdot X_2 + c \sim 0.$$

Für die andere Schlussrichtung der Aussage im Lemma überträgt sich direkt der Beweis von Bemerkung 2.10.16 auf einen linearen Zusammenhang bis auf stochastische Äquivalenz. ∎

In den Diagrammen in Abbildung 2.10.3 stellen wir jeweils für ein Zufallsexperiment die Punkte $(X_1(\omega), X_2(\omega))$ dar, wobei wir alle ω als gleich wahrscheinlich annehmen, und geben die Korrelation $\mathrm{Corr}(X_1, X_2)$ an.



$\mathrm{Corr}(X_1, X_2) \approx -0.11$ $\mathrm{Corr}(X_1, X_2) \approx -0.14$ $\mathrm{Corr}(X_1, X_2) \approx 0.99$

Abbildung 2.10: Korrelationen von Paaren von Zufallsvariablen.

2.10.4 Beweis der Cauchy-Schwarz-Ungleichung

Definition 2.10.25 *Sei V ein $\mathbb{R}$-Vektorraum. Ein **Skalarprodukt** auf V ist eine Abbildung*

$$\langle -, - \rangle : V \times V \to \mathbb{R}$$

1) die in beiden Argumenten linear ist, also

$$\langle a \cdot u + b \cdot v, w \rangle = a \cdot \langle u, w \rangle + b \cdot \langle v, w \rangle$$

für alle $a, b \in \mathbb{R}$ und $u, v, w \in V$, und analog im zweiten Argument,

2) die symmetrisch ist, also

$$\langle v, w \rangle = \langle w, v \rangle$$

für alle $v, w \in V$, und

3) die positiv definit ist, d.h.

$$\langle v, v \rangle \geq 0$$

und

$$\langle v, v \rangle = 0 \Leftrightarrow v = 0.$$

für alle $v \in V$. Die Abbildung

$$\|-\| : V \to \mathbb{R}$$
$$\|v\| = \sqrt{\langle v, v\rangle}$$

*bezeichnen wir als die von $\langle -, -\rangle$ induzierte **Norm**.*

Beispiel 2.10.26 *Auf $V = \mathbb{R}^2$ ist durch*

$$\langle v, w\rangle = v_1 w_1 + v_2 w_2$$

*für $v, w \in \mathbb{R}^2$ das sogenannte **Euklidische Skalarprodukt** gegeben (als Übung überprüfe mann, dass dies tatsächlich ein Skalarprodukt ist). Wir erhalten z.B.*

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 0$$
$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = 1.$$

Die induzierte Norm

$$\|v\| = \sqrt{v_1^2 + v_2^2}$$

*gibt genau die **Euklidische Länge** aus dem Satz von Pythagoras, siehe Abbildung 2.11. Wir erhalten z.B.*

$$\left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| = 1$$
$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\| = \sqrt{2}.$$

Lemma 2.10.27 (Cauchy-Schwarz-Ungleichung) *Sei $\langle -, -\rangle$ ein Skalarprodukt auf dem $\mathbb{R}$-Vektorraum V. Für alle $v, w \in V$ gilt*

$$\langle v, w\rangle^2 \leq \|v\|^2 \cdot \|w\|^2$$

und es gilt Gleicheit genau dann, wenn v und w linear abhängig sind.

Beweis. Für $w = 0$ sind die Aussagen klar.

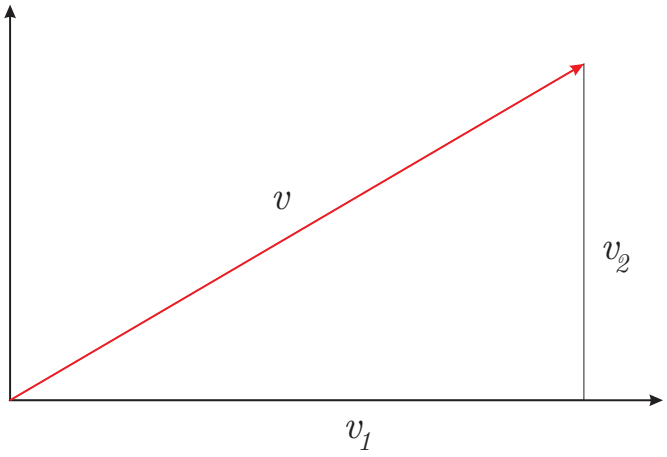


Abbildung 2.11: Euklidische Länge und der Satz von Pythagoras.

1) Für $w \neq 0$ und beliebiges $t \in \mathbb{R}$ ist

$$0 \leq \langle v - t \cdot w,\ v - t \cdot w \rangle = \|v\|^2 - 2t \langle v, w \rangle + t^2 \|w\|^2 =: f(t)$$

insbesondere ist die Funktion $f(t)$ eine nach oben geöffnete Parabel. Das Minimum von f wird daher angenommen für

$$t = \frac{\langle v, w \rangle}{\|w\|^2}$$

und hat den Wert

$$0 \leq \|v\|^2 - 2\frac{\langle v, w \rangle^2}{\|w\|^2} + \frac{\langle v, w \rangle^2}{\|w\|^4} \|w\|^2 = \|v\|^2 - \frac{\langle v, w \rangle^2}{\|w\|^2},$$

es gilt also die behauptete Ungleichung.

2) Sind v und w linear abhängig (und $w \neq 0$), dann ist für eine geeignete Wahl von t

$$v - t \cdot w = 0,$$

das Minimum der Parabel hat also den Wert

$$\langle v - t \cdot w, v - t \cdot w \rangle = 0.$$

3) Ist umgekehrt $\langle v, w\rangle^2 = \|v\|^2 \cdot \|w\|^2$, dann ist für alle t

$$\begin{aligned}\langle v - t\cdot w,\ v - t\cdot w\rangle &= \|v\|^2 \pm 2t\,\|v\|\cdot\|w\| + t^2\,\|w\|^2\\ &= (\|v\| \pm t\,\|w\|)^2\end{aligned}$$

und für geeignetes t ist der Ausdruck auf der rechten Seite 0. Es folgt $v = t\cdot w$, also sind v und w linear abhängig.

■

Definition 2.10.28 *Für $v, w \neq 0$ definieren wir den **Winkel** φ zwischen v und w durch*

$$\cos(\varphi) = \frac{\langle v, w\rangle}{\|v\|\cdot\|w\|}$$

was gemäß der Cauchy-Schwarz-Ungleichung wohldefiniert ist. Da der Cosinus 2π-periodisch ist, auf dem Intervall $[0, \pi]$ streng monoton fällt und mit dem Zwischenwertsatz jeden Wert zwischen 1 und -1 annimmt, gibt es ein eindeutiges

$$0 \leq \varphi \leq \pi,$$

das die obige Gleichung erfüllt. Siehe auch Abbildung 2.12. Dann

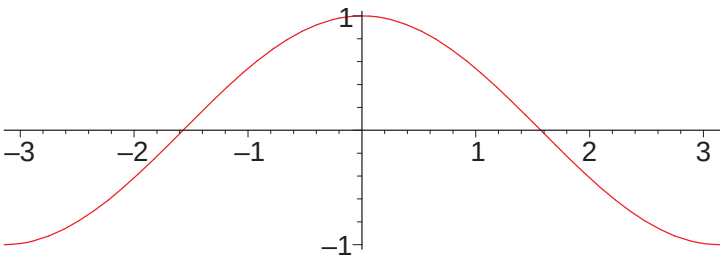


Abbildung 2.12: Cosinus.

haben wir z.B.

$$v \text{ und } w \text{ sind orthogonal} :\Longleftrightarrow \langle v, w\rangle = 0 \Longleftrightarrow \varphi = \frac{\pi}{2} = 90°.$$

Bemerkung 2.10.29 *Der Einfachheit halber nehmen wir $V = \mathbb{R}^2$ an (der allgemeine Fall geht genauso). Der Winkel φ erfüllt alle Eigenschaften, die man von einem Winkel erwarten würde:*

- *Vertauschen von v und w ändert φ nicht.*
- *Skalieren wir v oder w um eine positive Konstante, dann ändert sich φ nicht.*
- *Wir können damit annehmen, dass $\|v\| = \|w\| = 1$, also auf einem Einheitskreis liegen. Schreiben wir*

$$v = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} \qquad w = \begin{pmatrix} \cos\beta \\ \sin\beta \end{pmatrix}$$

mit $0 \leq \alpha, \beta < 2\pi$, dann ist

$$\cos(\varphi) = \langle v, w \rangle = \cos\alpha \cdot \cos\beta + \sin\alpha \cdot \sin\beta = \cos(\alpha - \beta),$$

wobei wir eines der Additionstheoreme verwendet haben. Der Winkel φ ist damit (der kleinste) Winkel zwischen α und β (Übung).

Auf einem Wahrscheinlichkeitsraum Ω betrachten wir jetzt den Vektorraum $V/\sim$ aller Zufallsvariablen X, für die $E(X^2)$ existiert, modulo stochastischer Äquivalenz $\sim$. Durch

$$\langle X, Y \rangle = E(X \cdot Y)$$

ist dann ein Skalarprodukt auf V gegeben. Anwendung der Cauch-Schwarz-Ungleichung aus Lemma 2.10.27 gibt direkt Lemma 2.10.20 und Bemerkung 2.10.24.

2.11 Bedingte Wahrscheinlichkeiten

2.11.1 Definition und Beispiele

Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$. In vielen Situationen will man die Wahrscheinlichkeit wissen, mit der ein Ergebnis ω auftritt unter der Voraussetzung, dass ein festgelegtes Ereignis E eintritt. Klar ist, dass für $\omega \notin E$, das Ergebnis nicht eintreten kann, die Wahrscheinlichkeit ist also 0. Für alle $\omega \in E$ sollte die Wahrscheinlichkeit $m'(\omega)$, dass ω eintritt unter der Voraussetzung,

dass E eintritt proportional zu $m(\omega)$ sein, das heißt es gibt ein $c \in \mathbb{R}$ mit

$$m'(\omega) = c \cdot m(\omega) \text{ für alle } \omega \in E.$$

Da andererseits die Wahrscheinlichkeit des Ereignisses E unter der Voraussetzung, dass E eintritt gleich 1 sein muss, gilt

$$1 = \sum_{\omega \in \Omega} m'(\omega) = \sum_{\omega \in E} m'(\omega) = c \cdot \sum_{\omega \in E} m(\omega)$$

also

$$c = \frac{1}{\sum_{\omega \in E} m(\omega)} = \frac{1}{P(E)}.$$

Definition 2.11.1 *Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$. Für ein Ereignis $E \subset \Omega$ mit $P(E) > 0$ und ein Ergebnis $\omega \in \Omega$ schreiben wir für die* ***bedingte Wahrscheinlichkeit*** *von ω unter der Voraussetzung, dass E eintritt,*

$$m(\omega \mid E) = \begin{cases} \frac{m(\omega)}{P(E)} & \text{falls } \omega \in E \\ 0 & \text{sonst} \end{cases}$$

Bemerkung 2.11.2 *Es gilt dann*

$$\sum_{\omega \in \Omega} m(\omega \mid E) = \sum_{\omega \in E} m(\omega \mid E) = \frac{P(E)}{P(E)} = 1,$$

durch

$$\begin{array}{rccc} m(- \mid E) : & \Omega & \to & \mathbb{R}_{\geq 0} \\ & \omega & \mapsto & m(\omega \mid E) \end{array}$$

ist also wieder eine Wahrscheinlichkeitsfunktion auf Ω gegeben.

Beispiel 2.11.3 *Wir werfen einen Würfel, also sei $\Omega = \{1, \ldots, 6\}$ und $m(\omega) = \frac{1}{6}$. Sei*

$$E = \{5, 6\}$$

das Ereignis, dass eine Zahl > 4 gewürfelt wird. Somit ist die bedingte Wahrscheinlichkeit, dass 5 gewürfelt wird unter der Voraussetzung, dass E eintritt,

$$m(5 \mid E) = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}.$$

Genauso ist

$$m(6 \mid E) = \frac{1}{2}$$

und

$$\sum_{\omega=1}^{6} m(\omega \mid E) = 0 + 0 + 0 + 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

Definition 2.11.4 *Für ein Ereignis $F \subset \Omega$ schreiben wir für die bedingte Wahrscheinlichkeit von F unter der Voraussetzung, dass E eintritt,*

$$P(F \mid E) = \sum_{\omega \in F} m(\omega \mid E)$$

Bemerkung 2.11.5 *Es gilt dann*

$$P(F \mid E) = \sum_{\omega \in F \cap E} \frac{m(\omega)}{P(E)} = \frac{P(F \cap E)}{P(E)},$$

insbesondere für $F \subset E$ ist

$$P(F \mid E) = \frac{P(F)}{P(E)},$$

also ist insbesondere

$$P(E \mid E) = 1,$$

wie erwartet.

Beispiel 2.11.6 *Unter allen Deutschen werden* 90% *der Männer mindestens* 60 *Jahre alt, und* 59% *mindestens* 80 *Jahre alt. Was ist die Wahrscheinlichkeit, dass ein mindestens* 60 *Jahre alter Mann mindestens* 80 *wird?*

Ist Ω die Menge aller Männer, E die Menge der Männer, die mindestens 60 *werden und F die Menge der Männer die mindestens* 80 *werden, dann müssen wir*

$$P(F \mid E) = \frac{P(F)}{P(E)} = \frac{0.59}{0.90} \approx 0.66$$

berechnen, wobei wir verwendet haben, dass $F \subset E$ (hier haben wir mit unserem Problem Glück gehabt, i.A. bräuchten wir noch Zusatzinformation über $F \cap E$).

Die Wahrscheinlichkeit eines mindestens 60*-jahrigen Mannes mindestens* 80 *zu werden ist also mit* 66% *deutlich höher als* 59% *für einen beliebigen Mann.*

Bemerkung 2.11.7 *Die Wahrschienlichkeiten, die an den Kanten eines Wahrscheinlichkeitsbaums stehen, sind nichts anderes als bedingte Wahrscheinlichkeiten, denn wir geben eine Wahrscheinlichkeit für ein Ergebnis an, unter der Voraussetzung, dass wir den Baum schon bis zu dem betrachteten Knoten verfolgt haben.*

Beispiel 2.11.8 *Bei einem Spiel haben wir zwei Töpfe mit Losen. Zunächst werfen wir eine Münze. Bei Kopf ziehen wir aus Topf A bei Zahl aus Topf B. Topf A enthält 5 Lose von denen 3 Gewinne sind und 2 Nieten. Topf B enthält einen Gewinn und eine Niete. Wir haben also*

$$\Omega = \{A, B\} \times \{G, N\}.$$

Die Wahrscheinlichkeiten der Ergebnisse können wir mittels des Baumdiagramms in Abbildung 2.13 darstellen, also

$$m((A,G)) = \tfrac{3}{10} \qquad m((A,N)) = \tfrac{1}{5}$$
$$m((B,G)) = \tfrac{1}{4} \qquad m((B,N)) = \tfrac{1}{4}.$$

Den Wahrscheinlichkeitsbaum können wir in Termen von bedingten Wahrscheinlichkeiten interpretieren: Das Gewinn- und das Verlustereignis ist

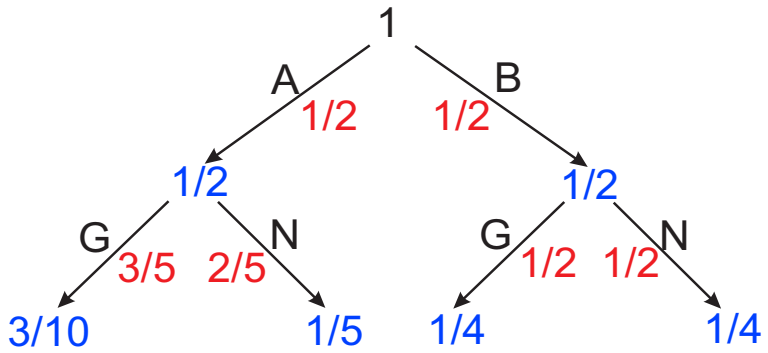


Abbildung 2.13: Wahrscheinlichkeitsbaum für zweistufiges Spiel

$$\mathcal{G} = \{(A,G),(B,G)\} \qquad \mathcal{N} = \{(A,N),(B,N)\}.$$

Weiter haben wir Ereignisse, die anzeigen, aus welchem Topf wir gezogen haben:

$$\mathcal{A} = \{(A,G),(A,N)\} \qquad \mathcal{B} = \{(B,G),(B,N)\}.$$

Am Wahrscheinlichkeitbaum können wir die Wahrscheinlichkeiten

$$P(\mathcal{A}) = \frac{1}{2} \qquad P(\mathcal{B}) = \frac{1}{2}$$

$$P(\mathcal{G} \mid \mathcal{A}) = \frac{3}{5} \qquad P(\mathcal{N} \mid \mathcal{A}) = \frac{2}{5}$$
$$P(\mathcal{G} \mid \mathcal{B}) = \frac{1}{2} \qquad P(\mathcal{N} \mid \mathcal{B}) = \frac{1}{2}$$

ablesen, und dann z.B. berechnen

$$\begin{aligned} m((A,G)) &= P(\mathcal{A} \cap \mathcal{G}) = P(\mathcal{A}) \cdot P(\mathcal{G} \mid \mathcal{A}) \\ &= \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10}. \end{aligned}$$

2.11.2 Bayes-Umkehrformel

In der Praxis hat man oft bedingte Wahrscheinlichkeiten $P(F \mid E)$ gegeben und möchte daraus die sogenannte inverse bedingte Wahrscheinlichkeit $P(E \mid F)$ berechnen.

Beispiel 2.11.9 *In Beispiel 2.11.8 können wir z.B. nach der Wahrscheinlichkeit fragen, dass wir aus Topf A gezogen haben unter der Voraussetzung, dass wir gewonnen haben. Dazu müssen wir*

$$P(\mathcal{A} \mid \mathcal{G})$$

bestimmen.

Satz 2.11.10 *Sei Ω ein diskreter Wahrscheinlichkeitsraum und sei*

$$\Omega = A_1 \cup \ldots \cup A_r$$

eine Partition von Ω in Ereignisse (insbesondere sind die A_i paarweise disjunkt) und sei B ein weiteres Ereignis. Dann gilt

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_1) \cdot P(B \mid A_1) + \ldots + P(A_r) \cdot P(B \mid A_r)}.$$

Beweis. Aus der Definition der bedingten Wahrscheinlichkeit haben wir

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

und genauso

$$P(A_j \cap B) = P(A_j) \cdot P(B \mid A_j)$$

für alle j. Da die A_j eine Partition von Ω bilden, ist

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \ldots + P(A_r \cap B) \\ &= P(A_1) \cdot P(B \mid A_1) + \ldots + P(A_r) \cdot P(B \mid A_r). \end{aligned}$$

■

Bemerkung 2.11.11 *Wir können und den Satz so vorstellen: Das Ereignis B liefert eine Evidenz, dass A_i vorliegt. Wir wissen mit welcher Wahrscheinlichkeit A_i eintritt. Weiter wissen wir, mit welcher Wahrscheinlichkeit $P(B \mid A_i)$ das Ereignis B eintritt unter der Voraussetzung, dass A_i vorliegt. Dann können wir die Wahrscheinlichkeit $P(A_i \mid B)$ von A_i bestimmen unter der Voraussetzung, dass B eintritt.*

Beispiel 2.11.12 *Die Bayes-Formel spielt eine zentrale Rolle in medizinischen Diagnosen: Mittels einer Studie finden wir heraus, welche Teilnehmer bestimmte Erkrankungen bekommen (Ereignisse A_i) und dass zu Beginn der Studie ein bestimmter Test ein positives Ergebnis geliefert hat (Ereignis B). Dadurch kennen wir die Wahrscheinlichkeiten $P(B \mid A_i)$, dass der Test ein positives Ergebnis liefert, unter der Voraussetzung, dass ein Patient eine bestimmte Erkrankung A_i bekommt.*

Jetzt kommt ein Patient zum Arzt und dieser führt Tests durch. Der Arzt (und natürlich auch der Patient) sind an der Wahrscheinlichkeit

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_1) \cdot P(B \mid A_1) + \ldots + P(A_r) \cdot P(B \mid A_r)}$$

interessiert, dass der Patient eine bestimmte Erkrankung bekommt unter der Voraussetzung, dass der Test positiv ist.

Ein Beispiel dazu betrachten wir in Übungsaufgabe 2.30.

Beispiel 2.11.13 *In Beispiel 2.11.8 erhalten wir*

$$\begin{aligned} P(\mathcal{A} \mid \mathcal{G}) &= \frac{P(\mathcal{A} \cap \mathcal{G})}{P(\mathcal{G})} \\ &= \frac{P(\mathcal{A}) \cdot P(\mathcal{G} \mid \mathcal{A})}{P(\mathcal{A}) \cdot P(\mathcal{G} \mid \mathcal{A}) + P(\mathcal{B}) \cdot P(\mathcal{G} \mid \mathcal{B})} \\ &= \frac{\frac{1}{2} \cdot \frac{3}{5}}{\frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{6}{11}. \end{aligned}$$

Natürlich könnten wir die bedingte Wahrscheinlichkeit auch direkt aus den Wahrscheinlichkeiten der einzelnen Ergebnisse ablesen: Mit

$$\begin{aligned} \mathcal{G} &= \{(A,G),(B,G)\} \\ \mathcal{A} &= \{(A,G),(A,N)\} \\ \mathcal{A} \cap \mathcal{G} &= \{(A,G)\} \end{aligned}$$

erhalten wir

$$\begin{aligned} P(\mathcal{G}) &= \frac{3}{10} + \frac{1}{4} = \frac{11}{20} \\ P(\mathcal{A} \cap \mathcal{G}) &= \frac{3}{10} \end{aligned}$$

also ist

$$P(\mathcal{A} \mid \mathcal{G}) = \frac{P(\mathcal{A} \cap \mathcal{G})}{P(\mathcal{G})} = \frac{\frac{3}{10}}{\frac{11}{20}} = \frac{6}{11}.$$

Mit der Bayes-Formel können wir auch

$$P(\mathcal{A} \mid \mathcal{N}) = \frac{P(\mathcal{A}) \cdot P(\mathcal{N} \mid \mathcal{A})}{P(\mathcal{A}) \cdot P(\mathcal{N} \mid \mathcal{A}) + P(\mathcal{B}) \cdot P(\mathcal{N} \mid \mathcal{B})} = \frac{\frac{1}{2} \cdot \frac{2}{5}}{\frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{4}{9}$$

bestimmen und erhalten den kompletten ***umgekehrten Wahrscheinlichkeitsbaum*** *in Abbildung 2.14.*

Eine weitere interessante Anwendung ist die Frage zwischen krank und gesund. Mathematisch ist das dasselbe Problem wie in Beispiel 2.11.12:

Beispiel 2.11.14 *Ein Arzt führt einen Test an einem Patienten durch. Aus einer Studie, weiss man, dass* $\frac{1}{2000}$ *aller Menschen*

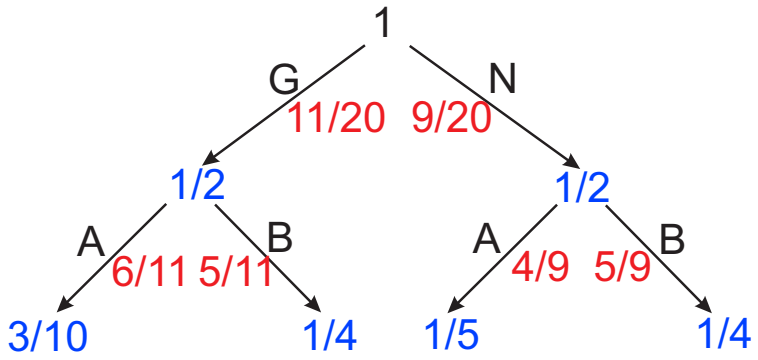


Abbildung 2.14: Umgekehrter Wahrscheinlichkeitsbaum für zweistufiges Spiel.

eine bestimmte Erkrankung haben. Ist ein Mensch erkrankt, dann ist der Test mit Wahrscheinlichkeit $\frac{98}{100}$ *positiv. Ist der Mensch nicht krank, dann ist der Test mit Wahrscheinlichkeit* $\frac{90}{100}$ *negativ. Wir schreiben* k *für krank,* g *für gesund,* 1 *für positiv und* 0 *für negativ. Wir wissen*

$$P(k) = \frac{1}{2000} \Longrightarrow P(g) = \frac{1999}{2000}$$
$$P(1 \mid k) = \frac{98}{100} \Longrightarrow P(0 \mid k) = \frac{2}{100}$$
$$P(0 \mid g) = \frac{90}{100} \Longrightarrow P(1 \mid g) = \frac{10}{100}$$

Fällt der Test nun positiv aus, was ist die Wahrscheinlichkeit, dass der Patient erkrankt ist? Die Bayes-Formel liefert

$$P(k \mid 1) = \frac{P(k) \cdot P(1 \mid k)}{P(k) \cdot P(1 \mid k) + P(g) \cdot P(1 \mid g)}$$
$$= \frac{\frac{1}{2000} \cdot \frac{98}{100}}{\frac{1}{2000} \cdot \frac{98}{100} + \frac{1999}{2000} \cdot \frac{10}{100}} = \frac{49}{10044}$$

Durch einen positiven Test ist also die Wahrscheinlichkeit, krank zu sein, von

$$\frac{1}{2000} = 0.0005$$

auf

$$\frac{49}{10044} \approx 0.0049$$

gestiegen. Die Wahrscheinlichkeit eines false positives ist mit

$$P(g \mid 1) = 1 - \frac{49}{10044} = \frac{9995}{10\,044} \approx 0.9951$$

sehr hoch. Diese Wahrscheinlichkeit können wir natürlich auch direkt mit der Umkehrformel bestimmen:

$$\begin{aligned} P(g \mid 1) &= \frac{P(g) \cdot P(1 \mid g)}{P(g) \cdot P(1 \mid g) + P(k) \cdot P(1 \mid k)} \\ &= \frac{\frac{1999}{2000} \cdot \frac{10}{100}}{\frac{1999}{2000} \cdot \frac{10}{100} + \frac{1}{2000} \cdot \frac{98}{100}} = \frac{9995}{10\,044}. \end{aligned}$$

Unter den positiv getesteten Patienten sind also 99.51% *gesund und nur* 0.49% *krank. Unter* 10000 *positiv getesteten Patienten finden wir also* 49 *Erkrankte während* 9951 *gesunde ohne Zusatznutzen zu weiteren Untersuchungen geschickt werden. Ein solcher Test kann also nur für extrem gefährliche Erkrankungen sinnvoll sein.*

Man muss sich allerdings auch klar machen, dass sich die Wahrscheinlichkeit $P(g \mid 1)$ *eben nur auf die positiv getesteten bezieht. Die Wahrscheinlichkeit negativ getestet zu werden, ist*

$$P(0) = \frac{1}{2000} \cdot \frac{2}{100} + \frac{1999}{2000} \cdot \frac{90}{100} = \frac{22489}{25000} \approx \frac{90}{100} = 0.9.$$

Unter 100000 *Getesteten haben* 10000 *einen weiteren Aufwand und davon sind* 9951 *gesund.*

Wieviele Erkrankungen werden nicht erkannt (false negative)? Dazu berechnen wir

$$\begin{aligned} P(k \mid 0) &= \frac{P(k) \cdot P(0 \mid k)}{P(k) \cdot P(0 \mid k) + P(g) \cdot P(0 \mid g)} \\ &= \frac{\frac{1}{2000} \cdot \frac{2}{100}}{\frac{1}{2000} \cdot \frac{2}{100} + \frac{1999}{2000} \cdot \frac{90}{100}} = \frac{1}{89956} \approx 0.00001. \end{aligned}$$

bei 100000 *Getesteten also etwa* 1 *Person.*

Wir diskutieren noch zwei Beispielprobleme zu bedingten Wahrscheinlichkeiten mit überraschenden Lösungen. Insbesondere hätte man vielleicht auf die Idee kommen können, dass wir in beiden Beispielen dieselbe Wahrscheinlichkeit bekommen.

Beispiel 2.11.15 *Von einer Familie mit 2 Kindern wissen wir, dass wenigstens eines der Kinder ein Junge ist. Was ist die bedingte Wahrscheinlichkeit, dass dann beide Kinder Jungen sind? Wir nehmen an, dass jedes Kind mit Wahrscheinlichkeit $\frac{1}{2}$ ein Junge oder ein Mädchen ist. Wir haben also den Wahrscheinlichkeitsraum*

$$\Omega = \{(J,J),(J,M),(M,J),(M,M)\}$$

wobei alle Ergebnisse Wahrscheinlichkeit $m(\omega) = \frac{1}{4}$ haben. Das Ereignis mindestens ein Junge ist

$$\mathcal{A} = \{(J,J),(J,M),(M,J)\}$$

das Ereignis zwei Jungen ist

$$\mathcal{B} = \{(J,J)\} \subset \mathcal{A}$$

und somit

$$P(\mathcal{B} \mid \mathcal{A}) = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Wir modifizieren dieses Beispiel leicht:

Beispiel 2.11.16 *Ein Vater hat zwei Kinder und läuft mit einem seiner Kinder im Park. Es ist ein Junge. Was ist die Wahrscheinlichkeit, dass das andere Kind auch ein Junge ist?*

Aus dem Wahrscheinlichkeitsbaum in Abbildung 2.15 können wir ablesen, dass das andere Kind mit Wahrscheinlicheit

$$\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{8}+\frac{1}{8}} = \frac{1}{2}$$

ein Junge ist.

Wichtig ist hier natürlich die Annahme, dass der Vater unter seinen beiden Kindern ein Kind zufällig mit gleicher Wahrscheinlichkeit $\frac{1}{2}$ zum Spazierengehen auswählt.

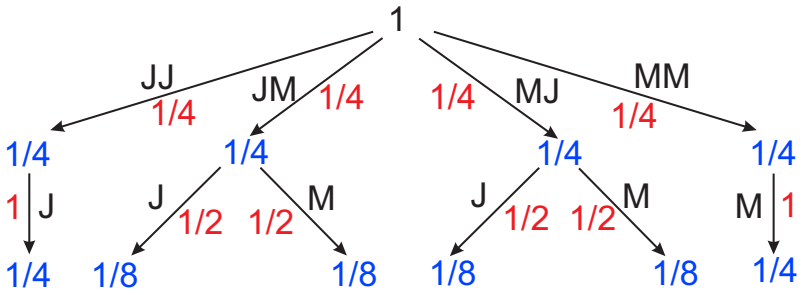


Abbildung 2.15: Kinder des Vater und ob er einen Jungen oder ein Mädchen mit in den Park nimmt.

2.12 Wahrscheinlichkeit einer Mindestabweichung vom Erwartungswert

2.12.1 Effektive Schranke: Markov- und Tschebyscheff-Ungleichung

Als Vorresultat zum Gesetz der großen Zahlen, das eine Beziehung zwischen Mittelwert und Erwartungswert herstellt, zeigen wir die Tschebyscheff-Ungleichung, die eine effektive Schranke für die Wahrscheinlichkeit einer festgelegten minimalen Abweichung vom Erwartungswert. Die Schranke ist gegeben in Abhängigkeit von der Varianz von X.

