Wahrscheinlichkeit – und andere Mysterien der Statistik

Autor: Rüdiger Stegen
20.02.2017 17:38:31

Inhalt

Vorwort

Nach meiner aktiven Berufszeit bei einem internationalen Energieversorgungsunternehmen habe ich mich im Sommer 2012 für einen Lehrauftrag für Wirtschaftsmathematik an der Hochschule Weserberg in Hameln beworben. Da ich Mathematik studiert hatte, schien mir diese Tätigkeit sehr attraktiv zu sein – und ich wurde nicht enttäuscht. Bei meinem Vorstellungsgespräch wurde ich gefragt, ob ich mir auch vorstellen könnte, einen Statistikkurs zu machen. Meine Antwort war klar: ich stehe für alles im Rahmen meiner Möglichkeiten zur Verfügung – aber bloß nicht für Statistik. Offenbar hatte ich schlechte Erinnerungen an dieses Fach aus meiner Studienzeit.

Als ich im Frühjahr 2013 nochmal gefragt wurde, dachte ich: ich kann es ja mal versuchen. Und siehe da – Statistik entpuppte sich als außerordentlich spannendes Thema, allerdings mit einer Besonderheit: es gibt eine ganze Reihe von Veröffentlichungen, die sich ausschließlich damit beschäftigen, wie statistische Methoden zu Fehlschlüssen verleiten können oder wie sie sogar bewusst für genehme Aussagen missbraucht werden (siehe z. B. [BK], [DH] oder [MP]). Etwas Vergleichbares kannte ich von anderen mathematischen Bereichen nicht. In der Konsequenz hieß das für mich, dass man statistischen Darstellungen kritisch begegnen sollte!

Dieses Buch ist auf der Basis eines Statistikkurses, den ich als Lehrbeauftragter seit WS 2013/14 an der Hochschule Weserbergland in Hameln halte, entstanden. Ursprünglich diente die Unterlage nur zur Erläuterung der Rechenverfahren bei den Übungsaufgaben. Im weiteren Verlauf stellte sich dann aber heraus, dass manches, was in Büchern oder im Internet zu lesen ist, nicht ausreichend oder nicht plausibel begründet und manchmal sogar falsch ist. Insbesondere grundlegende Begriffe wie Wahrscheinlichkeit, Zufall oder Experiment wurden in vielen Büchern und Darstellungen im Internet, die ich mir angesehen hatte, viel zu nebulös und zum Teil widersprüchlich definiert. Und demzufolge blieb oft unklar, was eine so einfache Aussage wie „beim Münzwurf beträgt die Wahrscheinlichkeit für Wappen 50 %“ konkret bedeutet – und darauf kommt es mir primär an.

Auf folgende Themen wird im vorliegenden Buch ausführlich eingegangen:

- Bei Klassierungen werden manchmal spekulative Annahmen gemacht, um bestimmte Zahlenwerte für das arithmetische Mittel und andere Parameter ableiten zu können, obwohl für die Ausgangsdaten nur Intervalle bekannt sind. Dabei bleibt aber offen, was diese Annahmen mit der Realität zu tun haben.
- Bei dem Versuch, den Begriff Wahrscheinlichkeit konkret zu definieren, werden oft nur verschiedene Interpretationsmöglichkeiten wie subjektive, objektive, axiomatische, … Wahrscheinlichkeit genannt, die den Leser mehr oder weniger ratlos zurücklassen. Insbesondere bleibt unklar, ob die Rechenregeln für alle diese Arten von Wahrscheinlichkeiten gleichermaßen gelten und ob es überhaupt eine übergreifende Definition gibt.
- Manchmal findet man sogar die Feststellung, dass leider nicht genau klar ist, was Wahrscheinlichkeit konkret bedeutet. Aber wenn man nicht weiß, was Wahrscheinlichkeit konkret bedeutet, dann weiß man auch nicht, was ein Satz der Art „die Wahrscheinlichkeit, eine 3 zu würfeln, beträgt $\frac{1}{6}$“ konkret bedeutet – solche Aussagen wären dann in der Praxis sinnlos. Und in der Konsequenz wären dann große Teile der Statistikbücher, die sich mit der konkreten Anwendung von Wahrscheinlichkeiten beschäftigen, ebenfalls sinnlos und damit hinfällig.
- Es bleibt oft unklar, was der Begriff Wahrscheinlichkeit in den Axiomen von Kolmogoroff mit dem umgangssprachlichen Begriff der Wahrscheinlichkeit zu tun hat. Wenn man einfach die axiomatische Definition unmittelbar auf praktische Fragestellungen anwendet, ignoriert

man, dass der Zusammenhang zwischen einem mathematischen Modell und der Realität erklärt werden muss, wenn man das mathematische Modell konkret anwenden will. Die Axiome von Kolmogoroff sind also kein Ersatz und für eine unzureichende oder sogar fehlende Definition der Wahrscheinlichkeit bei praktischen Fragestellungen. Kurz: man kann den Mangel an einer klaren praktischen Definition der Wahrscheinlichkeit nicht durch die Axiome ausgleichen.