Notation 2.12.1 *Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$, sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $c \in \mathbb{R}$. Wir schreiben*

$$P(X \geq c) = \sum_{\substack{\omega \in \Omega \\ X(\omega) \geq c}} m(\omega)$$

für die Wahrscheinlichkeit, dass X Werte $\geq c$ annimmt, also für die Wahrscheinlichkeit des Ereignisses

$$\{\omega \in \Omega \mid X(\omega) \geq c\}\,.$$

Satz 2.12.2 (Markov-Ungleichung) *Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to$*

$\mathbb{R}_{\geq 0}$ und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $X \geq 0$. Für alle $\varepsilon > 0$ gilt

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Beweis. Wegen $X(\omega) \geq 0$ haben wir

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega) \cdot m(\omega) \geq \sum_{\substack{\omega \in \Omega \\ X(\omega) \geq \varepsilon}} X(\omega) \cdot m(\omega) \\ &\geq \varepsilon \cdot \sum_{\substack{\omega \in \Omega \\ X(\omega) \geq \varepsilon}} m(\omega) = \varepsilon \cdot P(X \geq \varepsilon) \end{aligned}$$

■

Satz 2.12.3 (Tschebyscheff-Ungleichung) *Sei Ω ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Für alle $\varepsilon > 0$ gilt*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Beweis. Wir wenden die Markov-Ungleichung aus Satz 2.12.2 an auf

$$Y = (X - E(X))^2$$

und erhalten

$$P(|X - E(X)| \geq \varepsilon) = P(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2} = \frac{V(X)}{\varepsilon^2}.$$

■

Beispiel 2.12.4 *Es gilt*

$$P(|X - E(X)| \geq k \cdot \sigma(X)) \leq \frac{V(X)}{k^2 \cdot \sigma(X)^2} = \frac{1}{k^2}$$

die Wahrscheinlichkeit einer Abweichung des mindestens k-fachen der Standardabweichung ist also höchstens $\frac{1}{k^2}$. Eine Abweichung von $10 \cdot \sigma(X)$ hat also eine Wahrscheinlichkeit von weniger als 1%.

Beispiel 2.12.5 *Für den randomisierten Quicksort-Algorithmus auf einer 40-elementigen Menge haben wir die Erwartungswert und Varianz der Laufzeit X in Beispiel 2.8.5 und Bemerkung 2.8.9 berechnet als*

$$E(X) \approx 190$$
$$V(X) \approx 484.$$

Damit erhalten wir die folgenden oberen Schranken für Wahrscheinlichkeiten für Abweichungen $\geq \varepsilon$ vom Erwartungswert

ε	10	20	30	40	50	100
$\min\left\{1, \frac{V(X)}{\varepsilon^2}\right\}$	1	1	0.54	0.30	0.19	0.05

Im Folgenden wollen wir untersuchen, wie scharf die Schranke in der Tschbyscheff-Ungleichung ist.

2.12.2 Qualität der Abschätzung durch die Tschebyscheff-Ungleichung

Beispiel 2.12.6 *Für einen n-maligen Münzwurf (mit 0 für Kopf und 1 für Zahl) sei*

$$\Omega = \{0,1\}^n$$

mit

$$m(\omega) = \frac{1}{2^n}$$

die Zufallsvariable X_i gebe das Ergebnis des i-ten Wurfs an,

$$S_n = X_1 + \ldots + X_n$$

die Häufigkeit von Zahl und

$$Y_n = \frac{S_n}{n}$$

die relative Häufigkeit von Zahl. Damit ist mit Satz 2.6.12

$$E(Y_n) = \frac{1}{n} \cdot (E(X_1) + \ldots + E(X_n)) = \frac{1}{n} \cdot n \cdot \frac{1}{2} = \frac{1}{2}$$

und mit Satz 2.9.9, Satz 2.7.9 und

$$V(X_i) = E(X_i^2) - E(X_i)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

dass

$$V(Y_n) = \frac{1}{n^2} \cdot (V(X_1) + \ldots + V(X_n)) = \frac{1}{n^2} \cdot n \cdot \frac{1}{4} = \frac{1}{4n}.$$

Wir können die Wahrscheinlichkeit, dass X_n bzw. Y_n einen bestimmten Wert annimmt explizit bestimmen als

$$P(Y_n = \frac{a}{n}) = P(X_n = a) = \binom{n}{a} \cdot \frac{1}{2^n}.$$

Somit ist

$$\begin{array}{rcl} P(|Y_{100} - 0.5| < 0.1) & = & P(0.4 < Y_{100} < 0.6) \\ & = & P(40 < S_{100} < 60) \\ & = & \sum_{a=41}^{59} \binom{100}{a} \cdot \frac{1}{2^{100}} \\ & \approx & 0.943 \end{array}$$

und analog

$$P(|Y_{1000} - 0.5| < 0.1) \approx 0.999999999727.$$

Diese Rechnungen können wir in MAPLE *durchführen mit:*

```
P:=sum( binomial(100,a)*1/2^100, a = 41..59):
```

$\frac{747210360626560260812516050 11}{792281625142643375935439503 36}$

```
evalf(P,12);
```

0.943112066359

```
P:=sum( binomial(1000,a)*1/2^1000, a = 401..599):
evalf(P,12);
```

0.999999999727

Die Tschebyscheff-Ungleichung aus Satz 2.12.3 liefert die Abschätzungen

$$P(|Y_n - 0.5| \geq 0.1) \leq \frac{100}{4n}$$

also

$$P(|Y_n - 0.5| < 0.1) \geq 1 - \frac{100}{4n}$$

und damit

$$P(|Y_{100} - 0.5| < 0.1) \geq 0.75$$
$$P(|Y_{1000} - 0.5| < 0.1) \geq 0.975.$$

Wir müssen also n ziemlich groß wählen, um eine gute Abschätzung zu erhalten. Für Anwendungen z.B. im Machine-Learning ist diese Abschätzung nicht gut genug, da man schon für relativ kleine n mit sehr hoher Wahrscheinlichkeit eine Aussage über den Wert einer Zufallsvariable treffen will. Beispielsweise hat man pro Zeiteinheit nur eine begrenzte Anzahl von Messungen über die Position eines Autos auf der Strasse zur Verfügung, will aber damit schon mit geringer Fehlerwahrscheinlichkeit die wahre Position auf eine notwendige Präzision abschätzen. Im Folgenden zeigen wir zunächst das, was man intuitiv sofort glaubt: Je mehr Messwerte wir mitteln, desto geringer wird die Fehlerwahrscheinlichkeit.

2.13 Wahrscheinlichkeit einer Mindestabweichung eines Mittelwerts vom Erwartungswert

2.13.1 Qualitatives Verhalten und eine erste Abschätzung: Gesetz der großen Zahlen

Im Folgenden untersuchen wir die Wahrscheinlichkeit, dass ein Mittelwert von Zufallsvariablen um ein festgelegtes Maß vom Erwartungswert abweicht. Mit dem Gesetz der großen Zahlen erhalten wir auch schon eine erste Schranke, die von der Tschebyscheff-Ungleichung herrührt (so wie in Beispiel 2.12.6). Diese (nicht besonders scharfe) quantitative Schranke erlaubt zumindest schon einmal eine qualitative Aussage über das Verhalten des Mittelwerts zu treffen.

Definition 2.13.1 *Wir sagen, dass Zufallsvariablen $X_1, \dots, X_n : \Omega \to \mathbb{R}$ auf dem diskreten Wahrscheinlichkeitsraum Ω* ***unabhängig und identisch verteilt*** *sind, falls für alle i die Verteilungsfunktionen $a \mapsto P(X_i = a)$ übereinstimmen und*

$$P(X_1 = a_1, \dots, X_n = a_n) = P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n)$$

für alle $a_1, \dots, a_n$.

Beispiel 2.13.2 *Für den n-maligen Münzwurf mit Wahrscheinlichkeitsraum*

$$\Omega = \{0,1\}^n$$

mit

$$m(\omega) = \frac{1}{2^n}$$

für alle ω sind die Zufallsvariablen

$$X_i : \Omega \to \{0,1\},\ (\omega_1, \ldots, \omega_n) \mapsto \omega_i$$

die das Ergebnis des i-ten Wurfs angeben, unabhängig und identisch verteilt.

Bemerkung 2.13.3 *Stochastisch äquivalente Zufallsvariablen haben dieselbe Verteilung, denn*

$$P(X = n) = \sum_{\substack{\omega\in\Omega \\ X(\omega)=n}} m(\omega) = \sum_{\substack{\omega\in\Omega \\ Y(\omega)=n}} m(\omega) = P(Y = n)$$

da das Ereignis aller $\omega \in \Omega$ mit $X(\omega) \neq Y(\omega)$ Wahrscheinlichkeit 0 hat.

Die Umkehrung gilt jedoch nicht: Ist $\Omega = \{a,b,c,d\}$ und die Ergebnisse haben die folgenden Wahrscheinlichkeiten und Werte der Zufallsvariablen X_1 und X_2

ω	a	b	c	d
$m(\omega)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
X_1	1	2	1	3
X_2	2	1	1	4

dann sind X_1 und X_2 identisch verteilt

n	1	2	3	4
$P(X_1 = n)$	$\frac{2}{3}$	$\frac{1}{3}$	0	0
$P(X_2 = n)$	$\frac{2}{3}$	$\frac{1}{3}$	0	0

jedoch nicht stochastisch äquivalent, da

$$P(X_1 \neq X_2) = P(\{a,b,d\}) = \frac{1}{3} + \frac{1}{3} + 0 = \frac{2}{3}.$$

Bemerkung 2.13.4 *Identisch verteilte Zufallsvariablen haben nach Bemerkung 2.6.5 denselben Erwartungswert und auch dieselbe Varianz.*

Satz 2.13.5 (Gesetz der großen Zahlen) *Seien $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen auf dem diskreten Wahrscheinlichkeitsraum Ω, die identisch unabhängig verteilt sind und für die $E(X_i^2)$ existiere. Wir schreiben*

$$\mu = E(X_i)$$
$$\sigma = \sigma(X_i).$$

Dann gilt für alle $\varepsilon > 0$, dass

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2},$$

insbesondere ist für eine Folge von Zufallsvariablen X_i wie oben

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Beweis. Wie in Beispiel 2.12.6 haben wir mit Satz 2.6.12, dass

$$E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mu$$

und mit mit Satz 2.9.9 und 2.7.9, dass

$$V\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

Ebenso wie im Beispiel liefert die Tschebyscheff-Ungleichung aus Satz 2.12.3 angewendet auf

$$X = \frac{X_1 + \ldots + X_n}{n}$$

dann die Behauptung. ■

Bemerkung 2.13.6 *Äquivalent liefert Satz 2.13.5, für alle $\varepsilon > 0$, dass*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| < \varepsilon\right) = 1.$$

Beispiel 2.13.7 *Beschreiben $X_1, \ldots, X_n : \Omega^n \to \{0,1\}$ die n-malige unabhängige Durchführung eines Zufallsexperiments mit boolschem Ergebnis aus $\Omega = \{0,1\}$ und $m(1) = p$, dann ist*

$$\frac{X_1 + \ldots + X_n}{n}$$

die relative Häufigkeit, mit der unter den n Versuchen der 1 Wert aufgetreten ist (siehe Beispiel 2.12.6). Das Gesetz der großen Zahlen liefert

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - p\right| < \varepsilon\right) = 1,$$

besagt also, dass wir für großes n erwarten können, dass die relative Häufigkeit nahe bei der Wahrscheinlichkeit $m(1) = p$ liegt.

Eine unbekannte Wahrscheinlichkeit können wir also in der Praxis durch eine relative Häufigkeit approximieren (wobei man diese aus einer möglichst großen Stichprobe berechnet).

Wie oben diskutiert stellt sich natürlich die Frage, wie groß die Stichprobe sein muss, damit wir mit hinreichender Sicherheit eine vorgegebene Genauigkeit erreichen. Bei der Tschebyscheff-Ungleichung brauchten wir dazu aufgrund der linearen Abhängigkeit des Nenners eine ziemlich große Stichprobe. Wir zeigen nun noch eine Schranke, bei der der Nenner nicht linear, sondern exponentiell wächst.

2.13.2 Effektive Schranke: Die Hoeffding-Ungleichung

Um eine gute Schranke zu erhalten, ist es nicht ausreichend nur die Varianz zu betrachen, man verwendet alle Momente $E(X^i)$ für $i \in \mathbb{N}$. Alle Potenzen von X kann man in der Exponentialfunktion codieren: Wir setzen für $\lambda \in \mathbb{R}$

$$M_X(\lambda) = E(\exp(\lambda \cdot X)).$$

Im Allgemeinen muss dieser Erwartungswert nicht existieren.

Proposition 2.13.8 (Chernoff-Schranke) *Sei X eine Zufallsvariable auf dem diskreten Wahrscheinlichkeitsraum Ω. Dann gilt für alle $\varepsilon > 0$, dass*

$$P(X \geq E(X) + \varepsilon) \leq \min_{\lambda \geq 0} \left(M_{X-E(X)}(\lambda) \cdot \exp(-\lambda \cdot \varepsilon)\right)$$

und

$$P(X \leq E(X) - \varepsilon) \leq \min_{\lambda \geq 0} \left(M_{E(X)-X}(\lambda) \cdot \exp(-\lambda \cdot \varepsilon)\right)$$

Beweis. Für alle $\lambda > 0$ ist wegen der Monotonie der Exponentialfunktion

$$X \geq E(X) + \varepsilon$$

genau dann, wenn

$$\exp(\lambda \cdot X) \geq \exp(\lambda \cdot E(X) + \lambda \cdot \varepsilon)$$

d.h.

$$\exp(\lambda \cdot (X - E(X))) \geq \exp(\lambda \cdot \varepsilon).$$

Mit der Markov-Ungleichung aus Satz 2.12.2 folgt dann

$$\begin{array}{rcl} P(X - E(X) \geq \varepsilon) & = & P(\ \exp(\lambda \cdot (X - E(X))) \geq \exp(\lambda \cdot \varepsilon)\ ) \\ & \leq & \frac{E(\exp(\lambda \cdot (X - E(X))))}{\exp(\lambda \cdot \varepsilon)} \\ & = & M_{X-E(X)}(\lambda) \cdot \exp(-\lambda \cdot \varepsilon). \end{array}$$

Für $\lambda = 0$ erhalten wir

$$M_{X-E(X)}(\lambda) \cdot \exp(-\lambda \cdot \varepsilon) = \frac{E(\exp(0))}{\exp(0)} = E(1) = 1,$$

die obige Ungleichung ist also ebenso erfüllt.

Da die linke Seite der Ungleichung nicht von λ abhängt, können wir auf der rechten Seite das Minimum nehmen und die Ungleichung ist immer noch erfüllt. Das Minimum existiert, da die rechte Seite stetig von λ abhängt und von unten durch 0 beschränkt ist.

Der Beweis der zweiten Ungleichung geht analog. ■

Beispiel 2.13.9 *Wir betrachen eine Zufallsvariable mit*

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Dann ist

$$E(X^k) = \begin{cases} 0 & \text{für } k \text{ ungerade} \\ 1 & \text{für } k \text{ gerade} \end{cases}$$

und somit

$$M_X(\lambda) = E(\exp(\lambda \cdot X)) = \sum_{k=0}^{\infty} \frac{\lambda^k E(X^k)}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}.$$

Da für alle $k \in \mathbb{N}_0$

$$(2k)! \geq 2^k \cdot k!$$

gilt (Beweis mit Induktion), folgt

$$M_X(\lambda) \leq \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k \cdot k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda^2}{2}\right)^k = \exp(\frac{\lambda^2}{2}).$$

Seien nun $X_1, \ldots, X_n$ *wie oben unabhängig und identisch verteilt und*

$$S_n = X_1 + \ldots + X_n.$$

Dann ist

$$E(S_n) = E(X_1) + \ldots + E(X_n) = 0$$

und

$$\begin{aligned} M_{S_n}(\lambda) &= E(\exp(\lambda \cdot (X_1 + \ldots + X_n))) \\ &= E(\exp(\lambda \cdot X_1) \cdot \ldots \cdot \exp(\lambda \cdot X_n)) \\ &= E(\exp(\lambda \cdot X_1)) \cdot \ldots \cdot E(\exp(\lambda \cdot X_n)) \\ &\leq \exp(\frac{\lambda^2}{2})^n = \exp(\frac{n\lambda^2}{2}) \end{aligned}$$

(wobei wir Satz 2.9.7 verwendet haben). Somit gibt die Chernoff-Schranke aus Proposition 2.13.8, dass

$$P(S_n \geq \varepsilon) \leq \min_{\lambda \geq 0} \left(\exp(\frac{n\lambda^2}{2}) \cdot \exp(-\lambda \cdot \varepsilon)\right) = \min_{\lambda \geq 0} \left(\exp(\frac{n\lambda^2}{2} - \lambda \cdot \varepsilon)\right).$$

Das Minimum und damit die beste Schranke erhalten wir für das Minimum von

$$f(\lambda) = \frac{n\lambda^2}{2} - \lambda \cdot \varepsilon$$

also für

$$\lambda = \frac{\varepsilon}{n}.$$

Einsetzen gibt

$$P(S_n \geq \varepsilon) \leq \exp(-\frac{1}{2n}\varepsilon^2).$$

Für den Mittelwert erhalten wir

$$P(\frac{S_n}{n} \geq \varepsilon) \leq \exp(-\frac{n}{2}\varepsilon^2)$$

(indem wir ε *durch* $n \cdot \varepsilon$ *ersetzen). Diese Schranke geht mit* n *exponentiell gegen* 0.

Allgemein gilt (der Beweis geht ähnlich wie in Beispiel 2.13.9):

Satz 2.13.10 (Hoeffding-Ungleichung) *Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen, die Werte im Intervall $[a, b]$ annehmen, dann ist für alle $\varepsilon > 0$*

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)) \geq \varepsilon\right) \leq \exp(-\frac{2n}{(b-a)^2}\varepsilon^2)$$

und

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)) \leq -\varepsilon\right) \leq \exp(-\frac{2n}{(b-a)^2}\varepsilon^2)$$

Dieser Satz wird in der Informatik zentral im Machine-Learning eingesetzt, da er es erlaubt schon mit einer kleinen Stichprobe mit hoher Wahrscheinlichkeit einen Erwartungswert als Mittelwert zu schätzen.

Beispiel 2.13.11 *Für den Mittelwert Y_n der Anzahl von Zahl beim n-maligen Münzwurf in Beispiel 2.12.6 erhalten wir z.B.*

$$P(|Y_n - 0.5| \geq 0.1) \leq 2 \cdot \exp(-2 \cdot n \cdot 0.1^2)$$

und damit

$$P(|Y_{100} - 0.5| \geq 0.1) \leq 0.2706705664$$
$$P(|Y_{1000} - 0.5| \geq 0.1) \leq 0.0000000041$$

oder äquivalent

$$P(|Y_{100} - 0.5| < 0.1) \geq 0.7293294336$$
$$P(|Y_{1000} - 0.5| < 0.1) \geq 0.9999999959$$

für $n = 100$ ist also die Tschebyscheff-Schranke sogar noch leicht besser als die Hoeffding-Schranke, für $n = 1000$ erhalten wir aber mit der Hoeffding-Ungleichung aufgrund der exponentiellen Abnahme der Fehlerwahrscheinlichkeit eine viel bessere Abschätzung:

	$P(\lvert Y_{1000} - 0.5\rvert \geq 0.1)$	$\approx$	0.0000000003
Tschebyscheff	$P(\lvert Y_{1000} - 0.5\rvert \geq 0.1)$	$\leq$	0.025
Hoeffding	$P(\lvert Y_{1000} - 0.5\rvert \geq 0.1)$	$\leq$	$0.0000000041.$

2.14 Übungsaufgaben

Übung 2.1 *Wir wollen die Menge*

$$X = \{56, 64, 58, 61, 75, 86, 17, 62, 8, 50, 87, 99, 67, 10, 74\} \subset \mathbb{Z}$$

mit $n = 15$ *Elementen mit Hilfe des Quicksort-Algorithmus sortieren.*

1) *Finden Sie einen Baum von Pivotelementen, sodass der Algorithmus angewendet auf* X *mindestens* $\binom{n}{2}$ *Vergleiche benötigt.*

2) *Durch welchen Baum von Pivotelementen können Sie weniger als* $n \cdot \ln(n)$ *Vergleiche erreichen?*

Hinweis: Als Laufzeit bezeichnen wir hier die insgesamt notwendige Anzahl von Vergleichen.

Übung 2.2 *Wir wollen, wie in der linearen Algebra definiert, das Produkt von zwei Matrizen* $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{n \times n}$ *berechnen.*

1) *Argumentieren Sie, dass für* $x \in \mathbb{R}^n$ *die Multiplikation Matrix mal Vektor*

$$A \cdot x = \left(\sum_{j=1}^{n} a_{ij} x_j \right)_{i=1..n}$$

entsprechend der angegebenen Formel Laufzeit in $O(n^2)$ *hat.*

2) *Argumentieren Sie, dass die Multiplikation Matrix mal Matrix*

$$A \cdot B = \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right)_{i,k=1..n}$$

entsprechend der angegebenen Formel Laufzeit in $O(n^3)$ *hat.*

3) *Gegeben seien Matrizen $A, B, C \in \mathbb{R}^{n\times n}$. Entwickeln Sie eine Idee für einen Monte-Carlo-Algorithmus mit einseitigem Fehler der*

$$A \cdot B = C$$

mit Laufzeit $O(n^2)$ testet.

Bemerkung: Als Laufzeit betrachten wir die Anzahl der Multiplikationen in $\mathbb{R}$.

Übung 2.3 *Implementieren Sie den randomisierten Quicksort-Algorithmus.*

Erproben Sie die asymptotische Laufzeit Ihrer Implementierung, indem Sie für verschiedene n eine Menge von zufällig erzeugten Zahlen $M = \{x_1, \ldots, x_n\} \subset \mathbb{Z}$ sortieren.

Hinweise:

- *Die Laufzeit messen wir, indem wir in der Implementierung die Anzahl der durchgeführten Vergleiche zählen.*
- *Die* MAPLE*-Funktion* `rand(m)()` *liefert eine Zufallszahl in $\{0, \ldots, m-1\}$.*

Übung 2.4 *Auf der Menge $\Omega = \{1, 2, 3\}$ definieren wir die Funktion*

$$m : \Omega \to \mathbb{R}_{\geq 0},\ 1 \mapsto \frac{1}{4},\ 2 \mapsto \frac{5}{12},\ 3 \mapsto \frac{1}{3}.$$

1) *Zeigen Sie, dass m eine Wahrscheinlichkeitsfunktion auf die Ergebnismenge Ω ist.*

2) *Bestimmen Sie für jede Teilmenge von Ω die Wahrscheinlichkeit.*

Übung 2.5 *Ein Würfel mit 6 Seiten ist so manipuliert, dass die Wahrscheinlichkeit, die Zahl n zu würfeln, proportional zu n ist.*

1) *Bestimmen Sie für jedes $1 \leq n \leq 6$ die Wahrscheinlichkeit, die Zahl n zu würfeln.*

2) *Was ist die Wahrscheinlichkeit eine ungerade Zahl zu würfeln?*

Übung 2.6 *In einem Glücksspiel würfeln wir m-mal mit einem (unmanipulierten) Würfel mit 6 Seiten. Wir gewinnen 6 ,€ falls wenigstens eine 6 auftritt, ansonsten verlieren wir 6 .€*

1) *Bestimmen Sie die Wahrscheinlichkeit, dass keine 6 in dem Spiel gewürfelt wird.*

2) *Wie groß muss m gewählt werden, damit wir im Durchschnitt bei dem Spiel Gewinn machen?*

3) *Was ist der Erwartungswert für den Gewinn?*

Übung 2.7 *Sei $0 < r < 1$. Auf der Menge $\Omega = \mathbb{N}_0$ sei die Funktion*

$$\begin{array}{rccc} m: & \Omega & \to & \mathbb{R}_{\geq 0} \\ & j & \mapsto & (1-r)^j \cdot r \end{array}$$

gegeben. Zeigen Sie, dass m eine Wahrscheinlichkeitsfunktion auf der Ergebnismenge Ω ist.

Übung 2.8 *Die Chance, dass es morgen regnet, ist $1:2$ und die Chance, dass den ganzen Tag die Sonne scheint, ist $3:10$. Was ist die Chance, dass es entweder regnet oder den ganzen Tag die Sonne scheint.*

Übung 2.9 *In einem Glücksspiel würfeln wir m-mal mit einem Würfel mit n Seiten. Wir gewinnen, falls wenigstens eine 1 auftritt, ansonsten verlieren wir. Zeigen Sie:*

1) *Die Wahrscheinlichkeit in dem Spiel keine 1 zu würfeln ist*

$$\left(1 - \frac{1}{n}\right)^m$$

2) *Für $m = n \cdot \ln(2)$ Würfe ist das Spiel im Grenzwert $n \to \infty$ fair. Zeigen Sie dazu, dass*

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{n\cdot\ln(2)} = \frac{1}{2}.$$

Hinweise:

- *Für $x > 0$ und $a \in \mathbb{R}$ ist*

$$x^a := \exp(a \cdot \ln(x)).$$

- *Verwenden Sie die Regel von l'Hospital.*

Übung 2.10 *Sei Ω eine endliche Menge, $m : \Omega \to \mathbb{R}_{\geq 0}$ eine Wahrscheinlichkeitsfunktion und $M_1, \ldots, M_n \subset \Omega$ Ereignisse. Zeigen Sie, dass*

$$P(M_1 \cup \ldots \cup M_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{|T|=k} P(M_T)$$

mit

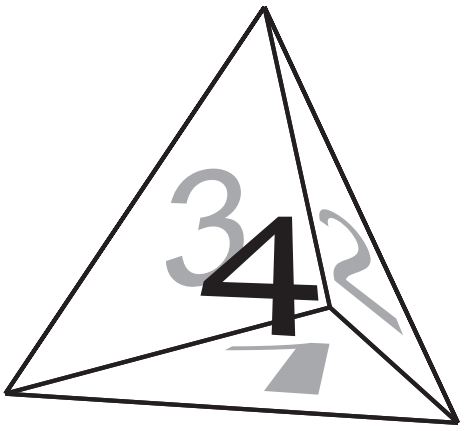
$$M_T = \bigcap_{i \in T} M_i$$

für $T \subset \{1, ..., n\}$.

Hinweis: Folgen Sie dem Beweis der Siebformel.

Übung 2.11 *Wir werfen eine Münze bis zum ersten Mal Kopf kommt. Wenn beim n-ten Wurf zum ersten Mal Kopf kommt, dann gewinnen wir 2^n €. Da es auf der Welt nur etwa 2^{47} € Geld gibt, gilt diese Regel nur für $n \leq 47$ und für $n > 47$ gewinnen wir ebenfalls 2^{47} €. Existiert der Erwartungswert für den Gewinn und, falls ja, welchen Wert hat er.*

Übung 2.12 *Wir spielen mit einem Tetraeder, dessen Seiten mit $1, \ldots, 4$ nummeriert sind. Wir werfen den Tetraeder, bis die Summe der geworfenen Zahlen ≥ 3 ist. Die Zufallsvariable X beschreibe die Anzahl der Würfe.*



1) *Bestimmen Sie den Wahrscheinlichkeitsbaum für dieses Zufallsexperiment.*

2) *Bestimmen Sie die Verteilung von X und den Erwartungswert $E(X)$.*

Übung 2.13 *Beim Spiel Seven Eleven wirft der Spieler zwei (hoffentlich unmanipulierte) Würfel.*

- *Ist die Augensumme $2, 3$ oder 12 verliert er.*
- *Ist die Augensumme 7 oder 11 gewinnt der Spieler.*
- *Ist die Augensumme $s \neq 7, 11$ (und der Spieler hat nicht verloren), dann würfelt der Spieler weiter bis entweder Augensumme s oder 7 auftritt. Im ersten Fall gewinnt er, im zweiten Fall verliert er.*

Gewinnt der Spieler, bekommt er $1 €$, anderenfalls verliert er $1 €$.

1) *Spielen Sie $N = 10$ Durchläufe des Spiels und berechnen Sie Ihren mittleren Gewinn.*

2) *Schreiben Sie ein Programm, das das Spiel implementiert. Bestimmen Sie für $N = 1000$ Durchläufe des Spiels Ihren mittleren Gewinn.*

3) *Erstellen Sie einen Wahrscheinlichkeitsbaum, der das Spiel beschreibt.*

4) *Was ist der Erwartungswert für den Gewinn?*

Übung 2.14 *Für $n \geq 1$ beschreibe die Zufallsvariable*

$$X : S_n \to \mathbb{N}_0$$

die Anzahl der Fixpunkte einer zufällig gewählten Permutation $\sigma \in S_n$, also die Anzahl der $i \in \{1, \ldots, n\}$ mit

$$\sigma(i) = i.$$

1) *Geben Sie eine Wahrscheinlichkeitsfunktion*

$$m : S_n \to \mathbb{R}_{\geq 0}$$

für die zufällige Wahl einer Permutation an, wobei wir annehmen, dass alle Permutationen mit gleicher Wahrscheinlichkeit gewählt werden.

2) *Bestimmen Sie für alle $\sigma \in S_3$ die Anzahl der Fixpunkte.*

3) *Berechnen Sie für $n = 3$ die Verteilung von X und die erwartete Anzahl $E(X)$ von Fixpunkten einer zufällig gewählten Permutation.*

4) *Sei nun $n = 10$. Schreiben Sie eine Funktion, die zufällig eine Permutation in S_n auswählt, und eine Funktion, die die Anzahl der Fixpunkte einer Permutation zählt. Bestimmen Sie für $N = 1000$ Durchläufe ihres Programms den Mittelwert der Anzahl der Fixpunkte. Was ist Ihre Vermutung für $E(X)$?*

Übung 2.15 *Für die zufällige Wahl einer Permutation in S_n definieren wir die Zufallsvariable*

$$\begin{array}{rccl} X_i : & S_n & \to & \{0,1\} \\ & \sigma & \mapsto & \begin{cases} 1 & \text{falls } \sigma(i) = i \\ 0 & \text{sonst} \end{cases} \end{array}$$

1) *Welche Größe wird durch die Zufallsvariable*

$$X = X_1 + \ldots + X_n$$

beschrieben?

2) *Bestimmen Sie den Erwartungswert von X.*

3) *Welche Schlußfolgerung können Sie über das Mischen von Spielkarten ziehen?*

Übung 2.16 *Sei $\Omega = \mathbb{N}$, $X(n) = n$ und $m(n) = \frac{a}{n^3}$ mit der endlichen Konstanten $a := \sum_{n=1}^{\infty} \frac{1}{n^3}$. Zeigen Sie, dass $E(X)$ existiert, $V(X)$ jedoch nicht.*

Übung 2.17 *1) Zeigen Sie für alle $n \in \mathbb{N}$, dass*

$$\sum_{i=1}^{n} i^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{6}$$

2) Die Zufallsvariable X beschreibe die zufällige Wahl einer Zahl aus der Menge $\{1, \ldots, n\}$. Zeigen Sie, dass für den Erwartungswert und die Varianz von X gilt

$$E(X) = \frac{n+1}{2}$$

und

$$V(X) = \frac{(n-1)(n+1)}{12}$$

Übung 2.18 *1) Wir spielen mit einem Würfel mit 6 Seiten. Auf die geworfene Augenzahl addieren wir 1 und multiplizieren das Ergebnis mit 2. Berechnen Sie Erwartungswert und Varianz des Ergebnisses.*

2) Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die die Varianz existiert und $c \in \mathbb{R}$. Zeigen Sie, dass dann

$$V(c \cdot X) = c^2 \cdot V(X)$$

und

$$V(X + c) = V(X).$$

Übung 2.19 *Sei Ω ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion $m : \Omega \to \mathbb{R}_{\geq 0}$ und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die*

$$E(X^{k+1}) = \sum_{\omega \in \Omega} X(\omega)^{k+1} \cdot m(\omega)$$

absolut konvergiert. Zeigen Sie, dass dann auch $E(X^k)$ absolut konvergiert.

Hinweis: Spalten Sie die Summe

$$\sum_{\omega \in \Omega} |X(\omega)|^k \cdot m(\omega)$$

in die Summanden mit $|X(\omega)| \leq 1$ und $|X(\omega)| > 1$ auf.

Übung 2.20 *Sei X eine Zufallsvariable deren Varianz existiert und sei*

$$Y = \frac{X - E(X)}{\sigma(X)}$$

Zeigen Sie, dass

$$E(Y) = 0$$

und

$$V(Y) = 1.$$

Übung 2.21 *Sind $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen für die jeweils die Varianz existiert, dann gilt*

$$V(X_1 \cdot X_2) = V(X_1) \cdot E(X_2) + E(X_1) \cdot V(X_2) + V(X_1) \cdot V(X_2).$$

Übung 2.22 *Bei einer Internetrecherche finden wir: Unter den Menschen, die an Lungenkrebs erkranken sind 16-mal mehr Raucher als Nichtraucher, 30% der Deutschen rauchen, 4% aller Deutschen erkranken in ihrem Leben an Lungenkrebs. Für eine Person ω sei $X_1(\omega) = 1$ falls die Person raucht und $X_1(\omega) = 0$ falls nicht, und $X_2(\omega) = 1$ falls die Person an Lungenkrebs erkrankt und $X_2(\omega) = 0$ falls nicht. Bestimmen Sie die Wahrscheinlichkeitsverteilung*

$$P(X_1 = n_1, X_2 = n_2)$$

für $(n_1, n_2) \in \{0, 1\}^2$.

Übung 2.23 *Sei Ω ein diskreter Wahrscheinlichkeitsraum. Für Zufallsvariablen $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ gilt*

$$\mathrm{Cov}(X_1, X_2) = E(X_1 \cdot X_2) - E(X_1) \cdot E(X_2).$$

Übung 2.24 *Wir würfeln zweimal mit einem 6-seitigen Würfel. Die Zufallsvariablen*

$$X_1 : \Omega \to \{1, \ldots, 6\},\ (a, b) \mapsto a$$
$$X_2 : \Omega \to \{1, \ldots, 6\},\ (a, b) \mapsto b$$

auf $\Omega = \{1, \ldots, 6\}^2$ geben das Ergebnis des ersten Wurfs bzw. zweiten Wurfs an.

1) *Bestimmen Sie die Wahrscheinlichkeitsverteilungen von* $X_1 + X_2$ *und* $X_1 - X_2$.

2) *Bestimmen Sie die Wahrscheinlichkeitsverteilung von* $(X_1 + X_2) \cdot (X_1 - X_2)$ *und daraus die Covarianz* $\mathrm{Cov}(X_1 + X_2, X_1 - X_2)$.

Übung 2.25 *Wir werfen drei Münzen, also* $\Omega = \{0,1\}^3$, *wobei* 0 *für Kopf und* 1 *für Zahl steht.*

- *Die Zufallsvariablen* X_1 *sei* 1 *falls die Anzahl von Kopf gerade ist und* 0 *sonst.*
- *Die Zufallsvariable* X_2 *sei* 1 *falls die ersten beiden Würfe dasselbe Ergebnis haben und* 0 *sonst.*
- *Die Zufallsvariable* X_3 *gebe die Anzahl der Würfe von Zahl an.*

Bestimmen Sie

1) *die Wahrscheinlichkeitsverteilungen von* $X_1 \cdot X_3$ *und* $X_2 \cdot X_3$.

2) *die Covarianzen* $\mathrm{Cov}(X_1, X_3)$ *und* $\mathrm{Cov}(X_2, X_3)$.

3) *die Korrelationen* $\mathrm{Corr}(X_1, X_3)$ *und* $\mathrm{Corr}(X_2, X_3)$.

Übung 2.26 *Sei* $\Omega = \{\omega_1, \ldots, \omega_r\}$ *mit Wahrscheinlichkeitsfunktion* m *gegeben durch* $m(\omega_i) := m_i$ *mit Zahlen* $m_i \geq 0$, *die* $\sum_{i=1}^{r} m_i = 1$ *erfüllen. Sei* X *ein Zufallsvariable auf* Ω *gegeben durch eine Prozedur, die* ω_i *den Wert* $X(\omega_i)$ *zuordnet.*

1) *Schreiben Sie ein Programm, das die Verteilung und den Erwartungswert von* X *berechnet.*

2) *Erweitern Sie Ihr Programm so, dass Sie auch die Varianz und für zwei Zufalls Variablen die Covarianz und Korrelation berechnen können.*

3) *Erproben Sie Ihr Programm an dem zweimaligen Münzwurf*

$$\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$$

mit $m(\omega) = \frac{1}{4}$, *und den Zufallsvariablen* $X(a,b) = (a+b) \cdot (a-b)$, $X(a,b) = a + b$ *und* $X(a,b) = a - b$.

4) Überprüfen Sie Ihre Ergebnisse aus Aufgabe 2.24 und 2.25 mit Hilfe Ihres Programms.

Hinweis: Bestimmen Sie zunächst alle Werte, die X auf Ω annehmen kann, und dann deren Wahrscheinlichkeiten.

Übung 2.27 *Sei Ω ein diskreter Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit positiver Varianz und $a, b \in \mathbb{R}$ mit $a \neq 0$. Zeigen Sie, dass*

$$\mathrm{Corr}(X,\ a \cdot X + b) = \begin{cases} 1 & \text{falls } a > 0 \\ -1 & \text{falls } a < 0 \end{cases}$$

Übung 2.28 *Seien A und B Ereignisse mit*

$$P(A \mid B) = P(A).$$

Zeigen Sie, dass

$$P(B \mid A) = P(B).$$

Übung 2.29 *Wir spielen mit einem Würfel und werfen diesen zwei Mal. Wir gewinnen, falls die die Augensumme kleiner als 7 ist. Was ist die Wahrscheinlichkeit noch zu gewinnen, unter der Voraussetzung, dass*

1) der erste Wurf eine 1 geliefert hat,

2) der erste Wurf eine 3 geliefert hat,

3) der erste Wurf eine 6 geliefert hat,

4) der erste Wurf eine Zahl kleiner als 5 geliefert hat.

Übung 2.30 *In einer medizinischen Studie haben die Teilnehmer jeweils eine der Erkrankungen A, B, C, D. Wir führen zwei Tests T_1, T_2 durch die das Ergebnis 0 (negativ) oder 1 (positiv) liefern können. Unter 1000 Teilnehmern der Studie haben die beiden Tests die Ergebnisse $00, 01, 10, 11$ mit den in der folgenden Tabelle angegebenen Häufigkeiten geliefert:*

	A	B	C	D
00	221	3	51	50
10	31	5	20	331
01	60	2	111	10
11	21	1	40	43

1) *Bestimmen Sie die relativen Wahrscheinlichkeiten der Erkrankungen unter der Voraussetzung, dass die Tests ein bestimmtes Ergebnis liefern.*

2) *Implementieren Sie die Berechnung dieser relativen Wahrscheinlichkeiten aus einem gegebenen Datensatz wie oben.*

Übung 2.31 *In einer populären Zeitschrift wurde das folgende Problem gestellt und hat zu großen Diskussionen geführt: Bei einem Fernsehquiz gibt es 3 Türen, hinter zweien befindet sich eine Ziege, hinter der anderen ein Auto. Der Spieler kann eine Tür auswählen. Der Moderator öffnet dann eine der anderen beiden Türen, hinter der sich eine Ziege befindet. Der Moderator gibt dem Spieler nun die Möglichkeit seine Wahl unter den beiden noch geschlossenen Türen zu ändern. Der Spieler bekommt dann den Preis hinter der Tür seiner Wahl. Ist es von Vorteil, von Nachteil, oder egal für den Spieler, die Wahl seiner Tür zu ändern?*

Hinweis: Zeichnen Sie den Wahrscheinlichkeitsbaum für die Abfolge: Positionierung des Autos, Wahl der Tür durch den Spieler, Wahl der Tür durch den Moderator, und lesen Sie daraus die Gewinnwahrscheinlichkeiten für die zwei Spielstrategien (Wechseln oder nicht Wechseln) ab.

Übung 2.32 *Schreiben Sie ein Programm, das das Spiel aus Aufgabe 2.31 simuliert.*

Übung 2.33 *Ein Freund fordert sie auf, mit einer Münze zu spielen, die er zufällig in der Tasche hat. Er wettet auf Kopf, Sie auf Zahl. Falls Zahl kommt, erhalten Sie* 1 €*, anderenfalls verlieren Sie* 1 €*. Nach* 100 *Würfen haben Sie* 30 € *verloren und den Verdacht, dass die Münze manipuliert ist. Bevor Sie Ihren Freund zur Rede stellen, wollen Sie sich Ihrer Sache sicher sein.*

1) *Wie oft haben Sie gewonnen und wie oft verloren?*

2) *Schätzen Sie die Wahrscheinlichkeit,* 30 € *oder mehr zu verlieren, mit der Tschebyscheff-Ungleichung ab.*

3) *Schätzen Sie die Wahrscheinlichkeit auch mit der Hoeffding-Ungleichung ab.*

4) Bestimmen Sie die Wahrscheinlichkeit, dass Sie bei den 100 *Würfen genau* 30 € *verlieren, und die Wahrscheinlichkeit, dass Sie* 30 € *oder mehr verlieren.*

Hinweis: Für die Rechnung können Sie die MAPLE*-Befehle* `sum` *und* `binomial` *verwenden.*

Übung 2.34 *1) Schreiben Sie eine Funktion, die zufällig* 0 *oder* 1 *mit Wahrscheinlichkeit von je* $\frac{1}{2}$ *zurückgibt. Wählen Sie mit Ihrer Funktion eine Stichprobe von* 250 *Zahlen und bestimmen Sie den Mittelwert.*

2) Wir führen das Experiment aus (1) nun 5000*-mal durch. Wie oft haben Sie eine Abweichung des Mittelwerts vom Erwartungswert von mindestens* 0.1 *bekommen?*

3) Vergleichen Sie die Häufigkeit aus (2) mit der Schranke aus der Hoeffding-Ungleichung.

Hinweis: Die MAPLE*-Funktion* `rand(m)()` *liefert eine Zufallszahl in* $\{0, \ldots, m-1\}$.

3

Wahrscheinlichkeitsdichten

3.1 Übersicht

Nicht jedes Zufallsexperiment hat ein Ergebnis in einem abzählbaren Wahrscheinlichkeitsraum. Beispielweise könnten wir zufällig einen Punkt in dem Intervall

$$[0,1] \subset \mathbb{R}$$

wobei jeder Punkt mit gleicher Wahrscheinlichkeit auftreten soll.

Beispiel 3.1.1 *In der Praxis können wir uns z.B. vorstellen, dass eine drehbare Scheibe, an deren Rand ein Punkt markiert ist, und ein feststehender Zeiger auf den Rand der Scheibe zeigt (Abbildung 3.1). Wir versetzen die Scheibe in Rotation. Kommt die Scheibe durch Reibung zum Stillstand, messen wir z.B. im Uhrzeigersinn den Abstand von der Markierung zu dem Zeiger. Hat die Scheibe Umfang* 1*, dann liefert dieses Experiment eine zufällige Zahl in* $[0,1]$*, wobei wir* 0 *und* 1 *identifizieren, d.h. eine Zahl in* $[0,1[$*.*

Bemerkung 3.1.2 *Mit dem* MAPLE*-Paket* `Statistics` *können wir n zufällige Zahlen in dem Intervall* $[0,1]$ *erzeugen, wobei alle Elemente von* $[0,1]$ *gleich wahrscheinlich sind. Im Folgenden erzeugen wir zwei Zufallszahlen:*

```
with(Statistics);
X:=RandomVariable(Uniform(0,1)):
Sample(X,3);
```

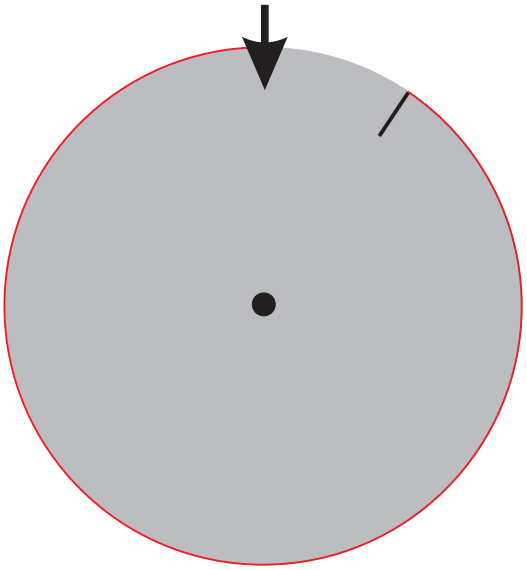


Abbildung 3.1: Rotierende Scheibe mit feststehendem Zeiger und Markierung.

0.402351616935948, 0.658856481072931, 0.126986816293506

Mit dem älteren und weitaus weniger leistungsfähigen MAPLE-*Paket* `stats` *geht dies auch mit einem einzigen Befehl:*

```
stats[random, uniform](n);
```

Natürlich kann eine derartige Funktion nur eine endliche Fließkommadarstellung zurückgeben und damit eine Zahl in $\mathbb{Q}$*. Mit der ganzzahligen Zufallsfunktion* `rand` *können wir ebenso eine Funktion schreiben, die eine solche Fließkommadarstellung erzeugt:*

```
N:=10^15:
rand(N+1)()/N;
```

$$\frac{69747391251049}{100000000000000}$$

Tatsächlich wählen wir also (in obigem Beispiel für $N = 10^{15}$*) aus der endlichen Menge*

$$M = \left\{ \frac{0}{N}, \frac{1}{N} \ldots, \frac{N}{N} \right\},$$

wobei alle Elemente mit gleicher Wahrscheinlichkeit

$$\frac{1}{N+1}$$

gezogen werden. Insbesondere muss also der Mittelwert

$$\sum_{j=0}^{N} \frac{j}{N} \cdot \frac{1}{N+1} = \frac{1}{N \cdot (N+1)} \cdot \frac{(N+1) \cdot N}{2} = \frac{1}{2}$$

sein. Wir erproben das an einer Stichprobe von $n = 1000$ *Werten (wobei wir mit* 16 *Fließkommastellen rechnen):*

```
N:=10^15:
n:=1000:
s:=0:
for j from 1 to n do
    s:=s+rand(N+1)();
od:
evalf(s/N/n,15);
```

0.501869387047372

Entsprechend mit dem `Statistics` *Paket:*

```
with(Statistics);
X:=RandomVariable(Uniform(0,1)):
L:=Sample(X,1000);
ExpectedValue(L);
```

0.502086525728241

3.2 Von der Summation zur Integration und zurück

Wir wollen, dass jeder Punkt im Intervall

$$[0,1] \subset \mathbb{R}$$

mit gleicher Wahrscheinlichkeit auftritt

$$m(n) = c$$

Nehmen wir an, dass $c > 0$ ist. Wir können leicht eine injektive Abbildung

$$g : \mathbb{N}_0 \to [0,1],\ x \mapsto \frac{1}{1+x}$$

angeben, d.h. $\mathbb{N}_0$ als Teilmenge von $[0,1]$ auffassen. Wahrscheinlichkeiten sollten aber additiv sein, d.h.

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) \cup P(A_2) \cup \ldots$$

für paarweise disjunkte Mengen $A_1, A_2, \ldots$. Damit folgt aber für die Wahrscheinlichkeit des Ereignisses $[0,1]$, dass

$$P([0,1]) \geq \sum_{n=0}^{\infty} m(g(n)) = \sum_{n=0}^{\infty} c = \infty$$

nicht konvergent, also schon gar nicht ≤ 1 ist. Wir erhalten also:

Bemerkung 3.2.1 *Treten alle*

$$n \in [0,1] \subset \mathbb{R}$$

in einem Zufallsexperiment mit gleicher Wahrscheinlichkeit

$$m(n) = c \in [0,1]$$

auf, dann muss

$$m(n) = 0$$

sein für alle $n \in [0,1]$.

Bemerkung 3.2.2 *Nach Bemerkung 3.2.1 muss gelten*

$$P([a,b]) = P(]a,b]) = P([a,b[) = P(]a,b[).$$

Die Beobachtung in 3.2.1 führt aber wiederum zu einem Problem mit unserer bisherigen Definition von Wahrscheinlichkeit von Ereignissen, denn

$$P([0,1]) = 1$$

erhalten wir sicher nicht als Summe von Nullen. Dieser scheinbare Widerspruch löst sich aber auf, da wir ja bisher nur abzählbare Summen, also Reihen betrachtet haben. Das heißt bisher konnte man eine `while`-Schleife schreiben, die über alle Elemente des Wahrscheinlichkeitsraums iteriert. Das Intervall $[0,1]$ ist aber nicht abzählbar, wir können also nicht mit einer Reihe über seine Elemente iterieren. Wie kann man dieses Problem lösen? Die Lösung kommt aus der Idee der Integralrechnung:

Bemerkung 3.2.3 *Wie schon in der Simulation der Gleichverteilung in Bemerkung 3.1.2 diskretisieren wir die Verteilung. Dazu unterteilen wir das Intervall* $[0,1]$ *in* n *gleich große Intervalle*

$$I_j := \left[\frac{j}{n}, \frac{j+1}{n}\right]$$

für $j = 0, \ldots, n-1$. *Dann sollte gelten*

$$P(I_j) = \frac{1}{n}.$$

Somit ist

$$P(\left[0, \frac{b}{n}\right]) = \sum_{i=1}^{b} \frac{1}{n} = \frac{b}{n}.$$

Nehmen wir nun $n = 10^k$. *Ist nun* $x \in \mathbb{R}$

$$x_k = \left\lfloor x \cdot 10^k \right\rfloor,$$

dann gibt die Dezimalbruchentwicklung

$$\frac{x_k}{10^k}$$

von a *mit* k *Stellen eine Folge mit*

$$\lim_{k\to\infty} \frac{x_k}{10^k} = x$$

d.h. die Cauchyfolge (x_k) *ist ein Repräsentant der reellen Zahl* $x = [(x_k)]$. *Dann sollte gelten*

$$P([0,x]) = \lim_{k\to\infty} P([0,x_k]) = \lim_{k\to\infty} \frac{x_k}{10^k} = x,$$

wenn wir annehmen, dass sich die Wahrscheinlichkeit $P([0,x])$ *stetig von* x *abhängt. Aus der Wahrscheinlichkeit* $P([0,x])$ *erhalten wir die Änderung der Wahrscheinlichkeit in Abhängigkeit von* x *durch die Ableitung*

$$P([0,x])' = (x)' = 1.$$

Nach dem Hauptsatz der Differential und Integralrechung ist also $P([0,x]) = x$ *eine Stammfunktion der Funktion* $f(x) := 1$, *wir erhalten also*

$$P([0,x]) = [x]_0^1 = \int_0^x 1 \, dx.$$

Die Funktion $f(x)$ bezeichnen wir als die Wahrscheinlichkeitsdichte. Durch Festlegen der Wahrscheinlichkeitsdichte können wir jeder Teilmenge $A \subset [0,1]$, über die wir integrieren können, eine Wahrscheinlichkeit zuordnen. Beispielsweise ist

$$P([0,1/4] \cup [3/4,1]) = \int_0^{1/4} 1dx + \int_{3/4}^{1} 1dx = \frac{1}{2}.$$

Für die Interpretation des Integrals und damit der Wahrscheinlichkeit als Fläche unter der Wahrscheinlichkeitsdichte, siehe Abbildung 3.2.

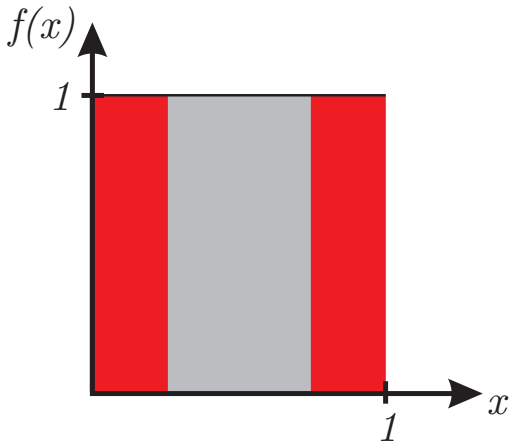


Abbildung 3.2: Wahrscheinlichkeit als Fläche unter der Wahrscheinlichkeitsdichte.

Bemerkung 3.2.4 *Die Berechnung des Integrals $\int_0^x 1\ dx$ als Riemannintegral (Ober- und Untersumme sind gleich für konstante Funktion 1) ist nichts anderes als die obige Berechnung von $P([0,x])$ als Grenzwert:*

$$\int_0^x 1\ dx = \lim_{k\to\infty} \sum_{i=1}^{x_k} \frac{1}{10^k} = \lim_{k\to\infty} \frac{x_k}{10^k} = \lim_{k\to\infty} P([0,x_k]) = P([0,x]).$$

Definition 3.2.5 *Sei $\Omega \subset \mathbb{R}^n$. Eine **Wahrscheinlichkeitsdichte** ist eine über Ω integrierbare Funktion $f : \Omega \to \mathbb{R}_{\geq 0}$ mit*

$$\int_\Omega f(\omega)d\omega = 1.$$

Für eine Teilmenge $A \subset \Omega$ über die f integrierbar ist, definieren wir

$$P(A) = \int_A f(\omega) d\omega.$$

Die Menge Ω zusammen mit der Wahrscheinlichkeitsdichte f und der Menge Σ aller erlaubten Teilmengen $A \subset \Omega$ bezeichnen wir als ***kontinuierlichen Wahrscheinlichkeitsraum****.*

Die Wahrscheinlichkeitsdichte übernimmt also im kontinuierlichen Fall die Rolle der Wahrscheinlichkeitsfunktion. Wahrscheinlichkeiten von Ereignissen erhalten wir statt durch Summation durch Integration.

diskret	kontinuierlich
$m : \Omega \to \mathbb{R}_{\geq 0}$	$f : \Omega \to \mathbb{R}_{\geq 0}$
$P(A) = \sum_{\omega \in A} m(\omega)$	$P(A) = \int_A f(\omega) d\omega$

Beispiel 3.2.6 *Für $\Omega = [0,1]$ und $A = [a,b] \subset \Omega$ ist*

$$P([a,b]) = \int_a^b f(\omega) d\omega.$$

Für $A = [0,1] \cap \mathbb{Q} \subset \Omega$ und Wahrscheinlichkeitsdichte $f(x) = 1$ können wir dem Riemannintegral über 1 also der Wahrscheinlichkeit $P(A)$ keinen sinnvollen Wert zuweisen: Zunächst lässt sich das Integral von 1 über A ausdrücken, indem wir die Funktion

$$g : [0,1] \to [0,1]$$
$$g(x) = \begin{cases} 0 & \text{für } x \in A \\ 1 & \text{für } x \notin A \end{cases}$$

über ganz $[0,1]$ integrieren, d.h.

$$\int_A 1 d\omega = \int_0^1 g(x) dx.$$

Die Obersumme von g ist 1, da jedes Intervall eine rationale Zahl enthält, während die Untersumme von g gleich 0 ist. Somit ist g nicht Riemannintegrierbar.

Bemerkung 3.2.7 *Generell stellt sich die Frage nach dem Begriff der Integrierbarkeit in Definition 3.2.5. Die Definition des Riemannintegrals können wir direkt von dem univariaten Fall übertragen: Wir unterteilen den Integrationsbereich in eine Vereinigung von Produkten von Intervallen*

$$[a_1, b_1] \times \ldots \times [a_n, b_n]$$

nähern das Volumen unter dem Funktionsgraphen durch

$$(b_1 - a_1) \cdot \ldots \cdot (b_n - a_n) \cdot f(x)$$

mit

$$x \in [a_1, b_1] \times \ldots \times [a_n, b_n]$$

an und machen Unterteilung immer feiner.[1] *Der* ***Satz von Fubini*** *besagt, dass sich dieses Integral auch durch iterierte univariate Integration berechnen läßt, z.B. ist*

$$\begin{aligned}\int_{[0,1]\times[0,1]} x_1 \cdot x_2 \; dx &= \int_0^1 \int_0^1 x_1 \cdot x_2 \; dx_1 dx_2 \\ &= \int_0^1 x_2 \int_0^1 x_1 \; dx_1 \; dx_2 \\ &= \int_0^1 \frac{x_2}{2} \; dx_2 = \frac{1}{4}.\end{aligned}$$

Wir bemerken noch: Für diskrete Wahrscheinlichkeitsräume existiert keine Wahrscheinlichkeitsdichte, denn Integrale über Punkte

$$\int_a^a f(x)dx = 0$$

verschwinden, damit aber auch über jede diskrete Menge, denn

$$\sum_{n=1}^{\infty} 0 = 0.$$

Wie man diskrete und kontinuierliche Wahrscheinlichkeitsräume in einem gemeinsame Konzept zusammenführt werden wir in einem kurzen Ausblick auf die axiomatische Wahrscheinlichkeitstheorie in Abschnitt 5.1 sehen.

[1] Für manche nicht Riemann-integrierbare Funktionen kann man immer noch das sogenannte Lebesgue-Integral definieren: Hier unterteilt man nicht den Definitionsbereich der Funktion, sondern den Wertebereich. Das Lebesgue-Integral für die Funktion g aus Beispiel 3.2.6 existiert und ist gleich 0.

Beispiel 3.2.8 *Die **Gleichverteilung** auf dem Intervall* $[c,d] \subset \mathbb{R}$ *ist gegeben durch die Wahrscheinlichkeitsdichte*

$$f(x) = \frac{1}{d-c}.$$

Analog kann man für kartesische Produkte von Intervallen im $\mathbb{R}^n$ *eine Gleichverteilung definieren.*

Auf $[0, \frac{1}{2}]$ *ist also die Gleichverteilung gegeben durch*

$$f(x) = 2$$

für alle x*. Wir bemerken insbesondere, dass Wahrscheinlichkeitsdichten (im Gegensatz zu Wahrscheinlichkeitsfunktionen) auch Werte* > 1 *annehmen können.*

Zum Abschluss des Abschnitts verallgemeinern wir noch Bemerkung 3.2.3 zur Diskretisierung auf beliebige stetige Wahrscheinlichkeitsdichten:

Bemerkung 3.2.9 *Für eine stetige Dichte*

$$f : [a,b] \to \mathbb{R}_{\geq 0}$$

können wir analog zu Bemerkung 3.2.3 diskretisieren: Für $n \in \mathbb{N}$ *sei*

$$t_i = a + (b-a) \cdot \frac{i}{n}$$

mit $i = 0, \ldots, n$ *und*

$$f_i = \int_{t_{i-1}}^{t_i} f(x)dx.$$

mit $i = 1, \ldots, n$*. Dann erhalten wir den diskreten Wahrscheinlichkeitsraum*

$$\Omega = \{1, \ldots, n\}$$

mit der Wahrscheinlichkeitsfunktion

$$m(i) = f_i.$$

Nach dem Mittelwertsatz der Differentialrechnung gibt es $\xi_i \in [t_{i-1}, t_i]$ *mit*

$$f(\xi_i) \cdot \frac{b-a}{n} = f_i.$$

Für eine erste praktische Näherung kann man

$$f_i = f(\frac{t_{i-1} + t_i}{2}) \cdot \frac{b-a}{n}$$

nehmen. Eine andere Näherung, die oft in numerischen Integrationsverfahren verwendet wird, ist

$$f_i = \frac{f(t_{i-1}) + f(t_i)}{2} \cdot \frac{b-a}{n}.$$

Für $n \to \infty$ konvergieren diese Diskretisierungen gegen die kontinuierliche durch die Dichte f gegebene Verteilung in dem Sinne, dass

$$P([a,x]) = \lim_{n\to\infty} P(\{i \in \Omega \mid t_i \leq x\})$$

(wobei Ω und die t_i von n abhängen).

Man beachte, dass eine analoge Diskretisierung auch funktioniert für $[a,\infty[$ oder $]-\infty,b]$ mit $\Omega = \mathbb{N}$ und für $\mathbb{R} =]-\infty,\infty[$ mit $\Omega = \mathbb{Z}$.

3.3 Erwartungswerte auf kontinuierlichen Wahrscheinlichkeitsräumen

Wie definiert man auf sinnvolle Weise nun einen Erwartungswert für eine Zufallsvariable? Wieder gibt die Diskretisierung den entscheidenden Hinweis:

Bemerkung 3.3.1 *Sei wie in Bemerkung 3.2.9*

$$f : [a,b] \to \mathbb{R}_{\geq 0}$$

eine stetige Dichte und für $n \in \mathbb{N}$ mit

$$t_i = a + (b-a) \cdot \frac{i}{n}$$

und

$$f_i = f(\frac{t_{i-1} + t_i}{2}) \cdot \frac{b-a}{n}$$

die Diskretisierung gegeben durch die Wahrscheinlichkeitsfunktion

$$m : \Omega = \{1, \ldots, n\} \to \mathbb{R}_{\geq 0}$$
$$m(i) = f_i$$

Eine stetige Zufallsvariable $X : [a, b] \to \mathbb{R}$ *können wir dann diskretisieren zu*

$$\widetilde{X} : \Omega \to \mathbb{R}$$
$$\widetilde{X}(i) = X(\frac{t_{i-1} + t_i}{2})$$

und erhalten den Erwartungswert

$$\begin{aligned} E(\widetilde{X}) &= \sum_{i=1}^{n} \widetilde{X}(i) \cdot m(i) \\ &= \frac{b-a}{n} \sum_{i=1}^{n} X(\frac{t_{i-1} + t_i}{2}) \cdot f(\frac{t_{i-1} + t_i}{2}). \end{aligned}$$

Mit dem Riemannintegral erhalten wir für $n \to \infty$ *dann*

$$E(\widetilde{X}) = \int_a^b X(x) \cdot f(x) dx.$$

Wir definieren daher:

Definition 3.3.2 *Sei* Ω *ein kontinuierlicher Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsdichte* f*. Für eine Zufallsvariable* $X : \Omega \to \mathbb{R}$ *defnieren wir den* ***Erwartungswert*** *als*

$$E(X) = \int_\Omega X(\omega) \cdot f(\omega) d\omega,$$

falls

$$\int_\Omega |X(\omega)| \cdot f(\omega) d\omega < \infty.$$

Analog definieren wir die ***Momente***

$$E(X^k) = \int_\Omega X(\omega)^k \cdot f(\omega) d\omega.$$

für alle $k \in \mathbb{N}$.