- Formulierungen, die die Stochastik als die „Lehre von den Gesetzmäßigkeiten des Zufalls“ (siehe z. B. [GH], Seite 1) oder als die „Mathematik des Zufalls“ (siehe z. B. [BE], Vorwort) bezeichnen, führen in die falsche Richtung. Wie im vorliegenden Buch ausführlich erläutert wird, spielt der Zufall – mit Ausnahme bestimmter Beispiele – keine Rolle in der Stochastik.
- Und dann werden noch Themen behandelt wie:
    - Der „ideale“ oder „faire“ Würfel ist als Beispiel weit verbreitet, aber irreführend.
    - Elementarereignisse sind keine Ereignisse – oder doch?
    - Zufallsvariable sind keine Variable, sondern Abbildungen – oder umgekehrt?

In manchen Büchern wird beklagt, dass Stochastik besonders unbeliebt bei den Studierenden ist oder als besonders schwierig empfunden wird (siehe z. B. [HR], Vorwort). Die oben dargestellten Themen legen nahe, dass die Widersprüche und Unklarheiten in Büchern und anderen Veröffentlichungen daran eine gewisse Mitschuld tragen können. Das vorliegende Buch soll einen Beitrag dazu leisten, diese Widersprüche und Unklarheiten zu benennen und zu beseitigen. Dazu werden zu allen oben angesprochenen Themen ausführlich begründete Lösungen dargestellt.

Der erste Teil des Buches befasst sich mit beschreibender Statistik. Besonders breiten Raum nehmen die Klassierungen ein, bei denen ohne die üblichen spekulative Annahmen unverbesserbare Aussagen zu Lage- und Streuparametern abgeleitet werden.

Der zweite größere Teil befasst sich mit der Definition, Interpretation und Berechnung von Wahrscheinlichkeiten in realen praktischen Situationen. Damit wird die in vielen Büchern zu Recht beklagte Unsicherheit zu diesen Themen – insbesondere zu der Frage, was Wahrscheinlichkeit konkret bedeutet (siehe z. B. [HR], Seite 6) – weitgehend beendet.

Im Ergebnis wird

- eine allgemeine übergreifende Definition der Wahrscheinlichkeit plausibel hergeleitet, die alle praktischen Aspekte berücksichtigt. Diese Definition basiert auf dem umgangssprachlichen Verständnis und benötigt nicht Begriffe wie Zufall, Ereignis oder Experiment. Die oft gestellte Frage, was Wahrscheinlichkeit konkret bedeutet, wird damit beantwortet.
- ein allgemeines Verfahren zur Bestimmung elementarer Wahrscheinlichkeiten beschrieben, das alle Varianten einschließlich Versuchsreihen und einmalige Vorgänge berücksichtigt.
- dem Begriff Wahrscheinlichkeit in der Praxis alles Mystische und Mehrdeutige genommen, da die Wahrscheinlichkeit auch nur so etwas wie ein Qualitätsurteil über Mozzarella, der Friedlichkeitsindex von Staaten oder eine Schulnote ist. Wahrscheinlichkeit ist nicht klarer oder unklarer als andere Bewertungen.
- ein Zusammenhang zwischen der allgemeinen Definition für praktische Fragestellungen und der axiomatischen Definition nach Kolmogoroff hergestellt und gezeigt, dass die axiomatische Definition nur bedingt etwas mit dem praktischen realitätsbezogenen Begriff Wahrscheinlichkeit zu tun hat. Für reale Probleme braucht man die Kolmogoroffschen Axiome nicht, wenn man die praktische Definition aus dem vorliegenden Buch anwendet.