Als Faustregel können wir also sagen, dass sich Summen im Fall von kontinuierlichen Wahrscheinlichkeitsräumen durch die entsprechenden Integrale ersetzen.

Bemerkung 3.3.3 *Mit Definition 3.3.2 übertragen sich direkt die Definitionen von* ***Varianz***

$$\begin{aligned} V(X) &= E((X - E(X))^2) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

und ***Standardabweichung***

$$\sigma(X) = \sqrt{V(X)}$$

einer Zufallsvariable X, und der ***Covarianz***

$$\mathrm{Cov}(X_1, X_2) = E(\ (X_1 - E(X_1)) \cdot (X_2 - E(X_2))\ )$$

und ***Korrelation***

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma(X_1) \cdot \sigma(X_2)}$$

von Zufallsvariablen X_1 und X_2.

Bemerkung 3.3.4 *Mit den Rechenregeln für Integrale folgt, dass wie im diskreten Fall*

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

für Zufallsvariablen X_i für die der Erwartungswert existiert (siehe Satz 2.6.12) und

$$E(c \cdot X) = c \cdot E(X)$$

für $c \in \mathbb{R}$ gilt (siehe Bemerkung 2.6.14).
Ebenso gilt

$$V(c \cdot X) = c^2 \cdot V(X)$$

und

$$V(X + c) = V(X)$$

für $c \in \mathbb{R}$.

Beweis. Für den Erwartungswert haben wir

$$\begin{aligned} E(X_1 + X_2) &= \int_\Omega (X_1 + X_2)(\omega) \cdot f(\omega) d\omega \\ &= \int_\Omega (X_1(\omega) \cdot f(\omega) + X_2(\omega) \cdot f(\omega)) d\omega \\ &= \int_\Omega X_1(\omega) \cdot f(\omega) d\omega + \int_\Omega X_2(\omega) \cdot f(\omega) d\omega \\ &= E(X_1) + E(X_2) \end{aligned}$$

und

$$\begin{aligned} E(c \cdot X) &= \int_\Omega c \cdot X(\omega) \cdot f(\omega) d\omega \\ &= c \cdot \int_\Omega X(\omega) \cdot f(\omega) d\omega \\ &= c \cdot E(X). \end{aligned}$$

Für die Varianz gilt mit den Rechenregeln für den Erwartungswert dann genau wie im diskreten Fall

$$\begin{aligned} V(c \cdot X) &= E(c^2 \cdot X^2) - E(c \cdot X)^2 \\ &= c^2 \cdot E(X^2) - c^2 \cdot E(X)^2 \\ &= c \cdot V(X) \end{aligned}$$

und

$$\begin{aligned} V(X + c) &= E((X + c)^2) - E(X + c)^2 \\ &= E(X^2) + 2c \cdot E(X) + c^2 - E(X)^2 - 2c \cdot E(X) - c^2 \\ &= E(X^2) - E(X)^2 = V(X). \end{aligned}$$

■

Beispiel 3.3.5 *Für die Gleichverteilung auf* $[0,1]$ *mit Dichte* $f : [0,1] \to \mathbb{R}_{\geq 0}$

$$f(x) = 1$$

ist der Erwartungswert von $X = \mathrm{id}$

$$E(X) = \int_0^1 x \cdot f(x) dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

und die Varianz

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \int_0^1 x^2 \cdot f(x)dx - \frac{1}{4} \\ &= \left[\frac{x^3}{3}\right]_0^1 - \frac{1}{4} \\ &= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

Beispiel 3.3.6 *Wir werfen auf eine Zielscheibe Ω mit Radius 1 und die Zufallsvariable $X : \Omega \to \mathbb{R}$ beschreibe den Abstand des Treffers vom Mittelpunkt. Die Wahrscheinlichkeit in Radius $\leq r$*

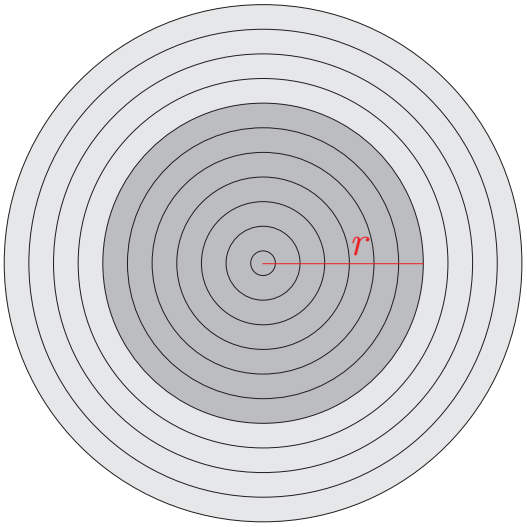


Abbildung 3.3: Zielscheibe und Trefferbereich mit Radius $\leq r$.

zu treffen ist proportional zur Fläche der Kreisscheibe mit Radius r (siehe Abbildung 3.3), d.h.

$$P(X \leq r) = P(\{(x, y) \mid x^2 + y^2 \leq r^2\}) = c \cdot \pi \cdot r^2.$$

Wir nehmen an, dass wir immer auf die Scheibe treffen. Dann ist

$$P(X \leq 1) = 1$$

also

$$c = \frac{1}{\pi}$$

und somit

$$P(X \leq r) = r^2.$$

Diese Wahrscheinlichkeitsverteilung können wir mit der Wahrscheinlichkeitsdichte

$$f(r) = 2r$$

darstellen als

$$P(X \leq r) = \int_0^r 2t \; dt = \int_0^r f(t) \; dt.$$

Damit können wir den Erwartungswert für den Trefferabstand vom Mittelpunkt berechnen als

$$\begin{aligned} E(X) &= \int_0^1 r \cdot f(r) \; dr = \int_0^1 2r^2 \; dr \\ &= \left[\frac{2}{3}r^3\right]_0^1 = \frac{2}{3}. \end{aligned}$$

Beispiel 3.3.7 *Für die* ***Exponentialverteilung***

$$f(x) = \lambda \exp(-\lambda x)$$

mit $\lambda > 0$ *auf* $\Omega = [0, \infty[$ *ist*

$$\int_0^\infty f(x)dx = [-\exp(-\lambda x)]_0^\infty = 0 - (-1) = 1,$$

somit definiert f *auf* Ω *einen Wahrscheinlichkeitsraum.*

Die Exponentialverteilung tritt oft bei Problemstellungen auf, bei denen ein Defekt auftritt, etwa bei einer Festplatte, SSD, LED oder einem Netzteil. Die Zufallsvariable $X = \mathrm{id}$ *gibt die Zeit bis zum Defekt an. Die Wahrscheinlichkeit eines Defekts bis Zeit* t *ist dann*

$$P(X \leq t) = \int_0^t f(x)dx = [-\exp(-\lambda x)]_0^t = 1 - \exp(-\lambda t).$$

Für $\lambda = \frac{1}{2}$ *erhalten wir z.B.*

$$P(X \leq 1) = 1 - \exp(-\frac{1}{2}) \approx 0.39.$$

Den Erwartungswert erhalten wir mit partieller Integration

$$\begin{aligned} E(X) &= \int_0^\infty x \cdot \lambda \cdot \exp(-\lambda x) dx \\ &= \lambda \cdot \int_0^\infty x \cdot \exp(-\lambda x) dx \\ &= \left[-x \cdot \exp(-\lambda x)\right]_0^\infty + \int_0^\infty \exp(-\lambda x) dx \\ &= 0 + \left[\frac{1}{-\lambda} \exp(-\lambda x)\right]_0^\infty = \frac{1}{\lambda} \end{aligned}$$

ebenso auch

$$\begin{aligned} E(X^2) &= \lambda \cdot \int_0^\infty x^2 \cdot \exp(-\lambda x) dx \\ &= \left[-x^2 \cdot \exp(-\lambda x)\right]_0^\infty + 2 \int_0^\infty x \cdot \exp(-\lambda x) dx \\ &= 2 \cdot \frac{1}{\lambda} \cdot E(X) = \frac{2}{\lambda^2} \end{aligned}$$

und damit die Varianz

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

In dem obigen Beispiel bedeutet $\lambda = \frac{1}{2}$ also, dass die erwartete Zeit bis zum Fehler bei 2 liegt (in einer festgelegten Einheit für die Zeitmessung, etwa Jahre).

3.4 De Buffons Nadelexperiment

Beispiel 3.4.1 *In de Buffons Nadelexperiment zeichnen wir äquidistant mit Abstand 1 parallele Geraden in eine große quadratische Fläche und werfen dann zufällig Nadeln der Länge 1. Wir zählen, wie oft eine Nadel eine Gerade schneidet, siehe Abbildung 3.4.*

Zu jeder Nadel assoziieren wir die nächste Gerade L, den Abstand m des Mittelpunkts der Nadel von L und den (kleinsten der beiden) Winkel φ, den die Nadel mit L einschließt. Das Experiment produziert damit Ergebnisse (m, φ) in

$$\Omega = \left[0, \frac{1}{2}\right] \times \left[0, \frac{\pi}{2}\right]$$

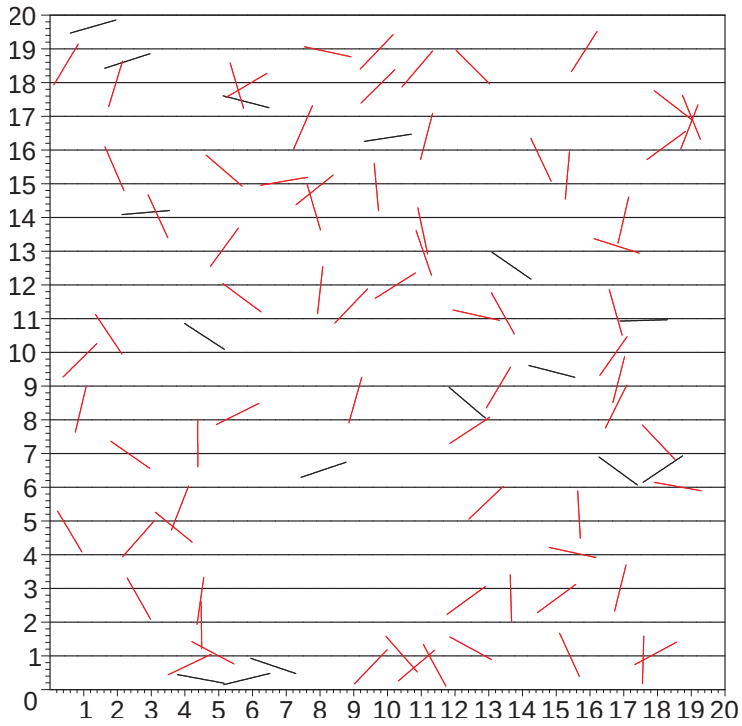


Abbildung 3.4: Buffons Nadelexperiment

und wir nehmen an, dass bei einem Wurf Elemente in Ω *zufällig gleich wahrscheinlich gewählt werden. Eine Nadel schneidet die nächste Gerade genau dann, wenn*

$$\frac{m}{\sin(\varphi)} \leq \frac{1}{2},$$

siehe Abbildung 3.5. Die günstigen Ereignisse nehmen also von der Gesamtfläche

$$\frac{1}{2} \cdot \frac{\pi}{2} = \frac{\pi}{4}$$

die Fläche

$$\int_0^{\frac{\pi}{2}} \frac{1}{2} \sin(\varphi) d\varphi = \frac{1}{2} \left[-\cos(\varphi) \right]_0^{\frac{\pi}{2}} = \frac{1}{2}$$

ein, siehe Abbildung 3.6. Die Wahrscheinlichkeit, dass die Nadel

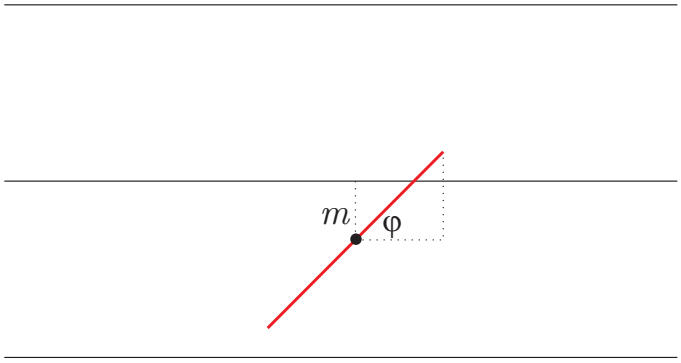


Abbildung 3.5: Winkel und Abstand zur nächsten Geraden.

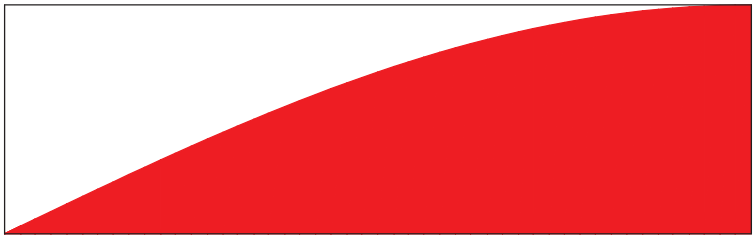


Abbildung 3.6: Günstige Ergebnisse im Nadelexperiment.

die nächste Gerade schneidet, ist mit dem Ereignis

$$A = \left\{ (m, \varphi) \in \Omega \mid m \leq \frac{1}{2} \sin(\varphi) \right\}$$

also wenn wir auf Ω *eine Gleichverteilung annehmen*

$$p = P(A) = \frac{\frac{1}{2}}{\frac{\pi}{4}} = \frac{2}{\pi}.$$

Somit haben wir für die Zufallsvariable

$$X(m, \varphi) = \begin{cases} 1 & \textit{Nadel schneidet} \\ 0 & \textit{sonst} \end{cases}$$

dass

$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p = \frac{2}{\pi}.$$

Wenn wir davon ausgehen, dass auch auf nicht-diskreten Wahrscheinlichkeitsräumen das Gesetz der großen Zahlen gilt, sollte nach Beispiel 2.13.7 die relative Häufigkeit, dass eine Nadel die nächste Gerade schneidet, den Erwartungswert $E(X) = \frac{2}{\pi}$ annähern. Wir haben also ein Verfahren zur Berechnung von π gefunden. In Aufgabe 3.5 implementieren wir das Buffonsche Nadelexperiment. Bei einer Implementierung sollte man die Verwendung von π und von trigonometrischen Funktionen vermeiden, da wir sonst schon implizit π als gegeben ansehen.

Bemerkung 3.4.2 *Um das Experiment etwas formaler mit unserer Notation von Wahrscheinlichkeitsdichten zu beschreiben, können wir also auf Ω die konstante Dichte*

$$f(m,\varphi) = \frac{4}{\pi}$$

einführen, die dann natürlich zu

$$\begin{aligned}\int_{\omega\in\Omega} f(\omega)d\omega &= \int_0^{\frac{\pi}{2}}\int_0^{\frac{1}{2}} \frac{4}{\pi}\, dm\, d\varphi = \frac{4}{\pi}\int_0^{\frac{\pi}{2}}\int_0^{\frac{1}{2}} 1\, dm\, d\varphi \\ &= \frac{4}{\pi}\int_0^{\frac{\pi}{2}} \frac{1}{2} d\varphi = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} 1 d\varphi \\ &= \frac{2}{\pi}\cdot\frac{\pi}{2} = 1\end{aligned}$$

integriert. Die Wahrscheinlichkeit, dass die Nadel, die nächste Gerade schneidet, ist damit

$$\begin{aligned}p = P(A) &= \int_A f(\omega)d\omega = \frac{4}{\pi}\cdot\int_A 1\, dm\, d\varphi \\ &= \frac{4}{\pi}\cdot\int_0^{\frac{\pi}{2}}\int_0^{\frac{1}{2}\sin(\varphi)} 1\, dm\, d\varphi = \frac{4}{\pi}\cdot\int_0^{\frac{\pi}{2}} \frac{1}{2}\sin(\varphi)\, d\varphi \\ &= \frac{4}{\pi}\cdot\frac{1}{2} = \frac{2}{\pi}.\end{aligned}$$

3.5 Unabhängigkeit im kontinuierlichen Fall

Um das Gesetz der großen Zahlen zu formulieren, müssen wir wieder den Begriff der **unabhängig und identisch verteilten** Zufallsvariablen einführen. Den Begriff der Unabhängigkeit müssen wir so wählen, dass er genau das leistet was wir brauchen, d.h. zu der Formel

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$$

und damit zu

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

führt, die essentiell im Beweis des Gesetzes der großen Zahlen verwendet werden. Unser bisheriger Beweis des Gesetzes der großen Zahlen überträgt sich dann direkt auf den kontinuierlichen Fall.

Beim Begriff der Unabhängigkeit gibt es ein ähnliches Problem wie bei dem Begriff der Wahrscheinlichkeit von Elementen. Die im diskreten Fall eingeführte Definition

$$P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1) \cdot P(X_2 = n_2)$$

macht zwar Sinn, falls X_1 und X_2 nur diskrete Werte annehmen. In diesem Fall gehorchen die Werte von X_1 und X_2 einer diskreten Verteilung und wir können direkt auf die Resultate aus Abschnitt 2 zurückgreifen. Im Allgemeinen kann das Bild von X_1 und X_2 aber auch kontinuierlich sein. In diesem Fall stellt obige Gleichung keine Bedingung, da typischerweise alle drei Wahrscheinlichkeiten in der Gleichung 0 sind. Wir betrachten daher für eine reellwertige Zufallsvariable die sogenannte **kumulative Wahrscheinlichkeit** $P(X \leq n)$ und definieren:

Definition 3.5.1 *Zwei Zufallsvariablen $X_1, X_2 : \Omega \to \mathbb{R}$ heißen **unabhängig**, wenn*

$$P(X_1 \leq n_1, X_2 \leq n_2) = P(X_1 \leq n_1) \cdot P(X_2 \leq n_2)$$

für alle n_1, n_2.

Bemerkung 3.5.2 *Im Fall, dass X_1 und X_2 nur abzählbar viele Werte annehmen (z.B. falls der Wahrscheinlichkeitsraum diskret ist) stimmt diese Definition mit unserer bisherigen überein.*

Beweis. Wir können dann annehmen, dass X_1 und X_2 ganzzahlige Werte annehmen. Dann folgt aus

$$P(X_1 \leq n_1, X_2 \leq n_2) = P(X_1 \leq n_1) \cdot P(X_2 \leq n_2)$$

dass

$$\begin{aligned} P(X_1 = n_1, X_2 = n_2) &= P(X_1 \leq n_1, X_2 \leq n_2) \\ &\quad -P(X_1 \leq n_1 - 1, X_2 \leq n_2) \\ &\quad -P(X_1 \leq n_1, X_2 \leq n_2 - 1) \\ &\quad +P(X_1 \leq n_1 - 1, X_2 \leq n_2 - 1) \\ &= P(X_1 \leq n_1) \cdot P(X_2 \leq n_2) \\ &\quad -P(X_1 \leq n_1 - 1) \cdot P(X_2 \leq n_2) \\ &\quad -P(X_1 \leq n_1) \cdot P(X_2 \leq n_2 - 1) \\ &\quad +P(X_1 \leq n_1 - 1) \cdot P(X_2 \leq n_2 - 1) \\ &= P(X_1 = n_1) \cdot P(X_2 \leq n_2) \\ &\quad -P(X_1 = n_1) \cdot P(X_2 \leq n_2 - 1) \\ &= P(X_1 = n_1) \cdot P(X_2 = n_2) \end{aligned}$$

und umgekehrt aus

$$P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1) \cdot P(X_2 = n_2)$$

mit dem Distributivgesetz, dass

$$\begin{aligned} P(X_1 \leq n_1, X_2 \leq n_2) &= \sum_{z_1 \leq n_1} \sum_{z_2 \leq n_2} P(X_1 = z_1, X_2 = z_2) \\ &= \sum_{z_1 \leq n_1} \sum_{z_2 \leq n_2} P(X_1 = z_1) \cdot P(X_2 = z_2) \\ &= \left(\sum_{z_1 \leq n_1} P(X_1 = z_1) \right) \cdot \left(\sum_{z_2 \leq n_2} P(X_2 = z_2) \right) \\ &= P(X_1 \leq n_1) \cdot P(X_2 \leq n_2). \end{aligned}$$

■

Beispiel 3.5.3 *Bei zweimaligem Würfeln mit den Ergebnissen X_1 und X_2 ist*

$$P(X_1 \leq 3, X_2 \leq 3) = 9 \cdot \frac{1}{36} = \frac{1}{4}$$

denn die zulässigen Ergebnisse sind

$$(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3).$$

Andererseits ist

$$P(X_1 \leq 3) = P(X_2 \leq 3) = \frac{1}{2}.$$

Bemerkung 3.5.4 *Die Unabhängigkeit von mehr als zwei Zufallsvariablen*

$$X_1, \ldots, X_r : \Omega \to \mathbb{R}$$

definiert man analog als

$$P(X_1 \leq n_1, \ldots, X_r \leq n_r) = P(X_1 \leq n_1) \cdot \ldots \cdot P(X_r \leq n_r).$$

3.6 Wahrscheinlichkeitsdichten von Zufallsvariablen

3.6.1 Kontinuierliche Zufallsvariablen

Wir schränken uns im Folgenden auf Zufallsvariablen ein, für die wir auf dem Wertebereich eine Wahrscheinlichkeitsdichte angeben können. Da wir, wie gerade gesehen, in der Untersuchung von Unabhängigkeit auch Wahrscheinlichkeiten der Form $P(X_1 \leq n_1, X_2 \leq n_2)$ betrachten müssen, d.h. die Verteilung der Zufallsvariable $(X_1, X_2) : \Omega \to \mathbb{R}^2$, führen wir vektorwertige Zufallsvariablen ein.

Definition 3.6.1 *Eine **kontinuierliche Zufallsvariable** ist eine Zufallsvariable*

$$X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$$

*für die es eine integierbare **Wahrscheinlichkeitsdichte***

$$f_X : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$$

gibt mit

$$\int_{\mathbb{R}^d} f_X(t)dt = 1$$

und

$$P(X \leq n) = \int_{t \leq n} f_X(t)dt$$

für alle $n \in \mathbb{R}^d$. *Hier sind die Ungleichungen komponentenweise zu verstehen.*

Wir sagen, dass zwei kontinuierliche Zufallsvariablen X_1 *und* X_2 ***identisch verteilt*** *sind, wenn*

$$f_{X_1} = f_{X_2}.$$

Die Dichte f_X im Wertebereich ist das Analogon zu der Verteilung

$$P(X = n) = \sum_{\substack{\omega \in \Omega \\ X(\omega)=n}} m(\omega).$$

im diskreten Fall, siehe Bemerkung 2.6.5. In Beispiel 2.6.6 hatten wir etwa die Verteilung der Anzahl X von Kopf bei einem 4-maligen Münzwurf bestimmt:

n	0	1	2	3	4
ω mit $X(\omega) = n$	1111	1110	1100	0001	0000
		1101	1001	0010	
		1011	0011	0100	
		0111	1010	1000	
			0101		
			0110		
$P(X = n)$	$\frac{1}{2^4}$	$\frac{4}{2^4}$	$\frac{6}{2^4}$	$\frac{4}{2^4}$	$\frac{1}{2^4}$

Die Analogie ist also:

diskret	kontinuierlich
$P(X \leq n) = \sum_{t \leq n} P(X = t)$	$P(X \leq n) = \int_{t \leq n} f_X(t)dt$

Beispiel 3.6.2 *Ist* $\Omega = [a, b]$ *ein Intervall und*

$$X : [a, b] \to \mathbb{R}$$

eine streng monoton wachsende (analog fallende) Zufallsvariable, sodass

$$P(X \leq n)$$

differenzierbar ist, dann ist nach dem Hauptsatz der Differential- und Integralrechnung

$$f_X : [X(a), X(b)] \to \mathbb{R}_{\geq 0}$$
$$f_X = \frac{d}{dn} P(X \leq n)$$

eine Wahrscheinlichkeitsdichte mit

$$\int_{X(a)}^{X(b)} f_X(t)dt = 1$$

und

$$P(X \leq n) = \int_{X(a)}^{n} f_X(t)dt$$

für alle n.

Beispiel 3.6.3 *In Beispiel 3.3.6 können wir*

$$\Omega = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$$

als die Einheitskreisscheibe nehmen mit der Dichte

$$f(x_1, x_2) = \frac{1}{\pi}$$

nehmen. Offenbar ist dann $P(\Omega) = 1$. Die Verteilung der Zufallsvariable

$$X(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$$

die den Abstand des Treffers vom Mittelpunkt beschreibt erhalten wir dann als

$$\begin{aligned} P(X \leq r) &= P(\{(x_1, x_2) \in \mathbb{R}^2 \mid X(x_1, x_2) \leq r\}) \\ &= P(\{(x_1, x_2) \in \mathbb{R}^2 \mid \sqrt{x_1^2 + x_2^2} \leq r\}) \\ &= \tfrac{1}{\pi}\pi r^2 = r^2 \end{aligned}$$

Mit dem Hauptsatz der Differential- und Integralrechnung folgt dann

$$f_X = \frac{d}{dr} P(X \leq r)$$

also

$$f_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$$
$$f_X(r) = 2r.$$

Bemerkung 3.6.4 *Die Unabhängigkeit von kontinuierlichen Zufallsvariablen X_1 und X_2, lässt sich in Termen der Dichten ausdrücken als*

$$\begin{aligned} P(X_1 \leq n_1, X_2 \leq n_2) &= P(X_1 \leq n_1) \cdot P(X_2 \leq n_2) \\ &= \int_{t_1 \leq n_1} f_{X_1}(t_1) dt_1 \cdot \int_{t_2 \leq n_2} f_{X_2}(t_2) dt_2 \\ &= \int_{t_1 \leq n_1} \int_{t_2 \leq n_2} f_{X_1}(t_1) f_{X_2}(t_2) dt_1 dt_2 \end{aligned}$$

d.h. (nach dem Hauptsatz der Differential- und Integralrechnung) sind die Zufallsvariablen X_1 und X_2 unabhängig genau dann, wenn die gemeinsame Verteilung von X_1 und X_2 gegeben ist durch das Produkt der Dichten von X_1 und X_2, also:

$$X_1 \text{ und } X_2 \text{ unabhängig} \Longleftrightarrow f_{(X_1,X_2)} = f_{X_1} \cdot f_{X_2}.$$

Beispiel 3.6.5 *Wollen wir gleichverteilt einen Punkt im Quadrat $\Omega = [0,2]^2$ wählen, indem wir gleichverteilt die beiden Koordinaten X_1 und X_2 wählen, dann ist*

$$f_{(X_1,X_2)} = f_{X_1} \cdot f_{X_2} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Die Wahrscheinlichkeit, dass die beide Koordianten ≤ 1 sind, erhalten wir dann als

$$P(X_1 \leq 1, X_2 \leq 1) = \int_0^1 \int_0^1 \frac{1}{4} dx_2 dx_1 = \frac{1}{4}.$$

3.6.2 Berechnung von Wahrscheinlichkeitsdichten von kontinuierlichen Zufallsvariablen

Im diskreten Fall war es einfach, aus der Wahrscheinlichkeitsfunktion m auf Ω mittels der Formel

$$P(X = n) = \sum_{\substack{\omega \in \Omega \\ X(\omega) = n}} m(\omega)$$

die Verteilung der Zufallsvariable X zu bestimmen. Im Fall einer reellwertigen Zufallsvariable auf einem Intervall wie in Beispiel 3.6.2 können wir ebenso direkt aus der Dichte auf dem Intervall die Dichte der Zufallsvariablen ausrechnen:

Bemerkung 3.6.6 *Ist $\Omega = \mathbb{R}$ mit Wahrscheinlichkeitsdichte $f : \Omega \to \mathbb{R}_{\geq 0}$ und $X : \Omega \to \mathbb{R}$ eine streng monoton wachsende Funktion, dann ist eine Wahrscheinlichkeitsdichte von X gegeben durch*

$$f_X = (f \circ X^{-1}) \cdot \left(X^{-1}\right)'.$$

Beweis. Schreiben wir

$$P(X \leq n) = \int_{-\infty}^{n} f_X(t)dt$$

dann ist nach dem Hauptsatz der Differential- und Integralrechung $P(X \leq a)$ eine Stammfunktion von f_X, d.h.

$$\frac{d}{dn}P(X \leq n) = f_X(n).$$

Andererseits ist

$$\begin{array}{rcl} P(X \leq n) & = & P(\{\omega \in \mathbb{R} \mid X(\omega) \leq n\}) \\ & = & P(\{\omega \in \mathbb{R} \mid \omega \leq X^{-1}(n)\}) \\ & = & \int_{\omega \leq X^{-1}(n)} f(\omega)d\omega \\ & = & \int_{t \leq n} f(X^{-1}(t)) \cdot (X^{-1})'(t)dt \end{array}$$

wobei wir in der letzten Gleichung die Substitutionsregel für Integrale verwendet haben. Wieder mit dem Hauptsatz der Differential- und Integralrechung erhalten wir

$$\frac{d}{dn}P(X \leq n) = f(X^{-1}(n)) \cdot (X^{-1})'(n)$$

und somit

$$f_X(n) = f(X^{-1}(n)) \cdot (X^{-1})'(n).$$

■

Die Analogie ist also:

diskret	kontinuierlich
$m : \Omega \to \mathbb{R}_{\geq 0}$	$f : \Omega \to \mathbb{R}_{\geq 0}$
$P(X = n) = \sum_{\omega \in X^{-1}(\{n\})} m(\omega)$	$f_X(n) = f(X^{-1}(n)) \cdot (X^{-1})'(n)$

Beispiel 3.6.7 *Stadtgründungen in Deutschland passieren in den letzten* 2000 *Jahren gleichverteilt. Auf* $\Omega = [0, 2000]$ *betrachten wir also die Wahrscheinlichkeitsdichte*

$$f(t) = \frac{1}{2000}$$

wobei t *für die Zeit seit der Stadtgründung steht. Die Stadtgrößen wachsen mit der Zeit exponentiell (mit Basis* λ *leicht größer als* 1*). Bis auf Skalierungsfaktoren wird die Stadtgröße bei einer Gründung vor* t *Jahren beschrieben durch die Zufallsvariable*

$$X(t) = \lambda^t = \exp(\ln(\lambda) \cdot t).$$

mit Umkehrfunktion

$$X^{-1}(x) = \frac{\ln(x)}{\ln(\lambda)}$$

Die Dichte von X *ist dann*

$$\begin{aligned} f_X(x) &= f(X^{-1}(x)) \cdot \left(X^{-1}\right)'(x) \\ &= \frac{1}{2000} \cdot \frac{1}{x \cdot \ln(\lambda)} \end{aligned}$$

Damit ist

$$\begin{aligned} P(a \le X \le b) &= \int_a^b \tfrac{1}{2000} \cdot \tfrac{1}{x \cdot \ln(\lambda)} dx \\ &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot \int_a^b \tfrac{1}{x} dx \\ &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot [\ln(x)]_a^b \\ &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot (\ln(b) - \ln(a)) \end{aligned}$$

Wir betrachten das Ruhrgebiet mit etwa 8000000 *Einwohnern als die größte deutsche Stadt. Tatsächlich wurde Köln als Zentrum des Ruhrgebiets vor fast genau* 2000 *Jahren gegründet. Wir können eine Näherung für* λ *bestimmen, indem wir die Gleichung*

$$\lambda^{2000} = 8000000$$

lösen und erhalten

$$\lambda = (8000000)^{\frac{1}{2000}} = \exp\left(\frac{\ln(8000000)}{2000}\right) \approx 1.008.$$

Damit ist dann

$$P(1 \leq X \leq 8000000) = 1$$

und z.B.

$$P([100000, 200000]) = \frac{1}{2000} \cdot \frac{1}{\ln(\lambda)} \cdot [\ln(x)]_{10^5}^{2 \cdot 10^5} \approx 0.0449$$
$$P([800000, 900000]) = \frac{1}{2000} \cdot \frac{1}{\ln(\lambda)} \cdot [\ln(x)]_{8 \cdot 10^5}^{9 \cdot 10^5} \approx 0.0076,$$

wir erwarten also, dass etwa 0.8% *aller Städte eine Größe zwischen* 800000 *und* 900000 *Einwohnern haben, aber* 4.5% *aller Städte eine Größe zwischen* 100000 *und* 200000 *Einwohnern.*

Tatsächlich gilt unabhängig von k *mit der Funktionalgleichung*

$$\ln(u \cdot v) = \ln(u) + \ln(v)$$

des Logarithmus, dass

$$\begin{aligned} P(10^k \leq X \leq 2 \cdot 10^k) &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot (\ln(2 \cdot 10^k) - \ln(10^k)) \\ &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot \ln(2) \approx 0.0449 \end{aligned}$$

und

$$\begin{aligned} P(8 \cdot 10^k \leq X \leq 9 \cdot 10^k) &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot (\ln(9 \cdot 10^k) - \ln(8 \cdot 10^k)) \\ &= \tfrac{1}{2000} \cdot \tfrac{1}{\ln(\lambda)} \cdot (\ln(9) - \ln(8)) \approx 0.0076. \end{aligned}$$

Daraus folgt etwas ziemlich seltsames: Städte mit einer Einwohnerzahl, die mit 1 *beginnt, sind wesentlich häufiger ist als Städte, deren Einwohnerzahl mit mit einer* 8 *beginnt. Dies kann man tatsächlich an den Einwohnerzahlen der deutschen Städte beobachten.*

3.6.3 Erwartungswerte von kontinuierlichen Zufallsvariablen

Wir schränken uns im Folgenden auf eine streng monotone Zufallsvariable auf $\Omega = \mathbb{R}$ ein, die Formel für den Erwartunswert gilt aber allgemeiner sofern man eine Dichte im Bildraum angeben kann und das Integral existiert.