Grundsätzlich hat man zwei Möglichkeiten bei der Betrachtung von Wahrscheinlichkeiten:

- Entweder beschäftigt man sich mit konkreten praktischen Fragestellungen. Dann kann die Darstellung im vorliegenden Buch als Grundlage genommen werden. Das Axiomensystem nach Kolmogoroff oder Ähnliches wird nicht benötigt.
- Oder man beschäftigt sich mit dem mathematischen Teilgebiet Maßtheorie. Dann stützt man sich auf ein Axiomensystem wie die Kolmogoroffschen Axiome, in dem der Begriff Wahrscheinlichkeitsmaß durch einen neutralen Begriff wie „normiertes Maß“ ersetzt werden sollte, um Verwirrungen zu vermeiden (Siehe z. B. [HB], Seite 6.). Wenn man das konsequent umsetzt, dann kommt der Begriff Wahrscheinlichkeit nicht mehr vor.

Darüber hinaus werden bestimmte gelegentlich vorkommende Ungereimtheiten korrigiert, wobei manche Korrekturen für das Verständnis wichtig sind und andere eher formalen Charakter haben. Zur besseren Verständlichkeit werden die Erläuterungen durch viele Beispiele untermauert.

Der Vollständigkeit halber noch ein Hinweis: schließende Statistik ist nicht Gegenstand des vorliegenden Buches.

Die einzelnen Themen werden nur bezüglich ihrer Grundlagen behandelt, das aber möglichst umfassend. Die dabei verwendeten mathematischen Methoden können mit Oberstufenmathematik verstanden werden. Das Buch ist für alle hilfreich, die sich mit grundlegenden Begriffen und Methoden der Statistik beschäftigen und in realen Situationen verstehen wollen, also Studierende, Lehrende und Anwender.

Bezeichnungen

$[a, b] = \{x \mid a \leq x \leq b\}$ abgeschlossenes Intervall
$]a; b[= \{x \mid a < x < b\}$ offenes Intervall
$]a; b] = \{x \mid a < x \leq b\}$ links halboffenes Intervall
$[a; b[= \{x \mid a \leq x < b\}$ rechts halboffenes Intervall

$\mathbb{N}$ $= \{1, 2, 3, 4, \ldots\}$ Menge der natürlichen Zahlen
$\mathbb{N}_0$ $= \{0, 1, 2, 3, 4, \ldots\}$ Menge der natürlichen Zahlen mit 0
$\mathbb{Z}$ Menge der ganzen Zahlen
$\mathbb{Q}$ Menge der rationalen Zahlen
$\mathbb{R}$ Menge der reellen Zahlen

$\lceil x \rceil$ $\lceil x \rceil$ ist die kleinste ganze Zahl mit $x \leq \lceil x \rceil$ (Gauss-Klammern)

[AB] Hinweis auf die Veröffentlichung [AB] im Quellenverzeichnis am Ende des Buches

1 Einführung

Zusammenfassung
In diesem Kapitel wird dargestellt, wie sehr statistische Methoden durch die Ausbreitung des Internet an Bedeutung gewonnen haben. Darüber hinaus werden grundlegende Begriffe wie Statistik und Stochastik definiert und die Vorgehensweise bei der Erhebung und Auswertung von Daten erläutert.
* * *

1.1 Vorbemerkungen

Täglich liest man etwas über Statistiken oder darüber, was wahrscheinlich passieren oder nicht passieren wird. Ob bei Finanzen, Wirtschaftsthemen, Sport oder Wissenschaften – überall wird man mit solchen Aussagen konfrontiert.

Durch die extrem ausgeweitete Nutzung des Internets ist die Menge der verfügbaren Daten in den letzten Jahren explosionsartig gestiegen. Die „Herrschaft der Algorithmen“ (siehe z. B. [KJ] oder [ZN]) oder die „Datendiktatur“ (siehe [HD]) scheinen einen (über)mächtigen Einfluss zu haben. Längst werden die Möglichkeiten, die die Daten im Internet bieten, in großem Stil kommerziell genutzt. Unternehmen wie Google, IBM, SAP, Blue Yonder oder Predata (www.predata.com) bieten staatlichen oder nichtstaatlichen Organisationen entsprechende Unterstützung zur Analyse von Daten an.