Proposition 3.6.8 *Den Erwartungswert einer kontinuierlichen monotonen Zufallsvariable $X : \Omega \to \mathbb{R}$ können wir mit Hilfe der Wahrscheinlichkeitsdichte $f_X : \Omega \to \mathbb{R}$ berechnen als*

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t) dt.$$

Beweis. Sei f die gegebene Wahrscheinlichkeitsdichte auf dem kontinuierlichen Wahrscheinlichkeitsraum $\Omega = \mathbb{R}$. Mit der aus der Analysis bekannten Substitutionsregel für Integrale[2] erhalten wir für die Substitution $\omega = X^{-1}(t)$, dass

$$\begin{aligned} E(X) &= \int_{\Omega} X(\omega) \cdot f(\omega) d\omega \\ &= \int_{-\infty}^{\infty} t \cdot f(X^{-1}(t)) \cdot (X^{-1})'(t) dt \\ &= \int_{-\infty}^{\infty} t \cdot f_X(t) dt, \end{aligned}$$

wobei wir in der letzten Gleichheit Bemerkung 3.6.6 verwendet haben. ■

Die Analogie zum diskreten Fall ist also

diskret	kontinuierlich
$E(X) = \sum_t t \cdot P(X = t)$	$E(X) = \int t \cdot f_X(t) dt.$

Beispielrechnungen haben wir schon den Beispielen 3.3.5, 3.3.6 und 3.3.7 gesehen.

Wie im diskreten Fall folgt mit der Proposition:

Satz 3.6.9 *Für unabhängige kontinuierliche Zufallsvariablen X_1 und X_2 gilt*

$$\begin{aligned} E(X_1 \cdot X_2) &= E(X_1) \cdot E(X_2) \\ V(X_1 + X_2) &= V(X_1) + V(X_2). \end{aligned}$$

[2] Wir erinnern uns an die Substitutionsregel: Ist $f : [r, s] \to \mathbb{R}$ stetig und $g : [a, b] \to [r, s]$ differenzierbar mit stetiger Ableitung, dann gilt

$$\int_r^s (f \circ g)(x) \cdot g'(x)\ dx = \int_{g(r)}^{g(s)} f(y)\ dy.$$

Dies folgt direkt aus der Kettenregel für Ableitungen.

Beweis. Mit Bemerkung 3.6.4 hat (X_1, X_2) auf dem Bildraum die Wahrscheinlichkeitsdichte

$$f(t_1, t_2) := f_{(X_1,X_2)}(t_1, t_2) = f_{X_1}(t_1) \cdot f_{X_2}(t_2)$$

und auf dem Bildraum von X_1 und X_2 hat $X_1 \cdot X_2$ den Wert

$$(X_1 \cdot X_2)(t_1, t_2) = t_1 \cdot t_2.$$

Den Erwartungswert von $X_1 \cdot X_2$ damit berechnen als

$$\begin{aligned} E(X_1 \cdot X_2) &= \int \int (X_1 \cdot X_2)(t_1, t_2) \cdot f(t_1, t_2)\ dt_2 dt_1 \\ &= \int \int t_1 \cdot t_2 \cdot f_{X_1}(t_1) \cdot f_{X_2}(t_2)\ dt_2 dt_1 \\ &= \int t_1 \cdot f_{X_1}(t_1) \cdot \int t_2 \cdot f_{X_2}(t_2)\ dt_2 dt_1 \\ &= \int t_2 \cdot f_{X_2}(t_2)\ dt_2 \cdot \int t_1 \cdot f_{X_1}(t_1)\ dt_1 \\ &= E(X_1) \cdot E(X_2). \end{aligned}$$

Die Gleichung für die Varianz folgt exakt wie in Satz 2.9.9 (aus der für den Erwartungswert und Bemerkung 3.3.4). ■

Bemerkung 3.6.10 *Man beachte, dass diese Formeln auch für unabhängige Zufallsvariablen gelten, die auf einem kontinuierlichen Wahrscheinlichkeitsraum diskrete Werte annehmen. Liegen die Werte von X_1 in der abzählbaren Menge N_1 und die von X_2 in N_2, dann haben wir (wie im Fall von Zufallsvariablen auf diskreten Wahrscheinlichkeitsräumen, siehe den Beweis von Satz 2.9.7)*

$$\begin{aligned} E(X_1 \cdot X_2) &= \sum_{n_1 \in N_1} \sum_{n_2 \in N_2} n_1 \cdot n_2 \cdot P(X_1 = n_1, X_2 = n_2) \\ &= \sum_{n_1 \in N_1} \sum_{n_2 \in N_2} n_1 \cdot n_2 \cdot P(X_1 = n_1) \cdot P(X_2 = n_2) \\ &= \sum_{n_1 \in N_1} n_1 \cdot P(X_1 = n_1) \cdot \sum_{n_2 \in N_2} n_2 \cdot P(X_2 = n_2) \\ &= E(X_1) \cdot E(X_2). \end{aligned}$$

Dies impliziert wiederum die Gleichung $V(X_1 + X_2) = V(X_1) + V(X_2)$ für die Varianz.

3.7 Mittelwerte von Zufallsvariablen

3.7.1 Gesetz der großen Zahlen

Auf einem kontinuierlichen Wahrscheinlichkeitsraum gilt auch wieder die Markov-Ungleichung:

Satz 3.7.1 (Markov-Ungleichung) *Sei Ω ein kontinuierlicher Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsdichte $f : \Omega \to \mathbb{R}_{\geq 0}$ und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $X \geq 0$. Für alle $\varepsilon > 0$ gilt*

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Beweis. Wir setzen

$$g(x) = \begin{cases} 1 & \text{falls } X(x) \geq \varepsilon \\ 0 & \text{sonst} \end{cases}$$

Wegen $X \geq 0$ haben wir

$$\begin{aligned} E(X) &= \int_\Omega X(\omega) \cdot f(\omega)\ d\omega \geq \int_\Omega X(\omega) \cdot f(\omega) \cdot g(\omega)\ d\omega \\ &\geq \varepsilon \cdot \int_\Omega f(\omega) \cdot g(\omega)\ d\omega = \varepsilon \cdot P(X \geq \varepsilon) \end{aligned}$$

■

Damit übertragen sich aus dem diskreten Fall ohne jede Änderung die Beweise der Tschebyscheffungleichung (Satz 2.12.3) und des Gesetzes der großen Zahlen (Satz 2.13.5), wobei wir für letzteres die Gleichung

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

aus Satz 3.6.9 und Bemerkung 3.6.10 verwenden:

Bemerkung 3.7.2 *Auch für kontinuierliche Zufallsvariabeln gilt die Tschebyscheffungleichung*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

und das Gesetz der großen Zahlen

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2},$$

insbesondere

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

wobei $X_1, \ldots, X_n$ identisch unabhängig verteilte Zufallsvariablen sind mit Erwartungswert $\mu = E(X_i)$ und Standardabweichung $\sigma = \sigma(X_i)$ und $\varepsilon > 0$.

Unsere Intuition bei Buffons Nadelexperiment, durch eine relative Häufigkeit eine Wahrscheinlichkeit anzunähern, war also korrekt.

3.7.2 Anwendung: Monte-Carlo-Integration

Bisher haben wir Integrale verwendet um kontinuierliche Zufallsprozesse zu beschreiben. Tatsächlich geht auch die Umkehrung: Man kann Integrale mittels eines Zufallsprozesses berechnen.

Beispiel 3.7.3 *Wir betrachten das Quadrat*

$$Q = [0,1]^2$$

und die Funktion

$$g(x) = x^2.$$

Wir wählen zufällig und gleich verteilt Punkte $(x,y) \in Q$ und

$$X(x,y) = \begin{cases} 1 & \text{falls } y \leq x^2 \\ 0 & \text{sonst} \end{cases}$$

Es ist dann

$$p = P(X = 1) = \frac{\int_0^1 x^2 dx}{1 \cdot 1} = \int_0^1 x^2 dx$$

gleich dem Anteil der Fläche von Q unter der Parabel, also

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = \int_0^1 x^2 dx.$$

Nach dem Gesetz der großen Zahlen können wir das Integral also annähern, indem wir n-mal X auswerten und den Mittelwert bilden.

Wir wählen nun mit dem folgenden MAPLE-*Programm* 10000 *Punkte in* Q *und bilden den Mittelwert der Werte von* X*, um damit* $E(X)$ *zu approximieren:*

```
N:=10000:
Px:=[stats[random, uniform](N)]:
Py:=[stats[random, uniform](N)]:
c:=0:
for j from 1 to N do
    if (Py[j]<=Px[j]^2) then c:=c+1;fi;
od:
evalf(c/N);
```

0.337

Siehe Abbildung 3.7.

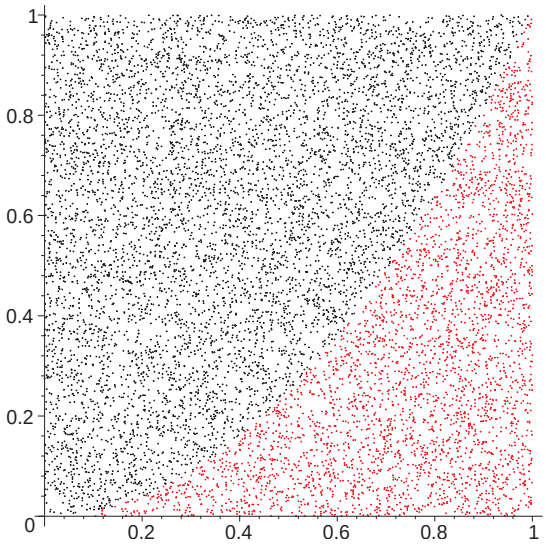


Abbildung 3.7: Montecarlo-Integration

Natürlich können wir hier exakt das Integral auswerten als

$$p = \int_0^1 x^2 dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}.$$

Das Verfahren lasst sich aber analog auf jede stetige Funktion anwenden und liefert dann eine Approximation der Fläche unter dem Funktionsgraphen, also des Integrals. Siehe dazu Übungsaufgabe 3.6.

Bemerkung 3.7.4 *Das Verfahren lässt sich noch wie folgt verbessern: Zur Integration der stetigen Funktion*

$$g : [a, b] \to \mathbb{R}$$

wählen wir mit der Zufallsvariable X zufällig gleichverteilt in $[a, b]$ Werte für x mit Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{b-a}.$$

Beschreibt die Zufallsvariable $Y = g(X)$ die entsprechenden Funktionswerte, dann ist

$$\begin{aligned} E(Y) &= \int_a^b Y(x) \cdot f(x)\ dx \\ &= \int_a^b g(x) \cdot \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \cdot \int_a^b g(x) dx. \end{aligned}$$

Beschreiben $X_1, \ldots, X_n$ die Ergebnisse einer zufälligen, unabhängigen, gleichverteilten Wahl von n Werten in $[a, b]$, dann können wir mit

$$\frac{b-a}{n} \sum_{i=1}^{n} g(X_i)$$

nach dem Gesetz der großen Zahlen das Integral annähern. Wären die Werte der X_i genau äquidistant auf das Intervall $[a, b]$ verteilt, dann wäre dies genau die Formel zur Riemannintegration.

Beispiel 3.7.5 *Für $g(x) = x^2$ können wir das Integral über $[0, 2]$ berechnen mit*

```
N:=10000:
Px:=[2*stats[random, uniform](N)]:
s:=0:
for j from 1 to N do
    s:=s+Px[j]^2;
od:
evalf(2*s/N);
```

2.688

Der exakte Wert ist

$$\int_0^2 x^2 dx = \frac{8}{3} = 2.66\ldots$$

Wir schätzen noch mit der Tschebyscheff-Ungleichung die Genauigkeit des Integrationsverfahrens ab:

Bemerkung 3.7.6 *Nehmen wir der Einfachheit halber an, dass g nur Werte in $[0,1]$ annimmt. Setzen wir weiter*

$$I = (b-a) \cdot Y$$

also

$$E(I) = \int_a^b g(x)dx$$

dann ist

$$V(I) = \int_a^b g(x)^2 dx - E(I)^2 \leq (b-a)^2$$

denn $(g(x)^2 \leq 1)$und somit mit der Schranke aus dem Gesetz der großen Zahlen

$$p = P\left(\left|\frac{b-a}{n}\sum_{i=1}^{n} g(X_i) - E(I)\right| \geq \varepsilon\right) \leq \frac{(b-a)^2}{n \cdot \varepsilon^2}.$$

Für das Integral von $g(x) = x^2$ über das Intervall $[0,1]$ ist die Wahrscheinlichkeit eines Fehlers von ≥ 0.02 nach $n = 10000$ Iterationen höchstens

$$\frac{1}{10000 \cdot 0.02^2} = 0.25.$$

Bemerkung 3.7.7 *Die Hoeffding-Ungleichung, die genau wie im diskreten Fall auch im kontinuierlichen Fall gilt, liefert eine Abschätzung, die unsere beobachtete Genauigkeit wesentlich besser abbildet mit*

$$p \leq 2\exp(-2 \cdot \varepsilon^2 \cdot n)$$

also in unserem Beispiel

ε	$2\exp(-2 \cdot \varepsilon^2 \cdot 10000)$
0.02	≈ 0.00068
0.01	≈ 0.27

3.8 Konvergenz von Verteilungen

In diesem Abschnitt wollen wir zwischen Verteilungen Beziehungen herstellen, die durch Grenzwertbildung entstehten. Wir beginnen mit einer uns schon bekannten diskreten Verteilung.

3.8.1 Binomialverteilung

Definition 3.8.1 *Bei einem **Bernoulliprozess** führen wir ein Zufallsexperiment mit boolschem Ergebnis n-mal durch. Die einzelnen Iterationen sind unabhängig und identisch verteilt und geben mit Wahrscheinlichkeit p true und mit Wahrscheinlichkeit $1-p$ false. Wir haben also*

$$\Omega = \{0,1\}^n$$

und

$$m(\omega) = p^j \cdot (1-p)^{n-j}$$

wobei j die Anzahl der Einträge von Einsen in ω ist.

Definition 3.8.2 *Wir interessieren uns nun für die Anzahl von Einsen in einem Bernoulliprozess. Ist Y_i eine Zufallsvariablen, die das Ergebnis des i-ten Wurfs liefert, also*

$$Y_i(\omega) = \omega_i$$

wird diese Anzahl gegeben durch die Zufallsvariable

$$X_n = Y_1 + \ldots + Y_n.$$

*Diese hat eine sogenannte **Binomialverteilung***

$$P(X_n = j) = \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j}$$

da wir j Möglichkeiten haben, die Positionen der Einsen aus den insgesamt n Positionen zu wählen.

Beispiel 3.8.3 *Das Musterbeispiel einer Binomialverteilung ist die Verteilung der Anzahl X von Kopf bei einem mehrfachen Münzwurf (wobei wir z.B. 0 für Kopf und 1 für Zahl schreiben). Wie in Beispiel 2.6.6 diskutiert erhalten wir die Verteilung*

j	0	1	2	3	4
ω *mit* $X(\omega) = j$	1111	1110	1100	0001	0000
		1101	1001	0010	
		1011	0011	0100	
		0111	1010	1000	
			0101		
			0110		
$P(X = j)$	$\frac{1}{2^4}$	$\frac{4}{2^4}$	$\frac{6}{2^4}$	$\frac{4}{2^4}$	$\frac{1}{2^4}$

also

$$P(X = j) = \binom{4}{j}\left(\frac{1}{2}\right)^4.$$

3.8.2 Poissonverteilung

Neben der Gleichverteilung und der Binomialverteilung ist eine der wichtigsten diskreten Verteilungen die Poissonverteilung. Wie wir sehen werden, kann man die Poissonverteilung als einen Limes der Binomialverteilung für $n \to \infty$ auffassen.

Nehmen wir an, wir wollen die Wahrscheinlichkeit beschreiben, wie oft ein Ereignis in einem Zeitintervall auftritt. Wir könnten z.B. ein radioaktives Material untersuchen, das bestimmte Teilchen aussendet, etwa Elektronen in einem β-Zerfall.

Ohne Einschränkung können wir das Zeitintervall $[0, 1]$ betrachten. Sei X eine Zufallsvariable, die die Anzahl der Ereignisse in $[0, 1]$ beschreibt. Falls die Ereignisse mit konstanter Rate λ von Ereignissen pro Sekunde passieren, dann erwarten wir in unserem Zeitintervall $\lambda \cdot 1 = \lambda$ Ereignisse. Wir groß ist die Wahrscheinlichkeit, dass in dem Zeitintervall $[0, 1]$ genau k Ereignisse passieren, also as ist $P(X = k)$?

Wir können nun das Intervall $[0, 1]$ wie oben in n Teile diskretisieren. Nehmen wir n groß genug, dann wird in jedem Teilintervall der Breite $\frac{1}{n}$ maximal 1 Ereignis liegen. Die Zufallsvariable X_n gebe die Gesamtzahl der Ereignisse. Die Wahrscheinlichkeit, dass ein Ereignis in einem Teilintervall auftritt ist dann

$$p = \frac{\lambda}{n}$$

und somit mit der Binomialverteilung

$$P(X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

insbesondere haben wir

$$P(X_n = 0) = (1 - \frac{\lambda}{n})^n$$

die Wahrscheinlichkeit, dass kein Ereignis eintritt. Idealerweise wollen wir aber keine Abhängigkeit von der Diskretisierung, wir sollten also den Limes $n \to \infty$ bilden. Dazu verwenden wir:

Lemma 3.8.4 *Für alle $x \in \mathbb{R}$ gilt*

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \exp(x).$$

Beweis. Nach Definition der allgemeinen Potenz gilt (falls wir n groß genug wählen, dass $1+\frac{x}{n} > 0$)

$$\left(1+\frac{x}{n}\right)^n = \exp(\ln(1+\frac{x}{n})\cdot n).$$

Weiter gilt mit der Regel von l'Hospital, dass

$$\lim_{n\to\infty}\left(\ln(1+\frac{x}{n})\cdot n\right) = \lim_{z\to 0}\frac{\ln(1+x\cdot z)}{z} = \lim_{z\to 0}\frac{1}{1+x\cdot z}\cdot x = x.$$

Mit der Stetigkeit der Exponentialfunktion folgt die Behauptung. ■

Corollar 3.8.5 *Im Grenzwert ist*

$$P(X=0) = \lim_{n\to\infty}(1-\frac{\lambda}{n})^n = \exp(-\lambda).$$

Was erhalten wir für k positiv? Da

$$\frac{P(X_n=k)}{P(X_n=k-1)} = \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1}} = \frac{n-k+1}{k}\frac{p}{1-p}$$
$$= \frac{n-k+1}{k}\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}} = \frac{\lambda-(k-1)\frac{\lambda}{n}}{k\cdot(1-\frac{\lambda}{n})}$$

haben wir

$$\frac{P(X=k)}{P(X=k-1)} = \lim_{n\to\infty}\frac{\lambda-(k-1)\frac{\lambda}{n}}{k\cdot(1-\frac{\lambda}{n})} = \frac{\lambda}{k}$$

also mit Induktion

$$P(X=k) = \frac{\lambda^k}{k!}\exp(-\lambda).$$

Definition 3.8.6 *Auf $\Omega = \mathbb{N}_0$ ist die **Poissonverteilung** gegeben durch*

$$P(X=k) = \frac{\lambda^k}{k!}\exp(-\lambda).$$

Nach Konstruktion lässt sich die Poissonverteilung sehr genau durch die Binomialverteilung annähern, indem wir n groß genug wählen.

Beispiel 3.8.7 *Für $\lambda = 1$ für die Poissonverteilung und $n = 100$ und*

$$p = \frac{1}{100}$$

für die Binomialverteilung erhalten wir

k	0	1	2	3	4
$P(X = k)$	0.36787	0.36787	0.18394	0.061313	0.015328
$P(X_n = k)$	0.36603	0.36972	0.18486	0.060999	0.014941

Beispiel 3.8.8 *In einer Blutprobe von 1 μl (d.h. 10^{-6} Liter) finden wir im Mittel 5 weiße Blutkörperchen. Was ist die Wahrscheinlichkeit, dass wir in einer solchen Blutprobe k weiße Blutkörperchen finden? Die Zufallsvariable X gebe diese Anzahl an. Mit der Poissonverteilung erhalten wir*

$$P(X = k) = \frac{5^k}{k!} \exp(-5).$$

Für die Verteilung siehe Abbildung 3.8. Eine Messung in $\{4, 5, 6\}$ tritt also nur mit Wahrscheinlichkeit $\approx 49.7\%$ auf, eine Messung mit Abweichung ≥ 2 mit Wahrscheinlichkeit $\approx 51.3\%$. Wir können die Messung mehrfach durchführen und erhalten gemäß dem Gesetz der großen Zahlen mit hoher Wahrscheinlichkeit eine kleinere Abweichung vom realen Wert.

Natürlich könnten wir das Problem auch mit der Binomialverteilung modellieren. Dazu brauchen wir allerdings zusätzlich die Wahrscheinlichkeit, dass ein weisses Blutkörperchen sich in der Probe befindet. Sei n die Anzahl der weißen Blutkörperchen in einem Menschen. Der Durchschnittsmensch hat insgesamt etwa 6 Liter Blut. Die Wahrscheinlichkeit, dass ein bestimmtes weißes Blutkörperchen sich in der Probe befindet, ist also

$$p = \frac{1}{6\ 000\ 000}$$

und der Mensch hat damit etwa

$$n = 30\ 000\ 000 = 5 \cdot 6\ 000\ 000$$

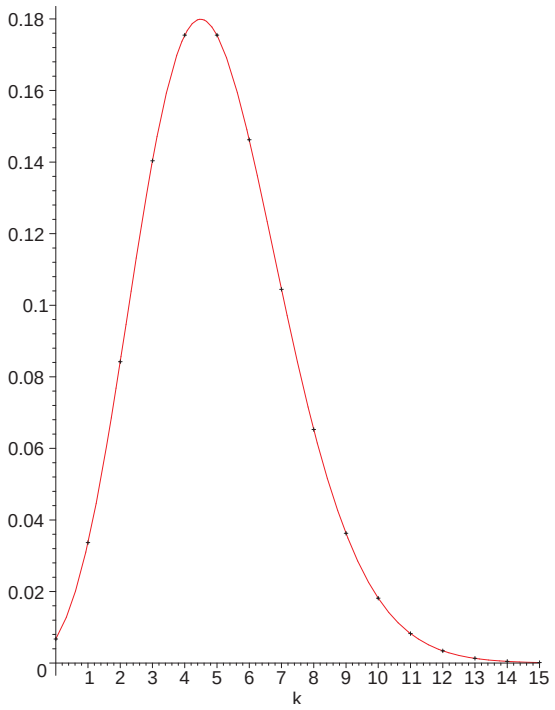


Abbildung 3.8: Poissonverteilung für $\lambda = 5$ und eine Interpolation der diskreten Verteilung durch eine stetige Funktion.

weiße Blutkörperchen. Abbildung 3.9 zeigt zusätzlich zu Abbildung 3.8 die Binomialnäherung. Ist p nicht bekannt (in der Praxis ist das tatsächlich so, da wir für einen spezifischen Menschen nicht genau die Blutmenge bestimmen können), dann ist die Binomialverteilung nicht verwendbar, die Poissonverteilung aber schon.

3.8.3 Normalverteilung

In Aufgabe 3.3 haben wir schon eine Idee für das Verhalten des Mittelwerts von 5 kontinuierlichen gleichverteilten Zufallsvariablen entwickelt. Es zeigte sich eine glockenförmige Wahrscheinlichkeitsverteilung. Wir könnten z.B. nach der Wahrscheinlichkeitsverteilung der Größe von Menschen fragen. Oft wird behauptet, dass die Wahrscheinlichkeitsverteilung der Größe von Menschen eine glockenförmige Verteilung besitzt. Dies ist auch

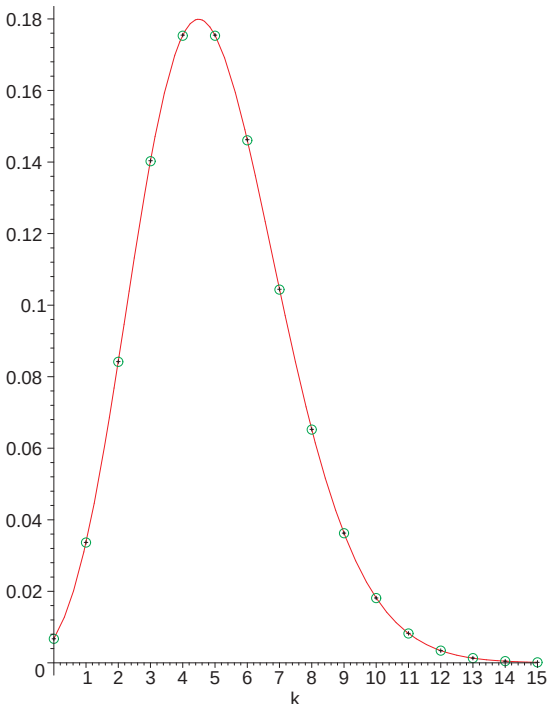


Abbildung 3.9: Poissonverteilung (schwarze Punkte) mit Binomialnäherung (grüne Kreise) und Interpolation der diskreten Werte durch eine stetige Funktion (rote Kurve).

empirisch der Fall, eine eindeutige Erklärung dafür gibt es aber nicht. Die wohl plausibelste ist, dass die Größe eines Menschen von einer Vielzahl von meist unabhängigen genetischen Faktoren bestimmt wird und jeder Faktor einen gewissen Beitrag zu der Körpergröße leistet. Wir werden sehen, dass unter einer solchen Annahme tatsächlich eine glockenförmige Verteilung zu erwarten ist. Erproben wir dies zunächst an einem Beispiel:

Beispiel 3.8.9 *Wir nehmen an, dass es* $G = 10$ *genetische Faktoren gibt, die jeweils gleichverteilt zu der Körpergröße eines erwachsenen Menschen zwischen* 0 *cm und* 10 *cm beitragen und, dass fast alle Menschen größer als* 125 *cm und kleiner als* 225 *cm sind. Ausgehend von einer Mindestgröße von* 125 *cm, liefert dies also Werte zwischen* 125 *cm und* 225 *cm. In* MAPLE *berechnen wir* $N = 20000$ *mal die Summe der Zufallsvariablen:*

```
N:=20000:
G:=10:
L:=[]:
for j from 1 to N do
    R:=10*[stats[random,uniform](G)]:
    l:=sum(R[i],i=1..G);
    L:=[op(L),sum(R[i],i=1..G)];
od:
```

Wir sortieren dann die Ergebnisse in die Intervalle

$$[125, 126[, \ldots, [223, 224[, [224, 225]$$

ein:

```
H:=[seq(0,jj=1..100)]:
for j from 1 to nops(H) do
    for k from 1 to N do
        if j-1<=L[k] and L[k]<j then
            H[j]:=H[j]+1;
        fi;
    od:
od:
H;
```

```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1,
3, 1, 2, 4, 4, 11, 15, 31, 39, 43, 62, 65, 79, 101, 161,
166, 208, 238, 299, 361, 413, 447, 518, 539, 654, 692,
712, 745, 828, 810, 854, 867, 873, 872, 808, 848, 753,
688, 675, 630, 564, 517, 507, 373, 370, 290, 227, 234,
168, 155, 108, 92, 72, 59, 40, 29, 20, 23, 8, 4, 5, 5,
1, 2, 3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0]
```

Abbildung 3.10 zeigt die Verteilung der Körpergrößen. Diese Darstellung lässt sich mit dem folgenden Code erzeugen:

```
with(plots):
pointplot([seq([j+125,H[j]],j=1..nops(H))]);
```

Ebenso zeigt die Binomialverteilung mit wachsendem n eine solche Verteilung.

Beispiel 3.8.10 *Für den $N = 100$ fachen Münzwurf mit Ergebnis 0 oder 1 sei X die Summe der Ergebnisse. Wir erhalten für*

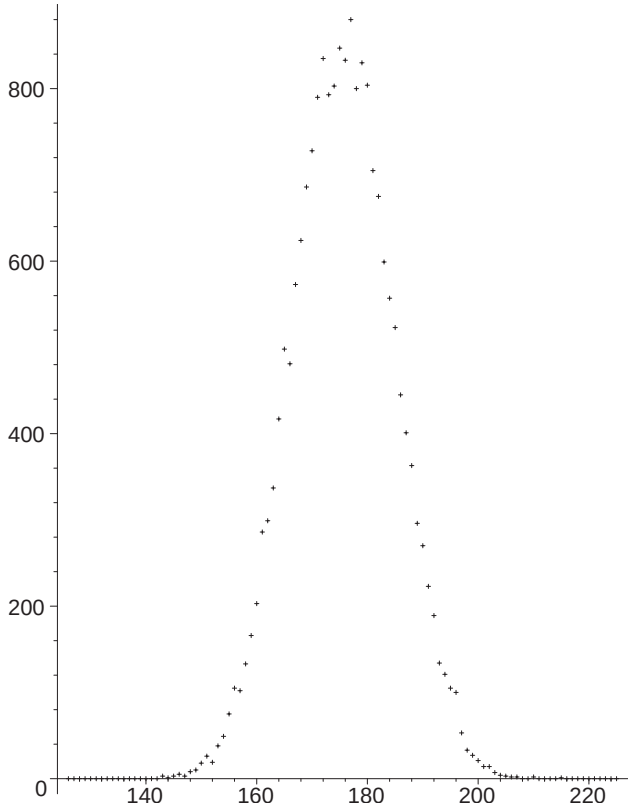


Abbildung 3.10: Häufigkeiten der Körpergrößen

X die Binomialverteilung in Abbildung 3.11. Diese Verteilung können wir in MAPLE *erzeugen mit:*

```
N:=100:
H:=[seq(binomial(N,j)*1/2^N,j=0..N)];
```

Andere Beispiele, in denen man ähnliche Wahrscheinlichkeitsdichten beobachtet sind die Verteilung von Messfehlern, die Verteilung von Abweichungen von Bauteilen vom gewünschten Maß oder die Beschreibung der Brownschen Bewegung von Gasmolekülen. Es ist zunächst ziemlich erstaunlich, dass diese Glockenkurve in so vielen Situationen auftritt. Den Grund hierfür liefert der Zentrale Grenzwertsatz, den wir im folgenden Abschnitt diskutieren.

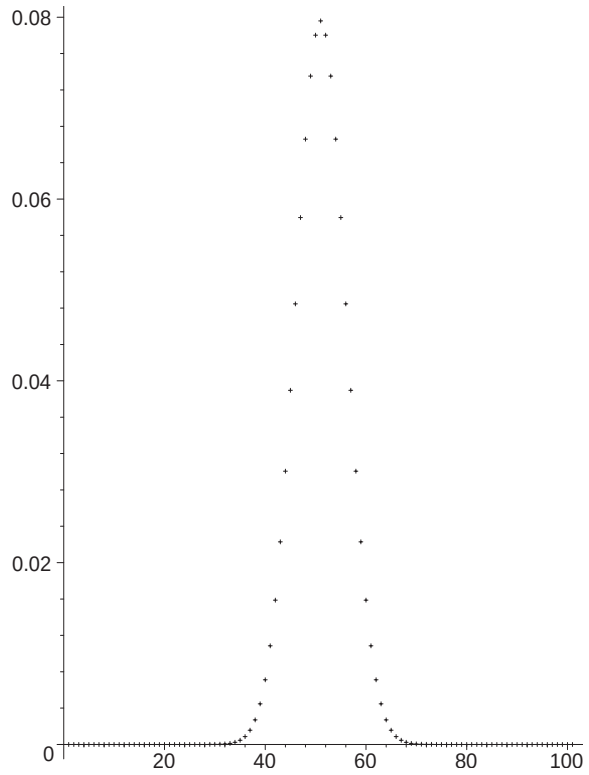


Abbildung 3.11: Binomialverteilung der Häufigkeit von Kopf beim N-fachen Münzwurf für großes N.

3.8.4 Zentraler Grenzwertsatz

Während das Gesetz der großen Zahlen uns etwas über den Erwartungswert eines Mittelwerts

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$$

von unabhänigig identisch verteilten Zufallsvariablen sagt, liefert der Zentrale Grenzwertsatz weit mehr Information. Er beschreibt die Verteilung der Mittelwerte von Stichproben um den Erwartungswert in Termen einer Wahrscheinlichkeitsverteilung. Im Gegensatz zur Hoeffding-Ungleichung gibt er nicht nur eine Schranke für Wahrscheinlichkeiten, sondern die exakte Verteilung im Grenzwert $n \to \infty$. Jedoch können wir damit eben

nur Wahrscheinlichkeiten im Grenzwert untersuchen und erhalten keine Ergebnisse für endliches n, was bei Machine-Learning-Anwendungen wichtig ist, da wir ja n so wählen wollen, dass für eine Mehrfachmessung eine bestimmte Fehlerwahrscheinlichkeit unter einer bestimmten Schranke liegt.

Satz 3.8.11 *Seien $X_1, \ldots, X_n$ identisch unabhängig verteilte Zufallsvariablen, für die die Varianz existiere. Wir schreiben*

$$\mu = E(X_i)$$
$$\sigma = \sigma(X_i).$$

Dann gilt für alle

$$-\infty \leq a < b \leq \infty,$$

dass

$$\lim_{n\to\infty} P\left(a \leq \frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \ldots + X_n}{n} - \mu\right) \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{x^2}{2})dx$$

Definition 3.8.12 *Die **Normalverteilung** auf $\Omega = \mathbb{R}$ mit Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{>0}$ ist gegeben durch die Wahrscheinlichkeitsdichte*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

*Die **Standardnormalverteilung** ist gegeben durch*

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}).$$

Für die Dichte der Standardnormalverteilung siehe Abbildung 3.12. Der Graph in Abbildung 3.13 zeigt die kummulative Wahrscheinlichkeit

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})dt.$$

Bemerkung 3.8.13 *Der Zentrale Grenzwertsatz sagt, dass die Verteilung von*

$$\frac{\sqrt{n}}{\sigma} \frac{X_1 + \ldots + X_n}{n}$$

im Grenzwert $n \to \infty$ einer Standardnormalverteilung gehorcht.

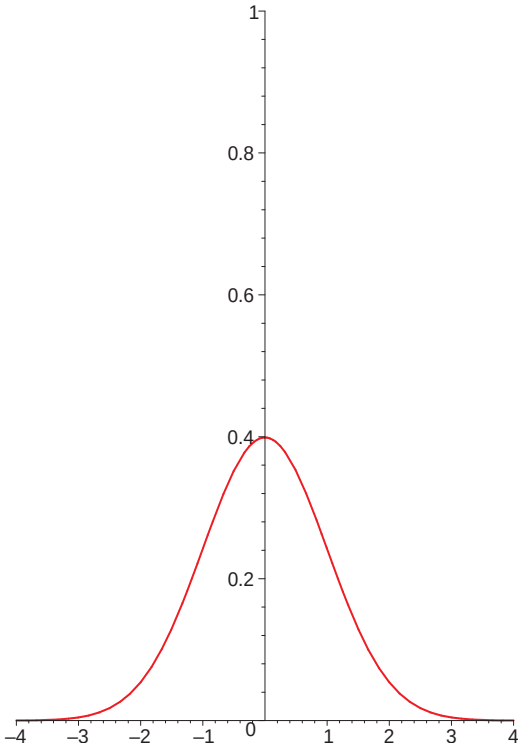


Abbildung 3.12: Dichte der Standardnormalverteilung.

Bevor wir den zentralen Grenzwertsatz beweisen, untersuchen wir die Eigenschaften der Normalverteilung. Die erste wichtige Beobachtung ist, dass die Normalverteilung tatsächlich eine Wahrscheinlichkeitsdichte gibt:

Satz 3.8.14 *Ist f die Wahrscheinlichkeitsdichte einer Normalverteilung, so gilt*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Beweis. Der Einfachheit halber beschränken wir uns auf den Fall der Standardnormalverteilung: Da die Dichte symmetrisch in x ist, reicht es zu zeigen, dass

$$\int_0^{\infty} \exp(-\frac{x^2}{2})dx = \sqrt{\frac{\pi}{2}}.$$

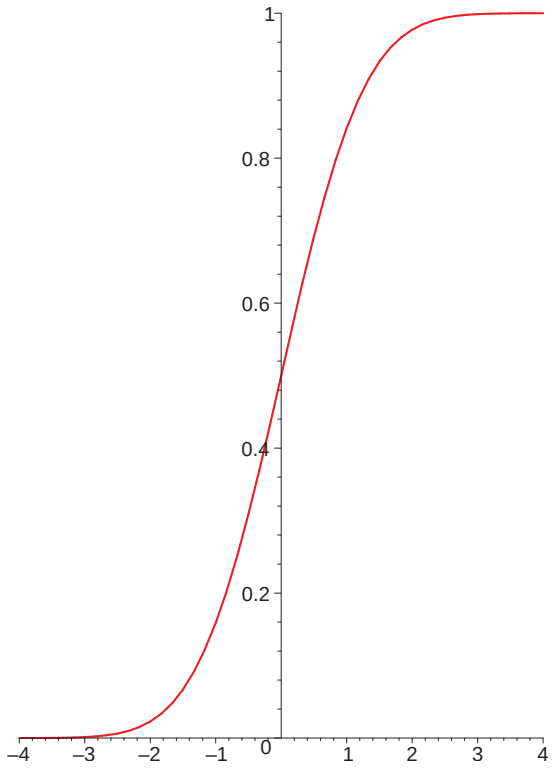


Abbildung 3.13: $P(X \leq x)$

Dazu betrachten wir das Quadrat und verwenden die Funktionalgleichung der Exponentialfunktion

$$\left(\int_0^\infty \exp(-\frac{x^2}{2})dx\right)^2 = \left(\int_0^\infty \exp(-\frac{x^2}{2})dx\right) \cdot \left(\int_0^\infty \exp(-\frac{y^2}{2})dy\right)^2$$
$$= \int_0^\infty \int_0^\infty \exp(-\frac{x^2+y^2}{2})dxdy$$

Der Integrand ist offensichtlich konstant auf Kreisen um den Ursprung. Deshalb wechseln wir in die sogenannten **Polarkoordinaten**

$$x = r \cdot \cos\varphi$$
$$y = r \cdot \sin\varphi$$

und müssen dann über den Radius $r = 0, \ldots, \infty$ und den Winkel $\varphi = 0, \ldots, \pi/2$ integrieren. Wie transformiert sich aber die Integration in die Polarkoordinaten? Bei einer Diskreditierung des

Riemannintegrals würden wir die x- als auch die y-Koordinate äquidistant unterteilen, ebenso die r und die φ Koordinate. Das Volumen eines solchen Diskretisierungselements im (r, φ)-Koordinatensystem nimmt allerdings mit wachsendem Radius r linear mit r zu, da der Umfang eines Kreises mit $2\pi r$ linear mit r wächst, siehe Abbildung 3.14.[3] Deshalb erhalten wir

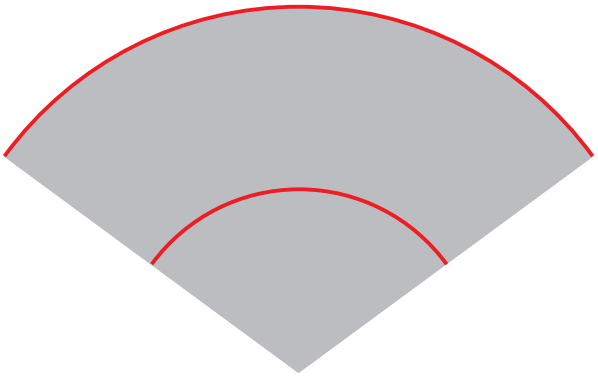


Abbildung 3.14: Bogenlänge in Polarkoordinaten für konstanten Winkel in Abhänigigkeit vom Radius..

$$\begin{aligned}\int_0^\infty \int_0^\infty \exp(-\frac{x^2+y^2}{2})dxdy &= \int_0^\infty \int_0^{\pi/2} \exp(-\frac{r^2}{2}) \cdot r \cdot drd\varphi \\ &= \frac{\pi}{2} \int_0^\infty \exp(-\frac{r^2}{2}) \cdot r \cdot dr \\ &= \frac{\pi}{2}\left[-\exp(-\frac{r^2}{2})\right]_0^\infty = \frac{\pi}{2}.\end{aligned}$$

■

Wir bestimmen noch den Erwartungswert und die Varianz:

Satz 3.8.15 *Ist hat X als Verteilung die Standardnormalverteilung, dann gilt*

$$\begin{aligned}E(X) &= 0 \\ V(X) &= 1.\end{aligned}$$

[3] Allgemeinen gilt als mehrdimensionale Verallgemeinerung der Substitutionsregel der Transformationssatz. Siehe dazu Satz 5.2.

Beweis. Da die Dichte der Normalverteilung symmetrisch zu $x = 0$ ist, haben wir

$$\begin{aligned}\int_{-\infty}^{\infty} x\exp(-\frac{x^2}{2})dx &= \int_0^{\infty} x\exp(-\frac{x^2}{2})dx + \int_{-\infty}^{0} x\exp(-\frac{x^2}{2})dx\\ &= \int_0^{\infty} x\exp(-\frac{x^2}{2})dx - \int_0^{\infty} x\exp(-\frac{x^2}{2})dx\\ &= 0.\end{aligned}$$

und damit $E(X) = 0$. Mit partieller Integration erhalten wir

$$\left[-x\exp(-\frac{x^2}{2})\right] = \int x^2\exp(-\frac{x^2}{2})dx - \int_{-\infty}^{\infty}\exp(-\frac{x^2}{2})dx$$

und damit für die Varianz

$$\begin{aligned}V(X) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2\exp(-\frac{x^2}{2})dx\\ &= \frac{1}{\sqrt{2\pi}}\left[-x\exp(-\frac{x^2}{2})\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\frac{x^2}{2})dx\\ &= 0 + 1 = 1\end{aligned}$$

wobei der erste Summand verschwindet, da $\exp(\frac{x^2}{2})$ für $x \to \pm\infty$ schneller wächst als x (wie man mit der Regel von l'Hospital sieht) und wir für den zweiten Summanden Satz 3.8.14 verwenden. ■

Satz 3.8.16 *Ist $\mu \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}_{\geq 0}$ und X standardnormalverteilt, dann ist*

$$Z = \sigma X + \mu$$

normalverteilt mit Parametern μ und σ^2.

Beweis. Wir haben

$$\begin{aligned}P(Z \leq a) &= P(X \leq \tfrac{a-\mu}{\sigma})\\ &= \int_{-\infty}^{\frac{a-\mu}{\sigma}}\exp(-\tfrac{x^2}{2})dx\\ &= \int_{-\infty}^{a}\tfrac{1}{\sqrt{2\pi}\sigma}\exp(-\tfrac{(x-\mu)^2}{2\sigma^2})dx\end{aligned}$$

mit der Substitutionsregel. ■

Satz 3.8.17 *Hat Z die Normalverteilung mit Parametern μ und σ, so ist*

$$E(Z) = \mu$$
$$V(Z) = \sigma^2.$$

Beweis. Da wir die Dichte von Z schreiben können als

$$f_Z(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$
$$= \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$$

folgt die Behauptung mit der Substitutionsregel mit den entsprechenden Integralen für die Standardnormalverteilung (Übung). ■

Beispiel 3.8.18 *Für die Normalverteilung erhalten wir*

$$\begin{array}{lcll} P(|X-\mu| \le \sigma) & = & P(|X| \le 1) & \approx & 0.6826 \\ P(|X-\mu| \le 2\sigma) & = & P(|X| \le 2) & \approx & 0.9544 \\ P(|X-\mu| \le 3\sigma) & = & P(|X| \le 3) & \approx & 0.9974 \end{array}$$

wie wir z.B. wie folgt mit MAPLE *sehen:*

```
int(1/sqrt(2*Pi)*exp(-x^2/2) , x=-1..1);
.6826894920
int(1/sqrt(2*Pi)*exp(-x^2/2) , x=-2..2);
.9544997360
int(1/sqrt(2*Pi)*exp(-x^2/2) , x=-3..3);
.9973002039
```

Bemerkung 3.8.19 *Der Zentrale Grenzwertsatz besagt also, dass die Verteilung des Mittelwerts*

$$\frac{X_1 + \ldots + X_n}{n}$$

für große n einer Normalverteilung mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$ gehorcht. Äquivalent gehorcht die Summe

$$X_1 + \ldots + X_n$$

einer Normalverteilung mit Erwartungswert $n \cdot \mu$ und Varianz $n \cdot \sigma^2$.

Beispiel 3.8.20 *In einer Studie haben wir die Körpergrößen der (männlichen) Teilnehmer gemessen, den Mittelwert und die Standardabweichung der Körpergrößen berechnet, und*

$$\mu = 178 \ cm$$
$$\sigma = 7.35 \ cm$$

erhalten. Wie man aus statistischen Daten die Standardabweichung schätzt werden wir im nächsten Kapitel diskutieren. Mit diesen statistisch bestimmten Daten können wir z.B. mit MAPLE *berechnen, dass*

$$\int_{-\infty}^{164} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \approx 0.028$$

d.h. 2.8% *der Männer sind kleiner als* 164cm.

Der Beweis von Satz 3.8.11 verwendet wie der Beweis der Hoeffding-Ungleichung Erzeugendenfunktionen für Momente. Wir geben den Beweis in Abschnitt 5.3.

3.9 Übungsaufgaben

Übung 3.1 *Basteln Sie eine rotierende Scheibe mit einem feststehenden Zeiger. Unterteilen Sie die Scheibe in* 3 *Tortenstücke, die den Umfang im Verhältnis* 3 : 2 : 1 *unterteilen und blau, gelb bzw. rot markiert sind, siehe Abbildung 3.15. Drehen Sie die Scheibe* 100*-mal und bestimmen Sie die relativen Häufigkeiten von rot, gelb und blau.*

Übung 3.2 *In das Quadrat* $[0,2]^2$ *sei ein Kreis mit Radius* 1 *einbeschrieben (siehe Abbildung 3.16)*

1) *Erzeugen Sie zufällig gleichverteilt* 100 *Punkte in dem Quadrat, indem Sie beide Koordinaten gleichverteilt in* $[0,2]$ *wählen.*

2) *Messen Sie die relative Häufigkeit, mit der ein Punkt in dem einbeschriebenen Kreis liegt. Welche Zahl approximiert Ihre Rechnung?*

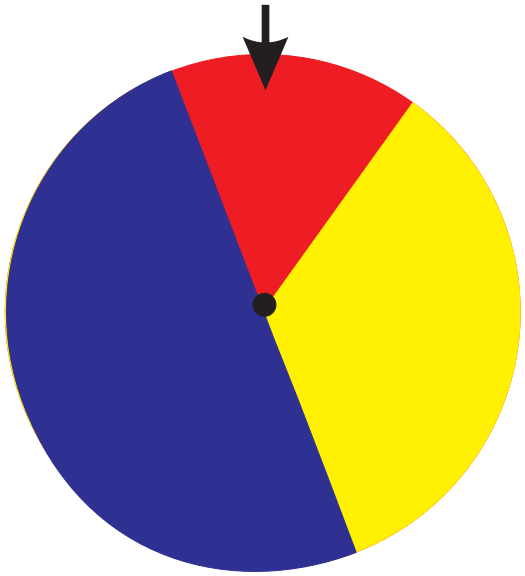


Abbildung 3.15: Rotierende Scheibe mit Zeiger und Unterteilung des Umfangs im Verhältnis $3:2:1$.

Hinweis: Die MAPLE*-Funktion* `stats[random, uniform](n)` *liefert n zufällige Zahlen in $[0,1]$, wobei alle Elemente von $[0,1]$ gleich wahrscheinlich sind.*

Übung 3.3 *1) Schreiben Sie ein Programm, das gleichverteilt 5 zufällige Zahlen $x_1,\ldots,x_5$ im Intervall $[0,1]$ bestimmt und deren Mittelwert*

$$S = \frac{x_1 + \ldots + x_5}{5}$$

bildet.

2) Führen Sie Ihr Programm 10000-mal aus und bestimmen Sie die relativen Häufigkeiten, dass S in den Intervallen

$$\left[0, \frac{1}{100}\right[, \ldots, \left[\frac{98}{100}, \frac{99}{100}\right[, \left[\frac{99}{100}, 1\right]$$

liegt. Erstellen Sie ein Diagramm mit den relativen Häufigkeiten.

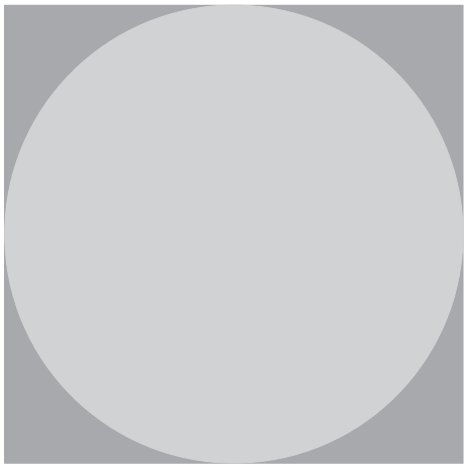


Abbildung 3.16: In ein Quadrat einbeschriebener Kreis.

Übung 3.4 *1) Die Wahrscheinlichkeit, dass eine Bremsleuchte Ihres Autos nach $t \geq 0$ Betriebsstunden defekt wird, ist beschrieben durch die Wahrscheinlichkeitsdichte*

$$p(t) = \lambda \cdot \exp(-\lambda t)$$

mit $\lambda > 0$.

2) Bestimmen Sie für $\lambda = \frac{1}{1000}$ die Wahrscheinlichkeit

$$\int_0^{t_0} p(t)dt$$

dass die Bremsleuchte in den nächsten t_0 Betriebsstunden defekt wird.

3) Was ist die Defektwahrscheinlichkeit in den nächsten 10, 100 und 1000 Stunden?

Übung 3.5 *1) Nehmen Sie eine Nähnadel und zeichen Sie auf einem großen Blatt Papier parallele Geraden, deren Abstand genau die Länge der Nadel ist.*

2) Werfen Sie die Nadel 100-mal auf das Papier. Bestimmen Sie für Ihren Wurf die relative Häufigkeit, dass eine der Geraden schneidet (siehe Abbildung 3.4).

3) Implementieren Sie das Nadelexperiment, ohne in Ihrem Programm die Zahl π oder trigonometrische Funktionen zu verwenden.

4) Führen Sie das Experiment 10000-mal durch und bestimmen Sie die relative Häufigkeit p, dass eine der Geraden schneidet, und berechnen Sie $\frac{2}{p}$.

Übung 3.6 *1) Entwickeln Sie einen expliziten Algorithmus zur Monte-Carlo-Integration einer Funktion*

$$f : [a, b] \to \mathbb{R}_{\geq 0}$$

wobei wir annehmen, dass $a, b \in \mathbb{Q}$.

2) Testen Sie Ihren Algorithmus an Polynomfunktionen. Vergleichen Sie mit der expliziten Auswertung des Integrals mittels einer Stammfunktion.

3) Berechnen Sie mit Ihrem Algorithmus

$$\int_0^1 \exp(x)dx.$$

Übung 3.7 *Sei $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{>0}$.*

1) Zeigen Sie, dass für die Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

der Normalverteilung gilt:

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

2) Sei Z eine kontinuierliche Zufallsvariable mit Wahrscheinlichkeitsdichte f. Zeigen Sie, dass

$$E(Z) = \mu$$
$$V(Z) = \sigma^2.$$

Hinweis: Verwenden Sie die entsprechenden Resultate für die Standardnormalverteilung.

4

Anwendungen aus der Statistik

4.1 Übersicht

Die Statistik ist ein zur Stochastik eng verwandtes Gebiet und befasst sich mit dem Sammeln und Analysieren von Daten. Während wir in der Stochastik untersuchen, was wir über die Eigenschaften des Resultats eines vorgegebenen datenerzeugenden Prozesses sagen können, ist die Fragestellung der Statistik das dazu inverse Problem: Gegeben eine Menge an Daten, ist die Frage, was wir über den Prozess sagen können, der diese Daten erzeugt. Das Gegenstück zur Statistik ist in der Informatik ist das Data Mining und das Machine Learning.

Wie wir in Abschnitt 2.13.1 und 3.7.1 gesehen haben, können wir durch Bilden eines Mittelwerts einer Messreihe einen Erwartungswert approximieren: Das Gesetz der großen Zahlen besagt, dass die Wahrscheinlichkeit einer großen Abweichung des (arithmetischen) Mittelwerts von dem Erwartungswert bei einer mehrfachen Durchführung eines Zufallsprozesses gegen 0 geht.

4.2 Statistische Größen aus stochastischen Größen

Die Begriffsbildung für die Statistik können wir mit der folgenden Beobachtung aus der Stochastik ableiten: Wie wir schon in

Beispiel 2.6.9 diskutiert, können wir jede Messreihe wieder als ein Zufallsexperiment auffassen. Sind etwa

$$x_1, \ldots, x_n$$

Messwerte, dann nehmen wir

$$\Omega = \{1, \ldots, n\}$$

mit der gleichvertielten Wahrscheinlichkeitsfunktion

$$m(i) = \frac{1}{n}$$

und definieren die Zufallsvariable $X : \Omega \to \mathbb{R}$ durch

$$X(i) = x_i$$

Als Erwartungswert erhalten wir dann z.B. den Mittelwert

$$\begin{aligned} E(X) &= \sum_i X(i)m(i) \\ &= \frac{1}{n}\sum_i x_i. \end{aligned}$$

Ebenso erhalten wir höhere Stichprobenmomente

$$\begin{aligned} E(X^k) &= \sum_i X(i)^k m(i) \\ &= \frac{1}{n}\sum_i x_i^k. \end{aligned}$$

eine Notation von Varianz als

$$V(X) = E((X - E(X))^2)$$

und der Standardabweichung

$$\sigma(X) = \sqrt{V(X)}.$$

Genauso übertragen sich die Definitionen der Covarianz und die Korrelation. Wenn wir für einen Datensatz von diesen Größen sprechen, dann meinen wir stets die entsprechenden Größen der zugeordneten Zufallsvariable X.

Beispiel 4.2.1 *Für die Messwerte*

$$6, 3, 5, 3, 1, 1, 3, 4, 2, 1$$

haben wir

i	1	2	3	4	5	6	7	8	9	10
$X(i)$	6	3	5	3	1	1	3	4	2	1

Als Erwartungswert erhalten wir dann z.B. den Mittelwert

$$E(X) = \frac{1}{10}(6+3+5+3+1+1+3+4+2+1) = \frac{29}{10}$$

der Messwerte.

Die so hergeleiteten Formel für eine Approximation des Erwartungswerts ist ein Beispiel eines konsistenten Schätzers. Bei den Formeln für die Varianz, Standardabweichung, Covarianz und Korrelation gibt es allerdings noch ein Detail zu beachten.

Definition 4.2.2 *Seien $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ identisch unabhängig verteilte Zufallsvariablen, deren Verteilung von einem Parameter $\mu \in \mathbb{R}$ abhängt. Ein **konsistenter Schätzer** ist eine Abbildung $F : \mathbb{R}^n \to \mathbb{R}$ mit*

$$E(F(X_1, \ldots, X_n)) = \mu.$$

Bemerkung 4.2.3 *In der Praxis wird ein Schätzer dann auf eine Stichprobe $x_1, \ldots, x_n$ angewandt, d.h.*

$$F(x_1, \ldots, x_n)$$

berechnet.

Beispiel 4.2.4 *Der Mittelwert*

$$M(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

ist ein konsistenter Schätzer für den Erwartungswert

$$\mu = E(X_i),$$

da mit Satz 2.6.12

$$E(M(X_1, \ldots, X_n)) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

gilt.

Beispiel 4.2.5 *Bei der Varianz passiert etwas Unerwartetes: Die Formel*

$$V(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M(x_1,\dots,x_n)\right)^2$$

ist kein konsistenter Schätzer für die Varianz, denn für identisch unabhängig verteilte Zufallsvariablen $X_1,\dots,X_n$ gilt

$$\begin{aligned}
E\Big(\sum_{i=1}^{n}(X_i - M)^2\Big) &= \sum_{j=1}^{n} E(X_j^2) - 2\sum_{j} E(X_j M) + n\cdot E\left(M^2\right)\\
&= n\cdot E(X_i^2) - \frac{2}{n}\sum_{j,l} E(X_j X_l) + \frac{1}{n}\cdot\sum_{j,l} E\left(X_j X_l\right)\\
&= n\cdot E(X_i^2) - \frac{2}{n}\sum_{j\neq l} E(X_j X_l) + \frac{1}{n}\cdot\sum_{j\neq l} E\left(X_j X_l\right)\\
&\quad - \frac{2}{n}\sum_{j} E(X_j^2) + \frac{1}{n}\cdot\sum_{j} E\left(X_j^2\right)\\
&= n\cdot E(X_i^2) - \frac{1}{n}\sum_{j\neq l} E(X_j X_l) - E\left(X_i^2\right)\\
&= (n-1)\cdot E(X_i^2) - \frac{n(n-1)}{n} E(X_i)^2\\
&= (n-1)\cdot V(X_i)
\end{aligned}$$

(für alle i). Aus unserer Rechnung erhalten wir dagegen:

Definition und Satz 4.2.6 *Die* ***Stichprobenvarianz***

$$V(x_1,\dots,x_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - M(x_1,\dots,x_n)\right)^2$$

mit

$$M(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

ist ein konsistenter Schätzer für die Varianz. Die ***Stichprobenstandardabweichung*** *ist definiert als*

$$\sigma = \sqrt{V(X)}.$$

Bemerkung 4.2.7 *Die Formel mit Nenner n kann dennoch in Situtationen verwendet werden, in denen man den große n oder den Grenzwert $n\to\infty$ betrachtet, denn*

$$\lim_{n\to\infty}\frac{n-1}{n} = 1.$$

Beispiel 4.2.8 *Für die Messwerte in Beispiel 4.2.1 ist*

$$\begin{aligned} V(x_1,\ldots,x_n) &= \frac{1}{9}((6-\frac{39}{10})^2+(3-\frac{39}{10})^2+(5-\frac{39}{10})^2+(3-\frac{39}{10})^2 \\ &+(1-\frac{39}{10})^2+(1-\frac{39}{10})^2+(3-\frac{39}{10})^2+(4-\frac{39}{10})^2 \\ &+(2-\frac{39}{10})^2+(1-\frac{39}{10})^2) \\ &= \frac{41}{10} \end{aligned}$$

und

$$\sigma(x_1,\ldots,x_n) = \sqrt{\frac{41}{10}} \approx 2.02.$$

4.3 Konfidenzintervall für den Erwartungswert

Nach Satz 5.3.8 ist der Mittelwert

$$M = \frac{X_1+\ldots+X_n}{n}$$

von unabhängig identisch normalverteilten Zufallsvariablen mit Parametern μ und σ^2 eine Normalverteilung mit Parametern μ und $\frac{\sigma^2}{n}$. Der Mittelwert M schätzt den Erwartungswert μ. Wir suchen eine Zahl $\varepsilon > 0$ sodass für vorgegebene Wahrscheinlichkeit

$$0 \leq \gamma \leq 1$$

gilt

$$P(|M-\mu| \leq \varepsilon) < \gamma.$$

Dazu bestimmen wir für eine Stichprobe $x_1,\ldots,x_n$ den Mittelwert

$$m = M(x_1,\ldots,x_n)$$

und die Stichprobenstandardabweichung

$$s = \sigma(x_1,\ldots,x_n)$$

und setzen

$$v_1 = m - \frac{\varepsilon \cdot s}{\sqrt{n}}$$
$$v_2 = m + \frac{\varepsilon \cdot s}{\sqrt{n}}$$

Es gilt dann

$$P(v_1 \le M \le v_2) = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} \exp(-\frac{x^2}{2}) dx$$

Anhand einer Tabelle für die Standardnormalverteilung oder auch mit Maple können wir dann ε mit

$$\frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} \exp(-\frac{x^2}{2}) dx = \gamma$$

und damit das **Konfidenzintervall zur Konfidenz** γ

$$[v_1, \ldots, v_2] = \left[m - \frac{\varepsilon \cdot s}{\sqrt{n}}, m + \frac{\varepsilon \cdot s}{\sqrt{n}} \right]$$

bestimmen.

Beispiel 4.3.1 *Wir wollen den mittleren Schaden in € schätzen, den ein Wildschwein bei Kollision mit einem Auto verursacht. Die Versicherung hat Aufzeichnungen über* 10 *solchen Unfällen mit einem Schaden von jeweils*

$$5500\ e, 5750\ e, 5000\ e, 4300\ e, 2100\ e,$$
$$8000\ e, 5750\ e, 4100\ e, 7000\ e, 3000\ e.$$

Wir bestimmen den Mittelwert

$$\begin{aligned} m &= \frac{1}{10}(5500 + 5750 + 5000 + 4300 + 2100 + 8000 + 5750 + 4100 + 7000 + 3000) \\ &= 5050 \end{aligned}$$

und die Stichprobenvarianz

$$\begin{aligned} s^2 = \frac{1}{9}&((5500 - 5050)^2 + (5750 - 5050)^2 + (5000 - 5050)^2 + (4300 - 5050)^2 \\ &+ (2100 - 5050)^2 + (8000 - 5050)^2 + (5750 - 5050)^2 + (4100 - 5050)^2 \\ &+ (7000 - 5050)^2 + (3000 - 5050)^2) \\ &= 3117800 \end{aligned}$$

und damit

$$s = \sqrt{3117800} \approx 1765$$

Ein 95% *Konfidenzintervall erhalten wir dann als*

$$v_1 = 5050 - \frac{1765\varepsilon}{\sqrt{10}}$$
$$v_2 = 5050 + \frac{1765\varepsilon}{\sqrt{10}}$$

die Gleichung

$$\frac{1}{\sqrt{2\pi}} \int_{\varepsilon}^{\varepsilon} \exp(-\frac{x^2}{2}) dx = 0.95$$

lösen. Dies können wir z.B. in MAPLE *mit*

```
L:=int(1/sqrt(2*Pi)*exp(-x^2/2),x=-epsilon..epsilon);
solve(L=0.95,epsilon);
```

```
1.96
```

Somit ist das Konfidenzintervall

$$\left[5050 - \frac{1765 \cdot 1.96}{\sqrt{10}}, 5050 + \frac{1765 \cdot 1.96}{\sqrt{10}}\right] = [3956, 6144]$$

Bemerkung 4.3.2 *Ist die Varianz nicht bekannt, dann kann man die Normalverteilung durch die sogenannte t-Verteilung ersetzen.*

4.4 Lineare Regression

Im Gesetz der großen Zahlen haben wir gesehen, dass der Mittelwert von identisch unabhängig verteilten Messwerten den Erwartungswert annähert. Oft hängen die Messwerte aber von einem Parameter, etwa der Zeit ab. Ein typisches Problem ist es dann, in einer Klasse von Funktionen eine Funktion zu finden, deren Funktionsgraph eine gegebene Datenmenge am besten beschreibt. Wir wollen hier nur den Fall einer linearen Funktion betrachten. Gegeben seien Datenpunkte

$$(x_1, y_1), \ldots, (x_n, y_n)$$

Unser Ziel ist es aus der Klasse der linearen Polynomfunktionen ein Element zu finden, das die gegebenen Datenpunkte bestmöglich approximiert, d.h. wir wollen e

$$f(x) = a \cdot x + b \in \mathbb{R}[x]$$

finden mit

$$f(x_i) = y_i$$

für alle i. Dies können wir im Allgemeinen aber nur für $n = 2$ erwarten, denn zwei Punkte in der Ebene legen eine eindeutige Gerade fest, die wir mittels Interpolation finden können (alternativ könnten wir auch das entsprechende lineare Gleichungssystem lösen). Für $n = 1$ gäbe es eine ganze Schar von möglichen Geraden durch den Messpunkt. Typischerweise wird n aber wesentlich größer sein. Da bei Messwerten immer Messfehler auftreten (sagen wir wir bestimmen die Position eines Autos auf einer Straße in Abhängigkeit der Zeit), können wir selbst bei einem exakten linearen Zusammenhang von x und y nicht erwarten, dass auch unsere Messwerte einen solche Relation erfüllen.

Das übliche Verfahren ist es, die Quadrate der Abweichungen zu minimieren, d.h. wir minimieren

$$F(a, b) = \sum_{i=1}^{n} (a \cdot x_i + b - y_i)^2$$

in Abhängigkeit von a und b. Um ein lokales Minimum zu finden, suchen wir die Nullstellen der Ableitungen nach a und b. Die Ableitungen sind

$$\frac{\partial F}{\partial a} = \sum_{i=1}^{n} 2 \cdot (a \cdot x_i + b - y_i) \cdot x_i$$
$$\frac{\partial F}{\partial b} = \sum_{i=1}^{n} (a \cdot x_i + b - y_i)$$

also erhalten wir die lineare Gleichungssystem

$$\left(\sum_{i=1}^{n} x_i\right) \cdot b + \left(\sum_{i=1}^{n} x_i^2\right) \cdot a = \sum_{i=1}^{n} x_i y_i$$
$$n \cdot b + \left(\sum_{i=1}^{n} x_i\right) \cdot a = \sum_{i=1}^{n} y_i$$

Durch eine Zeilenoperation ist das System äquivalent zu

$$n \cdot b + \left(\sum_{i=1}^{n} x_i\right) \cdot a = \sum_{i=1}^{n} y_i$$
$$\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) \cdot a = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

also

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

und damit

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot a\right).$$

Mit den Mittelwerten

$$\overline{X} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \qquad \overline{Y} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)$$

können wir dann schreiben:

Satz 4.4.1 *Die eindeutige Lösung des linearen Regressionsproblems*

$$y = a \cdot x + b$$

für die Messwerte $(x_1, y_1), \ldots, (x_n, y_n)$ *ist gegeben durch*

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{X} \cdot \overline{Y}}{\sum_{i=1}^{n} x_i^2 - n \cdot \overline{X}^2} \qquad b = \overline{Y} - a \cdot \overline{X}$$

sofern

$$\sum_{i=1}^{n} x_i^2 \neq n \cdot \overline{X}^2$$

Man beachte: Schreiben wir das lineare Gleichungssystem in Matrixform als

$$\underbrace{\begin{pmatrix} \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ n & \sum_{i=1}^{n} x_i \end{pmatrix}}_{A} \cdot \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

dann ist die Bedingung der eindeutigen Lösbarkeit die Bedingung, dass A invertierbar ist, äquivalent, dass $\det A \neq 0$.

Beispiel 4.4.2 *Für die Punkte*

$$(1,1),(2,3),(4,4)$$

erhalten wir das Gleichungssystem

$$\begin{pmatrix} 7 & 21 \\ 3 & 7 \end{pmatrix} \cdot \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 23 \\ 8 \end{pmatrix}$$

mit der Lösung

$$a = \frac{23 - 3 \cdot \frac{7}{3} \cdot \frac{8}{3}}{21 - 3 \cdot \left(\frac{7}{3}\right)^2} = \frac{13}{14}$$
$$b = \frac{8}{3} - \frac{13}{14} \cdot \frac{7}{3} = \frac{1}{2}$$

Siehe dazu Abbildung 4.1 für eine Darstellung der drei Punkte und der linearen Funktion $f(x) = a \cdot x + b$. Alternativ können wir mit MAPLE *das lineare Gleichungssystem lösen:*

```
with(LinearAlgebra):
A := <<7,3>|<21,7>>;
v := <23,8>;
LinearSolve(A, v);
```

$$\begin{pmatrix} \frac{1}{2} \\ \frac{13}{14} \end{pmatrix}$$

Bemerkung 4.4.3 *Lineare Regression mit Polynomfunktionen höheren Grades funktioniert analog, man erhält nur größere lineare Gleichungssysteme.*

Bemerkung 4.4.4 *In* MAPLE *gibt es umfangreiche Funktionalität zur linearen Regression (mit Polynomfunktionen beliebigen Grades). Die obige Rechnung können wir durchführen mit*

```
with(Statistics):
X:= Vector([1,2,4],datatype=float):
Y:= Vector([1,3,4],datatype=float):
LinearFit([1,x],X,Y,x);
```

```
0.499999999999999 + 0.928571428571429 x
```

Das erste Argument in `LinearFit` *spezifiziert hier die Terme, die in der zu findenden Funktion vorkommen dürfen, in unserem Fall* `[1,x]`.

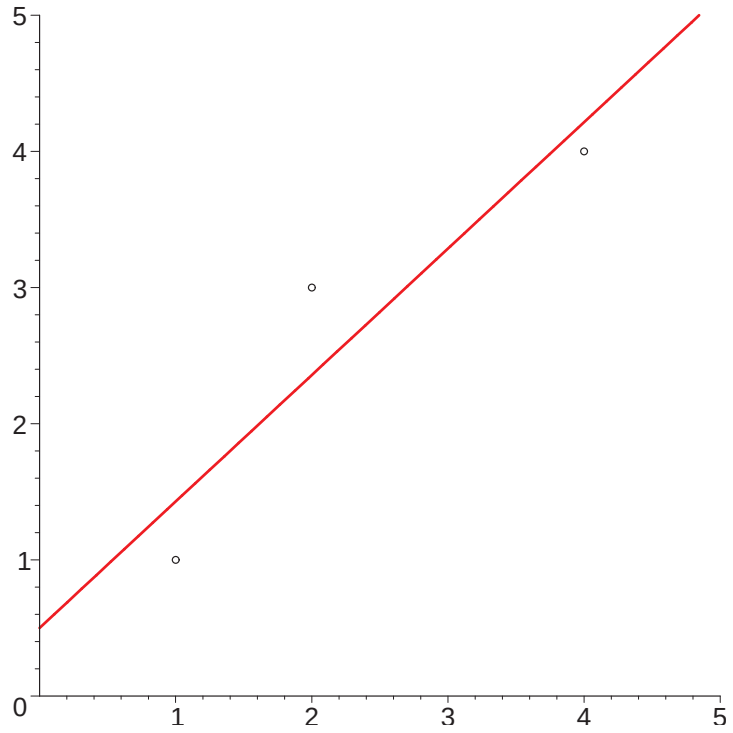


Abbildung 4.1: Lineare Regression.

4.5 Pseudozufallszahlen

In verschiedenen Beispielen haben wir Zufallszahlengeneratoren verwendet. Da Computer für feste Eingabe ein deterministisches Ergebnis liefern, stellt sich natürlich die Frage, woher diese Zufallszahlen kommen. In der Praxis verwendet man sogenannte **Pseudozufallszahlengeneratoren**. Diese liefern eine deterministische Folge von Zahlen, die aber die Eigenschaften von Zufallszahlen besitzen. Bei gleichem Input (diesen bezeichnet man als den **random seed**) erhalten wir aber immer dieselbe Folge. Akutelle Computerhardware enthält echte Zufallszahlengeneratoren, die auf der Messung von physikalischen Prozessen basieren. Deren Verwendung ist im Prinzip die weit bessere Lösung (ausser man möchte ein Programm debuggen und verwendet die Reproduzierbarkeit der Folge von Zufallszahlen). Allerdings ist bei einem Zufallszahlengenerator, der in Hardware implementiert

ist, sehr schwer zu verifizieren, dass dieser nicht in irgendeiner Form manipuliert ist. In jedem Fall sind kryptographische Verfahren, die auf Zufallszahlen basieren, darauf angewiesen, dass der Angreifer nicht die verwendeten Zufallszahlen reproduzieren oder statistisch vorhersagen kann.

Ein trivales Kriterium für einen Pseudozufallszahlengenerator ist, dass die erzeugten Zahlen im vorgegebenen Bereich gleichverteilt sind. Dies kann man z.B. mit einer iterierten Divison mit Rest erreichen:

Beispiel 4.5.1 *Man wählt Zahlen $a, b, N \in \mathbb{N}_{\geq 2}$ und einen random seed $x_0 \in M = \{0, \ldots, N-1\}$ und berechnet dann induktiv (mit Division mit Rest)*

$$x_i \in M$$

die die lineare Rekursion

$$x_i \equiv a \cdot x_{i-1} + b \bmod N$$

erfüllen. Einen solchen Zufallszahlengenerator bezeichnet man als ***linearen Kongruenzgenerator****. Da M endlich ist, muss sich sich die Zahl x_i irgendwann wiederholen und damit die Folge (x_i) periodisch werden. Tatsächlich kann man die Zahl $b = 0$ setzen ohne an Qualität des Zufallszahlengenerators zu verlieren. Man beachte, dass für N prim und $x_0 \neq 0$ die Zahl x_i nie 0 wird, da wir Zahlen $\neq 0$ in dem Körper $\mathbb{Z}/N$ multiplizieren. Man kann zeigen, dass für N prim die multiplikative Einheitengruppe*

$$(\mathbb{Z}/N)^\times = \{\overline{1}, \ldots, \overline{N-1}\}$$

zyklisch ist. Wählen wir $\overline{a} \in \mathbb{Z}/N$ als einen Erzeuger der zyklischen Einheitengruppe $(\mathbb{Z}/N)^\times$, also

$$(\mathbb{Z}/N)^\times = \langle \overline{a} \rangle$$

dann ist die Periode von (x_i) genau $N-1$ und die Folge ist dann offenbar auch gleichverteilt.

Bemerkung 4.5.2 *Hat man Zugriff auf genügend aufeinanderfolgende Elemente einer mit einem linearen Kongruenzgenerator erzeugten Folge, kann man die Parameter a und b bestimmen*

und damit weitere Elemente vorhersagen. Aus diesem Grund sind lineare Kongruenzgeneratoren für die Kryptographie nicht gut geeignet. Für die Verwendung in randomisierten Algorithmen spielen sie dennoch eine wichtige Rolle und sind in vielen Programmiersprachen vorhanden.

Beispiel 4.5.3 *In* MAPLE *können wir einen solchen Zufallszahlengenerator implementieren mit:*

```
x:=1;
a:=7;
N:=23;
L:=[]:
for j from 1 to 2200 do
    L:=[op(L),x];
    x:=a*x mod N;
od:
seq(L[j],j=1..10);
1, 7, 3, 21, 9, 17, 4, 5, 12, 15
```

Wir überprüfen, dass die Zahlen x_i tatsächlich gleichverteilt auftreten (man beachte 2200 *ist ein Vielfaches von der Gruppenordnung von $(\mathbb{Z}/23)^\times$ und $\overline{7}$ ist ein zyklischer Erzeuger):*

```
H:=[seq(0,jj=1..N)]:
for j from 1 to nops(L) do
        H[L[j]+1]:=H[L[j]+1]+1;
od:
H;
[100,100,100,100,100,100,100,100,100,100,100,100,
100,100,100,100,100,100,100,100,100,100]
```

Wie misst man nun die Qualität der Zufallszahlen? Ein Test auf Gleichverteiltheit ist nicht gut genug:

Beispiel 4.5.4 *Verwenden wir die lineare Rekursion*

$$x_i = x_{i-1} + 1 \bmod 7$$

dann sind die Pseudozufallszahlenßwar gleich verteilt, aber wir erhalten stets

$$\ldots 0, 1, 2, 3, 4, 5, 6, 0, 1, 2, 3, 4, 5, 6 \ldots$$

Beispiel 4.5.5 *In einer Zufallsfolge tritt eine streng monotone (aufsteigende oder absteigende) Sequenz der Länge $n \geq 2$ mit Wahrscheinlichkeit*

$$p_n \approx \frac{n}{(n+1)!}$$

auf, denn haben wir induktiv schon $n-1$ Zahlen (OE aufsteigend) gewählt, dann teilen diese die Menge M in n Teile und im Mittel $\frac{1}{n}$ Zahlen können die Folge fortsetzen (die Zahlen im obersten Teil). Wenn die Folge nach dem Schritt n endet, dann kann man aus den $n+1$ Teilen im Mittel $\frac{n}{n+1}$ für das nächste Element wählen (die Zahlen, die nicht im obersten Teil liegen). Bestimmen Sie als Übung auch die Wahrscheinlichkeit einer konstanten Folge der Länge n.

Beispiel 4.5.6 *Die Wahrscheinlichkeiten für streng monotone Folgen der Länge n sind dann*

n	2	3	4	5
p_n	$\frac{1}{3}$	$\frac{1}{8}$	$\frac{1}{30}$	$\frac{1}{144}$

In der mittels

Übung 4.1 `seq(rand(1..100)(),j=1..20);`
in MAPLE *erzeugten Zufallssequenz*

$$[53, 71, 33, 41, 17, 45, 85, 27, 49, 94, 11, 39, 87, 52, 58, 52, 49, 48, 20, 47]$$

finden wir 9 streng monotone Sequenzen der Länge 2 und 3 Sequenzen der Länge 3. Das stimmt mit der theoretischen Überlegung gut überein.

Beispiel 4.5.7 *In der Folge aus Beispiel 4.5.4 gibt nur streng monoton aufsteigende Sequenzen der Länge 7, jeweils gefolgt von einer absteigenden Sequenz der Länge 2. Entsprechend der obigen Beobachtung erkennen wir also (x_i) als eine schlechte Folge von Zufallszahlen.*

Bemerkung 4.5.8 *Eine geratene Sequenz erkennt man oft daran, dass die relativen Häufigkeiten der monotonen Sequenzen nicht stimmen. Lassen Sie als Übung eine mit dem Thema nicht vertraute Person eine Zufallsfolge raten und wenden Sie das obige Kritierum an (Übung 4.6).*

4.6 Bayes-Klassifizierer

Einer der wichtigsten Algorithmen im Machine-Learning ist (trotz seiner Einfachheit) der sogenannte Bayes-Klassifizierer, der auf der Bayes-Umkehrformel aus Abschnitt 2.11 basiert. Einsetzbar ist das Verfahren sowohl auf diskreten als auch kontinuierlichen Wahrscheinlichkeitsräumen. Praktisch verwendet man den Bayes-Klassifizierer zur Unterscheidung von Objekten anhand von Eigenschaften, beim natural language processing und insbesondere zur Klassifikation von Texten (z.B. Spam oder nicht Spam bei Emails) und in medizinischen Diagnosen. Das Grundprinzip haben wir schon in Beispiel 2.11.14 gesehen, wo wir Patienten in krank oder nicht krank klassifiziert haben. Allerdings hatten wir hier nur eine einzige Eigenschaft verwendet, das Ergebnis des durchgeführten Tests. Allgemein wird man aber mehrere Eigenschaften zur Klassifikation heranziehen wollen. Wir betrachten im Folgenden als ein Beispiel das Problem der Klassifikation eines Texts.

Zunächst zu der mathematischen Grundlage: Wir wollen zwischen möglichen Klassen unterscheiden, also teilen wir den Wahrscheinlichkeitsraum Ω, den wir im Folgenden als diskret annehmen, auf in eine Partition

$$\Omega = A_1 \cup \ldots \cup A_r.$$

Nach der Definition der bedingten Wahrscheinlichkeit gilt dann für jedes Ereignis $B \subset \Omega$

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$
$$P(B \mid A_i) = \frac{P(A_i \cap B)}{P(A_i)}$$

und damit

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A)}{P(B)}$$

(siehe Satz 2.11.10). In der Praxis wollen wir durch B mehrere Eigenschaften beschreiben. Dazu setzen wir

$$B = B_1 \cap \ldots \cap B_s$$

für

$$B_i = \{X_i = n_i\}$$

mit Zufallsvariablen X_i. Hier könnten z.B.

$$X_i : \Omega \to \{0, 1\}$$

angeben, ob das Wort Nummer i in einem gegebenen Satz vorkommt (Wert 1) oder nicht (Wert 0). In dem sogenannten **naiven Bayes-Klassifizierer** nehmen wir an, dass diese Eigenschaften in jeder Klasse A_i unabhängig vorkommen. Dazu führen wir analog zur Definition von Unabhängigkeit von Zufallsvariablen in Abschnitt 2.9 den folgenden Begriff ein:

Definition 4.6.1 *Zufallsvariablen $X_1, \ldots, X_s$ mit $X_i : \Omega \to N_i$ heißen **bedingt unabhängig gegeben das Ereignis** $A \subset \Omega$, falls*

$$P(X_1 = n_1, \ldots, X_s = n_s \mid A) = P(X_1 = n_1 \mid A) \cdot \ldots \cdot P(X_s = n_s \mid A)$$

für alle $n_1, \ldots, n_s$.

Man beachte, dass nach unserer Definition von bedingter Wahrscheinlichkeit

$$P(X = n \mid A) = \frac{P(X = n) \cap A}{P(A)} = \frac{P(\{\omega \in A \mid X(\omega) = n\})}{P(A)}$$

ist.

Satz 4.6.2 *Ist Ω ein diskreter Wahrscheinlichkeitsraum,*

$$\Omega = A_1 \cup \ldots \cup A_r$$

eine Partition von Ω in paarweise disjunkte Ereignisse und sind $X_i : \Omega \to N_i$, $i = 1, \ldots, s$ Zufallsvariablen, die bedingt unabhängig sind gegeben das Ereignis A_i für alle i, dann gilt

$$P(A_i \mid X_1 = n_1, \ldots, X_s = n_s) = \frac{P(A_i) \cdot P(X_1 = n_1 \mid A_i) \cdot \ldots \cdot P(X_s = n_s \mid A_i)}{P(X_1 = n_1, \ldots, X_s = n_s)}.$$

Beweis. Die Bayes-Formel und die bedingte Unabhängigkeit liefern

$$P(A_i \mid X_1 = n_1, \ldots, X_s = n_s) = \frac{P(A_i) \cdot P(X_1 = n_1, \ldots, X_s = n_s \mid A_i)}{P(X_1 = n_1, \ldots, X_s = n_s)}$$
$$= \frac{P(A_i) \cdot P(X_1 = n_1 \mid A_i) \cdot \ldots \cdot P(X_s = n_s \mid A_i)}{P(X_1 = n_1, \ldots, X_s = n_s)}$$

■

Der Klassifizierer funktioniert nun einfach durch den Vergleich von bedingten Wahrscheinlichkeiten:

- Aus einem Trainingsdatensatz bestimmen wir die bedingten Wahrscheinlichkeiten $P(X_j = n_j \mid A_i)$ als relative Häufigkeiten. Wir lesen auch die Wahrscheinlichkeiten $P(A_i)$ als relative Häufigkeiten in unserem Trainingsdatensatz ab.

  In unserem Anwendungsbeispiel zählen wir also, wie häufig bestimmte Worte in Texten vorkommen, die zu der Klasse A_i gehören.

- Für ein gegebenes Ereignis $\{\omega \mid X_1(\omega) = n_1, \ldots, X_s(\omega) = n_s\}$ (in unserem Anwendungsbeispiel ein Satz mit den Worten korrespondierend zu $X_1, \ldots, X_s$) vergleichen wir die Wahrscheinlichkeiten

  $$P(A_i \mid X_1 = n_1, \ldots, X_s = n_s)$$

  für verschiedene i und entscheiden uns für das A_i, für das die Wahrscheinlichkeit maximal wird.

Bemerkung 4.6.3 *In der Praxis muss man noch einen kleinen Trick anwenden: Es kann sein, dass in bestimmten Klassen A_i bestimmte Worte B_j nicht vorkommen (d.h. die relative Häufigkeit $P(X_j = n_j \mid A_i) = 0$). In diesem Fall nimmt die Formel in 4.6.2 den Wert 0. Zwischen verschiedenen Klassen mit Wert 0 ist dann keine Entscheidung möglich. Aus diesem Grund fügt man dem Trainingsdatensatz üblicherweise noch pro Klasse A_i je eine Stichprobe mit allen Elementen von Ω hinzu.*

Wir erproben den Klassifizierer an einem Beispiel:

Beispiel 4.6.4 *Unsere Trainingsdaten sind die folgenden Sätze, die den angegebenen Klassen zugeordnet worden sind:*

Stichproben	*Klasse*
ein faszinierender Sportwagen	A_1
der Zug ist weg	A_2
sehr schöner Wagen	A_1
ein schöner aber kleiner Sportwagen	A_1
dies ist ein schneller Zug	A_2

Wir wollen für einen Eingabesatz (der typischerweise nicht in den Trainingsdaten vorkommt) entscheiden, ob er in die Klasse A_1 (Auto) oder A_2 (Zug) gehört. Wir betrachten den Eingabesatz

ein sehr schneller Sportwagen

Insgesamt kommen die 14 Worte $W_1, \ldots, W_{14}$

ein, faszinierender, Sportwagen, der, Zug, ist, weg,
sehr, schöner, Wagen, aber, kleiner, dies, schneller

vor. Tupel (Sätze) aus diesen sind die Elemente des Wahrscheinlichkeitsraums Ω. Unser Eingabesatz enthält die Worte W_1, W_3, W_8, W_{14}. Worte aus der Klasse A_1 kommen in unserem Trainingsdatensatz mit der relativen Häufigkeit

$$P(A_1) = \frac{11}{20}$$

vor, Worte aus A_2 mit der relativen Häufigkeit

$$P(A_2) = \frac{9}{20}$$

Wir schreiben für das Ereignis $X_i = 1$, dass das i-te Wort W_i vorkommt, das jeweilige Wort. Wir erhalten dann die folgenden bedingten Wahrscheinlichkeiten

$$P(ein \mid A_1) = \frac{P(ein \cap A_1)}{P(A_1)} = \frac{\frac{2}{20}}{\frac{11}{20}} = \frac{2}{11}$$

Man beachte, dass wir die Häufigkeiten statt den relativen Häufigkeiten verwenden können, da sich der Nenner 20 wegkürzt.

Analog erhalten wir

W	$P(W \mid A_1)$	$P(W \mid A_2)$
ein	$\frac{2}{11}$	$\frac{1}{9}$
sehr	$\frac{1}{11}$	$\frac{0}{9}$
schneller	$\frac{0}{11}$	$\frac{1}{9}$
Sportwagen	$\frac{2}{11}$	$\frac{0}{9}$

Hier zeigt sich das oben angesprochene Problem, dass manche der bedingten Wahrscheinlichkeiten 0 *sind. Um dies zu umgehen, fügen wir dem Trainingsdatensatz sowohl für* A_1 *als auch* A_2 *eine Stichprobe mit allen* 14 *Worten hinzu und erhalten*

W	$P(W \mid A_1)$	$P(W \mid A_2)$
ein	$\frac{3}{25}$	$\frac{2}{23}$
sehr	$\frac{2}{25}$	$\frac{1}{23}$
schneller	$\frac{1}{25}$	$\frac{2}{23}$
Sportwagen	$\frac{3}{25}$	$\frac{1}{23}$

Hiermit können wir nun berechnen

$$
\begin{aligned}
&P(A_1 \mid W_1, W_3, W_8, W_{14}) \\
&\quad = \frac{P(A_1) \cdot P(W_1 \mid A_1) \cdot P(W_3 \mid A_1) \cdot P(W_8 \mid A_1) \cdot P(W_{14} \mid A_1)}{P(B)} \\
&\quad = \frac{\frac{11}{20} \cdot \frac{3}{25} \cdot \frac{2}{25} \cdot \frac{1}{25} \cdot \frac{3}{25}}{P(B)} = \frac{\frac{99}{3906\,250}}{P(B)} \approx \frac{0.000025}{P(B)}
\end{aligned}
$$

und

$$
\begin{aligned}
&P(A_2 \mid W_1, W_3, W_8, W_{14}) \\
&\quad = \frac{P(A_2) \cdot P(W_1 \mid A_2) \cdot P(W_3 \mid A_2) \cdot P(W_8 \mid A_2) \cdot P(W_{14} \mid A_2)}{P(B)} \\
&\quad = \frac{\frac{9}{20} \cdot \frac{2}{23} \cdot \frac{1}{23} \cdot \frac{2}{23} \cdot \frac{1}{23}}{P(B)} = \frac{\frac{9}{1399\,205}}{P(B)} \approx \frac{0.0000064}{P(B)}
\end{aligned}
$$

Man beachte, dass der Nenner $P(B)$ *in beiden Ausdrücken identisch und damit für den Vergleich irrelevant ist. Diese bedingten Wahrscheinlichkeiten weisen klar darauf hin, dass der zu klassifizierende Satz der Klasse* A_1 *zuzuordnen ist.*

4.7 Übungen

Übung 4.2 *Der Gewinn eines Unternehmens war in den angegebenen Jahren wie folgt:*

Jahr	2015	2016	2017	2018	2019
Gewinn	13	20	28	38	48

(gemessen in Millionen €).

1) *Bestimmen Sie mit der Methode der kleinsten Quadrate eine Regressionsgerade, die die obigen Werte approximiert.*

2) *Schätzen Sie den Gewinn des Unternehmens im Jahr* 2022.

Übung 4.3 *In der folgenden Tabelle sind die Häufigkeiten für Cholesterinwerte und Blutdruck in einer Studie mit* 80 *Teilnehmern angegeben:*

	$[100, 120[$	$[120, 140[$	$[140, 160[$	*Blutdruck*
$[170, 190[$	9	4	1	
$[190, 210[$	9	13	4	
$[210, 230[$	7	8	8	
$[230, 250[$	1	5	11	
Cholesterin				

1) *Bestimmen Sie die Randverteilungen.*

2) *Finden Sie eine Regressionsgerade, die den Cholesterinwert in Abhängigkeit vom Blutdruck beschreibt. Verwenden Sie für die Datenpunkte die jeweiligen Intervallmittelpunkte.*

3) *Welchen Cholesterinwert erwarten Sie für eine Person mit einem Blutdruck von* 170.

4) *Finden Sie eine Regressionsgerade, die den Blutdruck in Abhängigkeit von dem Cholesterinwert beschreibt.*

5) *Welchen Blutdruck erwarten Sie für eine Person mit einem Cholesterinwert von* 100.

Übung 4.4 *Ein neuer Algorithmus wird auf* 50 *Eingabedaten angewendet und liefern eine mittlere Laufzeit von* 9.27 *Sekunden mit einer Standardabweichung von* 0.21 *Sekunden. Der bisherige Algorithmus hatte eine Laufzeit von* 9.31 *Sekunden. Testen Sie mit einem Signifikanzniveau von* 1% *die Hypothese, dass der neue Algorithmus schneller ist als der alte.*

Hinweis: Welche Wahrscheinlichkeitsverteilung würden Sie in dieser Situation annehmen?

Übung 4.5 *1) An einer Universität wird bei* 50 *Beschäftigten (hoffentlich anonym) die Zeit gemessen, die die Mitarbeiter auf den Internetseiten Facebook und Ebay zubringen. Pro Arbeitstag von* 12 *Stunden sind dies* 32 *Minuten mit einer Standardabweichung von* 9 *Minuten. Bestimmen Sie ein Konfidenzintervall zum Konfidenzniveau* 99% *für den Mittelwert der Facebook- und Ebayzeit, wobei Sie eine Normalverteilung für die Zeit annehmen.*

2) Eine Bank möchte bis auf 25 € *das mittlere monatliche Guthaben auf den Kundenkonten abschätzen. Unter der Annahme, dass die Standardabweichung* $\sigma = 250$ € *ist, finden Sie den minimalen Stichprobenumfang für ein Konfidenzniveau von* 99.8%.

Übung 4.6 *Lassen Sie als eine Person eine Zufallsfolge von ganzen Zahlen im Intervall* $[0, 50]$ *raten und überprüfen Sie die Qualität der Zufallsfolge anhand der relativen Häufigkeiten von auf- bzw. absteigenden Sequenzen verschiedener Länge.*

Übung 4.7 *Implementieren Sie eine naiven Bayes-Klassifizierer und wenden Sie diesen auf die Texte aus Beispiel 2.11.8 an. Denken Sie sich weitere Eingabesätze aus den gegebenen Worten aus und erproben Sie den Klassifizierer.*

5

Anhang

5.1 Ausblick: Axiomatische Wahrscheinlichkeitsräume

In Abschnitt 3.2 haben wir schon gesehen, dass man im Allgemeinen nicht erwarten kann, dass man einer beliebigen Teilmenge eines Wahrscheinlichkeitsraums Ω eine sinnvolle Wahrscheinlichkeit zuordnen kann. Die Idee ist einfach nur Teilmengen zuzulassen, für die das eben geht. Das Musterbeispiel ist hier, dass die Wahrscheinlichkeit eines Intervalls $[a, b] \subset \mathbb{R}$ gegeben ist durch das Integral einer Wahrscheinlichkeitsdichte.

Definition 5.1.1 *Eine* ***Sigma-Algebra*** *(auch oft geschrieben σ-Algebra) ist eine Menge*

$$\Sigma \subset 2^{\Omega}$$

von Teilmengen von Ω mit

1) $\Omega \in \Sigma$

2) Σ *ist abgeschlossen unter Komplementen:*

$$A \in \Sigma \Rightarrow \Omega \backslash A \in \Sigma.$$

Wir schreiben auch kurz

$$\overline{A} := \Omega \backslash A.$$

3) Σ ist abgeschlossen unter abzählbaren Vereinigungen:

$$A_1, A_2, \ldots \in \Sigma \Rightarrow A_1 \cup A_2 \cup \ldots \in \Sigma.$$

Bemerkung 5.1.2 *Die bekannten de Morganschen Gesetze*

$$\overline{A_1 \cup A_2} = \overline{A_1} \cap \overline{A_2}$$
$$\overline{A_1 \cap A_2} = \overline{A_1} \cup \overline{A_2}$$

für Mengen A_i gelten offensichtlich auch für unendliche (abzählbare oder überabzählbare) Durchschnitte und Vereinigungen: Ist I eine Menge und ist A_i eine Familie von Mengen mit $i \in I$, dann ist

$$\begin{aligned}\overline{\bigcup_{i\in I} A_i} &= \{\omega \in \Omega \mid \omega \notin A_i \forall i\} \\ &= \left\{\omega \in \Omega \mid \omega \in \overline{A_i} \forall i\right\} \\ &= \bigcap_{i\in I} \overline{A_i}.\end{aligned}$$

Insbesondere sind Sigmaalgebren abgeschlossen unter abzählbaren Durchschnitten, denn

$$\bigcap_i A_i = \overline{\bigcup_i \overline{A_i}}$$

für $A_i \in \Sigma$.

Beispiel 5.1.3 *1) Die Menge Σ aller Teilmengen von Ω, d.h. die Potenzmenge $\Sigma = 2^\Omega$, ist eine Sigma-Algebra.*

2) $\{\varnothing, \Omega\}$ ist eine Sigma-Algebra.

Wir interessieren uns hauptsächlich für die folgende Sigma-Algebra:

Beispiel 5.1.4 *Die kleinste Sigma-Algebra Σ, die alle Intervalle der Form*

$$]-\infty, b] \subset \Omega = \mathbb{R}$$

enthält, heißt ***Borel-Sigma-Algebra****.*

Als Komplemente davon enthält Σ dann auch alle Intervalle der Form

$$]b,\infty[\;= \overline{]-\infty,b]}.$$

Nach Bemerkung 5.1.2 enthält Σ auch alle Intervalle der Form

$$]-\infty,b[\;= \bigcup_{n=1}^{\infty} \left]-\infty, b-\frac{1}{n}\right]$$

damit alle

$$[b,\infty[\;= \overline{]-\infty,a[}.$$

Damit enthält Σ alle

$$]a,b[\;=\;]-\infty,b[\;\cap\;]a,\infty[$$

und alle

$$[a,b] = \;]-\infty,b] \cap [a,\infty[.$$

Im Wesentlichen dieselbe Sigma-Algebra können wir auch auf Teilmengen verwenden:

Bemerkung 5.1.5 *Ist Σ die Borel-Sigma-Algebra auf $\Omega = \mathbb{R}$ und $\Omega' \subset \Omega$ eine Teilmenge, dann ist*

$$\Sigma' = \{\Omega' \cap A \mid A \in \Sigma\}$$

eine Sigma-Algebra auf Ω', die wir wieder als Borel-Sigma bezeichnen.

Bemerkung 5.1.6 *Ist Σ_1 eine Sigma-Algebra auf Ω_1 und Σ_2 eine Sigma-Algebra auf Ω_2, dann ist die **Produkt-Sigma-Algebra** auf $\Omega_1 \times \Omega_2$ erzeugt von allen $A_1 \times A_2$ mit $A_i \in \Sigma_i$.*

Das Resultat dieser Konstruktion bezeichenen wir als die Borel-Sigma-Algebra auf $\Omega = \mathbb{R}^n$.

Mit dem Konzept der Sigma-Algebra können wir nun unsere Definition eines Wahrscheinlichkeitsraums verallgemeinern:

Definition 5.1.7 *Ein **Wahrscheinlichkeitsraum** ist Ergebnismenge Ω zusammen mit einer Sigma-Algebra Σ auf Ω und einer Funktion*

$$P : \Sigma \to [0,1]$$

mit

1) $P(\Omega) = 1$

2) und für paarweise disjunkte $A_1, A_2, \ldots \in \Sigma$

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

Die Funktion P *bezeichnen wir als* ***Wahrscheinlichkeitsverteilung*** *oder* ***Wahrscheinlichkeitsmaß****.*

Einen Wahrscheinlichkeitsraum ist also ein Tupel (Ω, Σ, P)*.*

Bemerkung 5.1.8 *Es gilt dann offenbar*

$$P(\varnothing) = 0.$$

Beispiel 5.1.9 *Ist* Ω *endlich oder abzählbar, dann ist* $\Sigma = 2^\Omega$ *eine Sigma-Algebra. Eine Wahrscheinlichkeitsfunktion*

$$m : \Omega \to \mathbb{R}_{\geq 0}$$

definiert einen Wahrscheinlichkeitsraum (Ω, Σ, P)*, da jedes* $A \in \Sigma = 2^\Omega$ *sich als abzählbare disjunkte Vereinigung*

$$A = \bigcup_{\omega \in A} \{\omega\}$$

schreiben lässt und wir daher das Wahrscheinlichkeitsmaß

$$P(A) = \sum_{\omega \in A} m(\omega)$$

definieren können.

Umgekehrt liefert ein Wahrscheinlichkeitsmaß $P : \Sigma \to \mathbb{R}_{\geq 0}$ *eine Wahrscheinlichkeitsfunktion* $m : \Omega \to \mathbb{R}_{\geq 0}$ *durch*

$$m(\omega) := P(\{\omega\}).$$

Die erweiterte Definition 5.1.7 stimmt also für abzählbare Ω *mit unserer bisherigen Definition 2.3.1 überein.*

Beispiel 5.1.10 *Sind* $(\Omega_1, \Sigma_1, P_1)$ *und* $(\Omega_2, \Sigma_2, P_2)$ *Wahrscheinlichkeitsräume, dann auch* $\Omega_1 \times \Omega_2$ *mit der Produkt-Sigma-Algebra* Σ *und dem Produkt-Wahrscheinlichkeitsmaß*

$$P(A_1 \times A_2) = P(A_1) \cdot P(A_2).$$

Beispiel 5.1.11 *Auf $\Omega \subset \mathbb{R}^n$ zusammen mit der Borel-Sigma-Algebra Σ und einer integrierbaren Funktion*

$$f : \Omega \to \mathbb{R}$$

mit

$$\int_\Omega f(x)dx = 1,$$

genannt Wahrscheinlichkeitsdichte, ist durch

$$P(A) = \int_A f(x)dx$$

für $A \in \Sigma$ ein Wahrscheinlichkeitsraum gegeben.

Definition 5.1.12 *Für eine Zufallsvariable $X : \Omega \to N$ betrachten wir Wahrscheinlichkeiten $P(X \in A)$ nur für Mengen $A \subset N$ mit*

$$X^{-1}(A) \in \Sigma$$

und setzen dann

$$P(X \in A) := P(X^{-1}(A)) = P(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

Durch

$$\Sigma' = \left\{A \in 2^N \mid X^{-1}(A) \in \Sigma\right\}$$

ist dann auf N eine Sigma-Algebra gegeben und durch

$$P(A) = P(X \in A)$$

für $A \in \Sigma'$ eine Wahrscheinlichkeitsverteilung. Diese bezeichnen wir als die ***Verteilung der Zufallsvariable*** *von X.*

5.2 Zur Integration: Substitutionsregel und Transformationsformel

Bei der Untersuchung der Normalverteilung hatten wir statt rechtwinkligen Koordinaten, Polarkoordinaten verwendet. Den Wechsel von Koordinaten bei der Integration beschreibt allgemein der Transformationssatz für Integrale. Dieser ist eine mehrdimensionale Verallgemeinerung der Substitutionsregel:

Satz 5.2.1 *Ist $f : [r,s] \to \mathbb{R}$ stetig und $g : [a,b] \to [r,s]$ differenzierbar mit stetiger Ableitung, dann gilt*

$$\int_r^s (f \circ g)(x) \cdot g'(x)\ dx = \int_{g(r)}^{g(s)} f(y)\ dy.$$

Dies folgt direkt aus der Kettenregel für Ableitungen.

Die mehrdimensionale Verallgemeinerung können wir wie folgt formulieren:

Satz 5.2.2 *Ist $M \subset \mathbb{R}^n$ eine offene Menge*[1] *und*

$$\Phi : M \to N \subset \mathbb{R}^n$$

stetig differenzierbar und bijektiv, dann ist f auf $\Phi(M)$ integrierbar genau dann, wenn $f \circ \Phi$ auf M integrierbar ist, und

$$\int_{\Phi(M)} f(y)dy = \int_M f(x) \cdot T(x)\ dx$$

mit

$$T(x) = \det\left(\frac{\partial \Phi_i(x)}{\partial x_j}\right)_{i,j}.$$

Beispiel 5.2.3 *Im Fall der Polarkoordinaten ist*

$$\begin{aligned} T(r,\varphi) &= \det\begin{pmatrix} \cos\varphi & -r\cdot\sin\varphi \\ \sin\varphi & r\cdot\cos\varphi \end{pmatrix} \\ &= r\cdot(\cos\varphi)^2 + r\cdot(\sin\varphi)^2 = r. \end{aligned}$$

5.3 Beweis des Zentralen Grenzwertsatzes

In diesem Abschnitt wollen wir einen Beweis des Zentralen Grenzwertsatzes (Satz 3.8.11) skizzieren. Dieser besagt, dass für identisch unabhängig verteilte Zufallsvariablen $X_1, \ldots, X_n$ mit Erwartungswert

$$\mu = E(X_i)$$

[1] Eine offene Menge ist eine Verallgemeinerung des offenen Intervalls $]a,b[$. Produkte von solchen Intervallen sind z.B. offen.

und Varianz

$$\sigma = \sigma(X_i)$$

für alle

$$-\infty \leq a < b \leq \infty$$

gilt, dass

$$\lim_{n\to\infty} P\left(a \leq \frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \ldots + X_n}{n} - \mu\right) \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{x^2}{2})dx.$$

Ohne Einschränkung können wir annehmen, dass $\mu = 0$ ist (anderenfalls kann man X_i durch $X_i - \mu$ ersetzen). Für

$$Y_n := \frac{X_1 + \ldots + X_n}{\sigma\sqrt{n}}$$

gilt dann

$$E(Y_n) = 0$$

nach Satz 2.6.12 und nach Satz 2.7.9 wegen der Unabhängigkeit

$$V(Y_n) = 1.$$

Wie im Beweis der Hoeffding-Ungleichung codieren wir wieder alle Momente in der Erzeugendenfunktion

$$M_X(\lambda) = E(\exp(\lambda \cdot X)) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} E(X^n)$$

wobei wir X für die gemeinsame Verteilung der X_i schreiben und annehmen, dass alle Momente existieren. Wir verwenden die folgenden Hilfsaussagen, die im wesentlichen sagen, dass M die Eigenschaften der Exponentialfunktion erbt:

Lemma 5.3.1 *Sind X_1 und X_2 unabhängig, dann gilt*

$$M_{X_1+X_2}(\lambda) = M_{X_1}(\lambda) \cdot M_{X_2}(\lambda)$$

für alle $\lambda \in \mathbb{R}$.

Beweis. Mit der Funktionalgleichung der Exponentialfunktion und da sich der Erwartungswert in Produkte von unabhängigen Variablen hineinzieht, haben wir

$$\begin{aligned} M_{Z_1+Z_2}(\lambda) &= E(\exp(\lambda \cdot (X_1 + X_2))) \\ &= E(\exp(\lambda \cdot X_1) \cdot \exp(\lambda \cdot X_2)) \\ &= E(\exp(\lambda \cdot X_1)) \cdot E(\exp(\lambda \cdot X_2)) \\ &= M_{X_1}(\lambda) \cdot M_{X_2}(\lambda). \end{aligned}$$

■

Mit Induktion folgt aus Lemma 5.3.1:

Corollar 5.3.2 *Sind $X_1, \ldots, X_n$ identisch wie X und unabhängig verteilt, dann ist*

$$M_{X_1+\ldots+X_n}(\lambda) = M_X(\lambda)^n$$

Weiter folgt aus Lemma 5.3.1:

Corollar 5.3.3 *Für*

$$Y = aX + b$$

mit $a, b \in \mathbb{R}$ gilt

$$M_Y(\lambda) = M_X(a\lambda) \cdot \exp(b\lambda)$$

Mit Corollar 5.3.2 und 5.3.3 gilt dann

$$M_{Y_n}(\lambda) = \left(M_X(\frac{\lambda}{\sigma\sqrt{n}}) \right)^n$$

für alle $\lambda \in \mathbb{R}$. Mit der Annahme $E(X) = 0$ folgt

$$\begin{aligned} M_X(\frac{\lambda}{\sigma\sqrt{n}}) &= \sum_{j=0}^{\infty} \frac{\lambda^j}{j! \cdot \sigma^j \cdot \sqrt{n}^j} E(X^j) \\ &= 1 + \frac{1}{2}\frac{\lambda^2}{n} + \frac{R(n)}{n} \end{aligned}$$

wobei $\lim_{n\to\infty} R(n) = 0$. Somit ist

$$M_{Y_n}(\lambda) = \left(1 + \frac{1}{2}\frac{\lambda^2}{n} + \frac{R(\lambda, n)}{n} \right)^n$$

und mit einer kleinen Erweiterung von Lemma 3.8.4 in der Form

$$\lim_{n\to\infty}\left(1+\frac{x}{n}+\frac{R(n)}{n}\right)^n = \exp(x)$$

für $\lim_{n\to\infty} R(n) = 0$, folgt

$$\lim_{n\to\infty} M_{Y_n}(\lambda) = \exp(\frac{1}{2}\lambda^2).$$

Nach dem folgenden Lemma ist dies die Erzeugendenfunktion der Standardnomalverteilung:

Lemma 5.3.4 *Ist X standardnormalverteilt, dann gilt*

$$M_X(\lambda) = \exp(\frac{1}{2}\lambda^2).$$

Beweis. Es ist

$$\begin{aligned}\exp(\frac{1}{2}\lambda^2) &= \sum_{n=0}^{\infty}\frac{\lambda^{2n}}{n!}\frac{1}{2^n}\\ &= \sum_{m=0}^{\infty}\frac{\lambda^m}{m!}E(Z^m)\end{aligned}$$

denn wie wir aus Satz 3.8.15 schon wissen ist $E(1) = 1$, $E(Z) = 0$, $E(Z^2) = V(Z) = 1$ und man kann genauso mit partieller Integration zeigen, dass allgemein

$$E(Z^m) = \begin{cases} 0 & \text{für } m \text{ ungerade}\\ 1\cdot 3\cdot 5\cdot\ldots\cdot(m-1) & \text{für } m \text{ gerade}\end{cases}$$

und mit Induktion, dass

$$\frac{(2n)!}{n!2^n} = 1\cdot 3\cdot 5\cdot\ldots\cdot(2n-1).$$

■

Beispiel 5.3.5 *Wir entwickeln die Erzeugendenfunktion für die Momente der Standardnormalverteilung in* MAPLE *bis zur Ordnung* 10*:*

```
taylor(exp(1/2*t^2),10);
```

$1 + \frac{1}{2}t^2 + \frac{1}{8}t^4 + \frac{1}{48}t^6 + \frac{1}{384}t^8$

Es bleibt dann noch zu zeigen:

Satz 5.3.6 *Die Erzeugendenfunktion der Momente*

$$M_X(\lambda) = E(\exp(\lambda \cdot X))$$

legt die Verteilung von X eindeutig fest.

Beweis. Nehmen wir der Einfachheit halber an, dass X nur Werte in $\{1, \ldots, m\}$ annimmt. Dann ist

$$\begin{aligned} M_X(\lambda) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} E(X^n) \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \sum_{j=1}^{m} j^n \cdot P(X = j) \\ &= \sum_{j=1}^{m} P(X = j) \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} j^n \\ &= \sum_{j=1}^{m} P(X = j) \cdot \exp(j \cdot \lambda) \end{aligned}$$

Mit

$$T = \exp(\lambda)$$

ist also

$$M_X(\lambda) = \sum_{j=1}^{m} P(X = j) \cdot T^j$$

ein Polynom in T. Zwei Polynome sind aber gleich genau dann, wenn die Koeffizienten gleich sind. ■

Wir bemerken noch:

Corollar 5.3.7 *Für eine normalverteilte Variable Z mit Erwartungswert μ und Varianz σ^2 ist*

$$M_Z(\lambda) = \exp(\frac{1}{2}\sigma^2 \cdot \lambda^2 + \mu \cdot \lambda).$$

Beweis. Wir können schreiben

$$Z = \sigma X + \mu$$

mit $a, b \in \mathbb{R}$ und X standardnormalverteilt. Dann gilt mit Corollar 5.3.3 und Lemma 5.3.4, dass

$$\begin{aligned} M_Z(\lambda) &= M_X(\sigma\lambda) \cdot \exp(\mu \cdot \lambda) \\ &= \exp(\frac{1}{2}\sigma^2 \cdot \lambda^2 + \mu \cdot \lambda). \end{aligned}$$

■

Corollar 5.3.8 *Der Mittelwert*

$$X = \frac{1}{n}(X_1 + \ldots + X_n)$$

von identisch unabhängig normalverteilten Zufallsvariablen $X_1, \ldots, X_n$ mit Erwartungswert μ und Varianz σ^2 ist normalverteilt mit Erwartungswert

$$E(X) = \mu$$

und Varianz

$$V(X) = \frac{\sigma^2}{n}.$$

Beweis. Mit Corollar 5.3.2, 5.3.3 und 5.3.7 ist

$$\begin{aligned} M_X(\lambda) = M_{\frac{1}{n}X_i}(\lambda)^n = M_{X_i}(\frac{1}{n}\lambda)^n = \exp(\frac{1}{2}\sigma^2 \cdot \frac{\lambda^2}{n^2} + \mu \cdot \frac{\lambda}{n})^n \\ = \exp(\frac{1}{2}\frac{\sigma^2}{n} \cdot \lambda^2 + \mu \cdot \lambda). \end{aligned}$$

Mit Satz 5.3.6 folgt die Behauptung. ■

5.4 Computeralgebra

5.4.1 Überblick

Für die Kombinatorik, Analysis und elementares Programmieren ist ein Computeralgebrasystem mit allgemeiner Funktionalität am besten geeignet, da es alle drei Themengebiete gemeinsam abdeckt. Im kommerziellen Bereich sind MAPLE [13], und MATHEMATICA [15] verfügbar, ebenso die Open-Source-Systeme MAXIMA [14], REDUCE [18], und AXIOM [1], die allerdings einen deutlich kleineren Funktionsumfang besitzen.

Speziell für die Anwendung in der Algebra (exaktes Rechnen) gibt es deutlich leistungsfähigere Systeme, wie z.B. die Open-Source-Systeme SINGULAR [19], MACAULAY2 [9] und GAP [8], und das kommerzielle System MAGMA [10]. Dasselbe gilt für die Numerik (Rechnen mit floating point Zahlen), in der MATLAB [16] den Standard darstellt.

Wir wollen zunächst ausgehend von einfachen Fragestellungen einen kurzen Überblick über MAPLE geben, das sowohl in der Kombinatorik als auch in der Analysis eine umfangreihe Funktionalität bereitstellt.

5.4.2 Maple

MAPLE kann sowohl in der Kommandozeile als auch in einem graphischen Frontend verwendet werden. Die Ausgabe von Graphik ist natürlich nur in letzterem möglich, wobei die Kommandozeilenversion Graphiken in Dateien schreiben kann. In beiden Benutzeroberflächen folgt Output auf Input. Eine neue Zeile für mehrzeiligen Input erhält man durch `Shift-Return`, ein neues Eingabefeld durch `Strg-j`. Jeder Befehl wird mit einem Strichpunkt abgeschlossen und durch `Return` ausgewertet. Ersetzt man den Strichpunkt durch einen Doppelpunkt wird der Output unterdrückt. Durch `quit;` verlassen wir MAPLE.
Zuweisungen erfolgen mit:

```
i:=0;
0
```

Bedingte Anweisungen haben folgende Syntax:

```
if i=0 then print(null");fi;
null"
```

Mengen erzeugt man durch geschweifte Klammern:

```
M:={1,1,2,3,2};
M:={1,2,3}
```

und Listen durch eckige Klammern:

```
L:=[1,1,2,3,-1];
L:=[1,1,2,3,-1]
```

An eine Liste hängt man an durch

```
L:=[op(L),2];
L:=[1,1,2,3,-1,2]
```

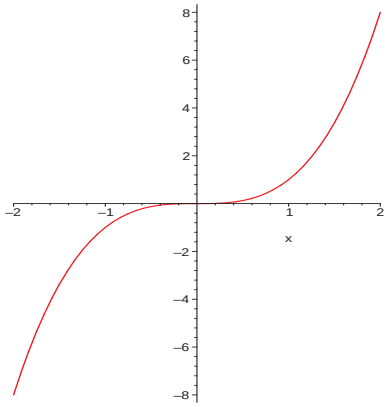
und genauso für Mengen.

Abbildungen (oder Prozeduren) werden auf die Elemente einer Menge oder die Einträge einer Liste angewendet durch:

```
map(x->x^2,L);
[1, 1, 4, 9, 1, 4]
```

Abbildungen $f : \mathbb{R} \to \mathbb{R}$ lassen sich plotten mit:

```
plot(x^3, x=-2..2);
```



Die Ausgabe wird nach dem Befehl

```
plotsetup(jpeg, plotoutput=‘plot.jpg‘, plotoptions
                          =‘portrait,noborder,color‘);
```

in eine Datei umgeleitet. Für eine Postscript-Ausgabe kann man jpeg durch ps ersetzen. Auf dem Bildschirm werden Plots wieder ausgegeben nach:

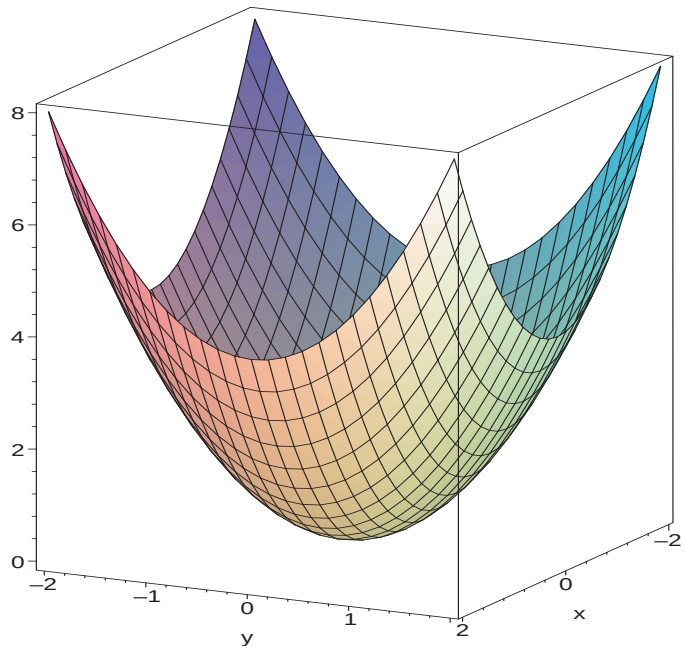
```
plotsetup(default);
```

Den Graphen der Abbildung

$$f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto x^2 + y^2$$

erhalten wir mit:

```
plot3d(x^2+y^2, x=-2..2,y=-2..2);
```



Bei der graphischen Ausgabe sind viele Optionen verfügbar, siehe dazu die Hilfe-Funktion unter `plot,options`.

Ein Beispiel für eine Prozedur, die

$$\sum_{k=1}^{n} k$$

berechnet ist (lokale Variablen werden mit `local` deklariert):

```
summe:=proc(n)
  local k,s;
  s:=0;
  for k from 1 to n do
      s:=s+k;
  od;
  return(s);
end proc:
```

Damit erhalten wir:

```
summe(5);
15
```

Tatsächlich gibt es eine Funktion die Summen und Produkte direkt auswertet:

```
sum(k,k=1..5);
```
15

gibt

$$\sum_{k=1}^{5} k = 1 + 2 + 3 + 4 + 5 = 15$$

und

```
product(k,k=1..5);
```
120

liefert

$$\prod_{k=1}^{5} k = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

Dies funktioniert (in vielen Fällen) auch für unbestimmte Grenzen:

```
sum(k,k=1..n);
```
$\frac{(n+1)^2}{2} - \frac{n}{2} - \frac{1}{2}$

Durch Vereinfachen mit der sehr mächtigen Funktion `simplify` sieht man, dass der Ausdruck mit der bekannten Formel übereinstimmt (wobei sich `%` auf die letzte Ausgabe bezieht):

```
simplify(%);
```
$\frac{1}{2}n^2 + \frac{1}{2}n$

Man kann auch Summenformeln eingeben, ohne sie auszuwerten

```
s:=Sum(k,k=1..n);
```
$s := \sum_{k=1}^{n} k$

damit weiterrechnen, z.B. n durch einen konkreten Wert ersetzen

```
s:=subs(n=5,s);
```
$s := \sum_{k=1}^{5} k$

und schließlich die Formel auswerten:

```
value(s);
```
15

In der Division mit Rest von a durch b mit Rest r

$$a = q \cdot b + r$$

erhalten wir q und r in MAPLE wie folgt, z.B. für $a = 36$ und $b = 15$:

```
iquo(36,15);
```
2
```
irem(36,15);
```
6

Diese Funktionen können Sie z.B. verwenden, um eine Prozedur zur Berechnung der Binärdarstellung zu schreiben. Vergleichen Sie auch mit der schon vorhandenen Funktion:

```
convert(23,binary);
```

```
10111
```

Weitere Anwendungsbeispiele werden wir jeweils in Zusammenhang mit den theoretischen Resultaten diskutieren.

Index

Literaturverzeichnis

[1] The Axiom Group: *Axiom*, http://www.axiom-developer.org/ (2012).

[2] J. Böhm: Grundlagen der Algebra und Zahlentheorie, Springer (2016).

[3] J. Böhm: Mathematik für Informatiker: Algebraische Strukturen, Lecture Notes (2018).

[4] J. Böhm, M. Marais: Introduction to algebraic structures, Lecture Notes (2019).

[5] J. Böhm: Mathematik für Informatiker: Kombinatorik und Analysis, Lecture Notes (2018).

[6] J. Böhm: Mathematik für Informatiker: Analysis, Lecture Notes (2019).

[7] O. Forster: *Analysis I*, Vieweg (2010).

[8] The GAP Group, *GAP – Groups, Algorithms, and Programming, Version 4.4.12*; http://www.gap-system.org, (2008).B. Kreußler, G. Pfister: Mathematik für Informatiker: Algebra, Analysis, Diskrete Strukturen, Springer (2009).

[9] Grayson, D. R.; Stillman, M. E.: *Macaulay2, a software system for research in algebraic geometry*, available at http://www.math.uiuc.edu/Macaulay2/ (2009).

[10] Bosma, W.; Cannon J.; Playoust C.: *The Magma algebra system. I. The user language*, J. Symbolic Comput., 24 (1997), 235–265.

[11] K. Königsberger: *Anaysis I*, Springer (2008).

[12] B. Kreußler, G. Pfister: *Mathematik für Informatiker: Algebra, Analysis, Diskrete Strukturen*, Springer (2009).

[13] Maple (Waterloo Maple Inc.): Maple 16, http://www.maplesoft.com/ (2012).

[14] Maxima: *Maxima, a Computer Algebra System.* Version 5.25.1, available at http://maxima.sourceforge.net/ (2011).

[15] Wolfram Research, Inc.: *Mathematica Edition: Version 7.0* (2008).

[16] MATLAB. Natick, Massachusetts: The MathWorks Inc., http://www.mathworks.de/products/matlab/ (2013).

[17] The On-Line Encyclopedia of Integer Sequences, http://oeis.org/.

[18] Hearn, A. C.: *REDUCE 3.8*, available at http://reduce-algebra.com/ (2009).

[19] Decker, W.; Greuel, G.-M.; Pfister, G.; Schönemann, H.: SINGULAR *4-1-1 — A computer algebra system for polynomial computations.* http://www.singular.uni-kl.de (2015).