Bei der Auswertung dieser Daten spielen statistische Methoden eine zentrale Rolle. Es werden Zusammenhänge zwischen Daten ermittelt und Wahrscheinlichkeiten für künftige Entwicklungen abgeleitet. Auch wenn man es oft nicht merkt: statistische Methoden haben einen starken und wachsenden Einfluss auf unser tägliches Leben. Dabei werden gerade auch die sozialen Netzwerke intensiv mit Social analytics-Methoden durchforstet. So wurde zum Beispiel in China ein Kleid, das eine Schauspielerin in einer Fernsehshow getragen hatte, in den sozialen Medien besonders positiv bewertet und daraufhin von einer Firma innerhalb weniger Tage in gebietsspezifischen Variationen auf den Markt gebracht (siehe [SD]).

Das vorliegende Buch befasst sich mit grundlegenden statistischen Methoden. Dabei wird besonderer Wert auf eine umfassend begründete Interpretation der Begriffe und Ergebnisse in der Praxis gelegt.

1.2 Definitionen

Statistik befasste sich ursprünglich nur mit Daten, die den Zustand und die Entwicklung von Staaten beschreiben. Nimmt man z. B. den Brockhaus von 1868 zur Hand, so findet man unter dem Begriff „Statistik“ folgende Darstellung (siehe [BF]):

„Statistik (vom neulat. statista*, Staatsmann) (…) Nach der gegenwärtig üblichen Auffassung versteht man darunter die Darstellung des Zustandes und der Veränderungen von Staaten und staatlichen Gebilden oder sonstigen menschlichen Gemeinschaften in einer Form, welche das Individuum verschwinden läßt und Gleichartiges zusammenfaßt. (…)*
… und die Gefahr falscher Schlußfolgerungen auf Grund unrichtiger Verbindung von Thatsachen so nahe liegt, daß man den paradoxen Satz aufstellen durfte: mit Hülfe der Statistik lasse sich alles beweisen, das Richtige wie das Verkehrte."

Im Meyers Konversationslexikon von 1909 (siehe [MK]) wird der Begriff Statistik bereits weiter gefasst und bezieht sich auch auf andere Gebiete wie *„Naturerscheinungen (Gewitter, Hagelschläge, etc.)"*.

In beiden Lexika taucht Stochastik als eigenständiger Begriff noch nicht auf.

Stochastik kommt aus dem Griechischen στοχαστικός (stochastikos) und bedeutet „im Erraten geschickt", στοχαστικὴ τέχνη (stochastike techne) bedeutet „Kunst des Ratens" oder „Kunst des Vermutens".

Im „Meyers enzyklopädisches Lexikon" von 1978 steht unter dem Begriff „Stochastik" (siehe [ME]):
„Stochastik [gr.], in der Statistik und Wahrscheinlichkeitsrechnung (…) die mit wahrscheinlichkeitstheoret. Methoden vorgenommene Untersuchung zufallsabhängiger Ereignisse (z. B. von Stichproben), im Gegensatz zur Methodik der deskriptiven Statistik."

Und im aktuellen Duden ist definiert (siehe [DU]):
„Stochastik – Teilgebiet der Statistik, das sich mit der Untersuchung vom Zufall abhängiger Ereignisse und Prozesse befasst".

Im Gegensatz dazu wird in manchen Büchern definiert, dass Stochastik der Oberbegriff von Wahrscheinlichkeitstheorie und Statistik sei (siehe z. B. [GH], Seite 3).

Nimmt man das ursprüngliche griechische Wort als Grundlage, so kann die beschreibende Statistik kein Teil der Stochastik sein. In der beschreibenden Statistik werden exakte Parameter ermittelt, so dass kein Platz zum „Raten" oder „Vermuten" bleibt.

(1.2-01) ***Definition***
Statistik (engl.: *statistics*) ist die wissenschaftliche Disziplin, die sich mit den Methoden der Gewinnung, Analyse und Interpretation von Daten beschäftigt. Sie besteht aus den beiden Bereichen „beschreibende (deskriptive) Statistik" und „schließende (induktive) Statistik".
In der beschreibenden Statistik werden bestimmte Parameter aus einem vollständigen Satz von Daten abgeleitet.
In der schließenden Statistik wird mit Hilfe der Wahrscheinlichkeitsrechnung aus den Eigenschaften einer Stichprobe auf die Eigenschaften der Gesamtheit geschlossen.
Eine Statistik (engl.: *statistic*) ist eine Aufstellung von gleichartigen Daten, also in der Regel einfach eine Tabelle oder eine daraus abgeleitete Grafik.
* * *

Wenn man also sagt: „Statistik ist interessant", dann ist die Wissenschaft gemeint. Wenn man aber sagt „diese Statistik ist interessant", dann ist eine Tabelle oder eine Grafik gemeint.

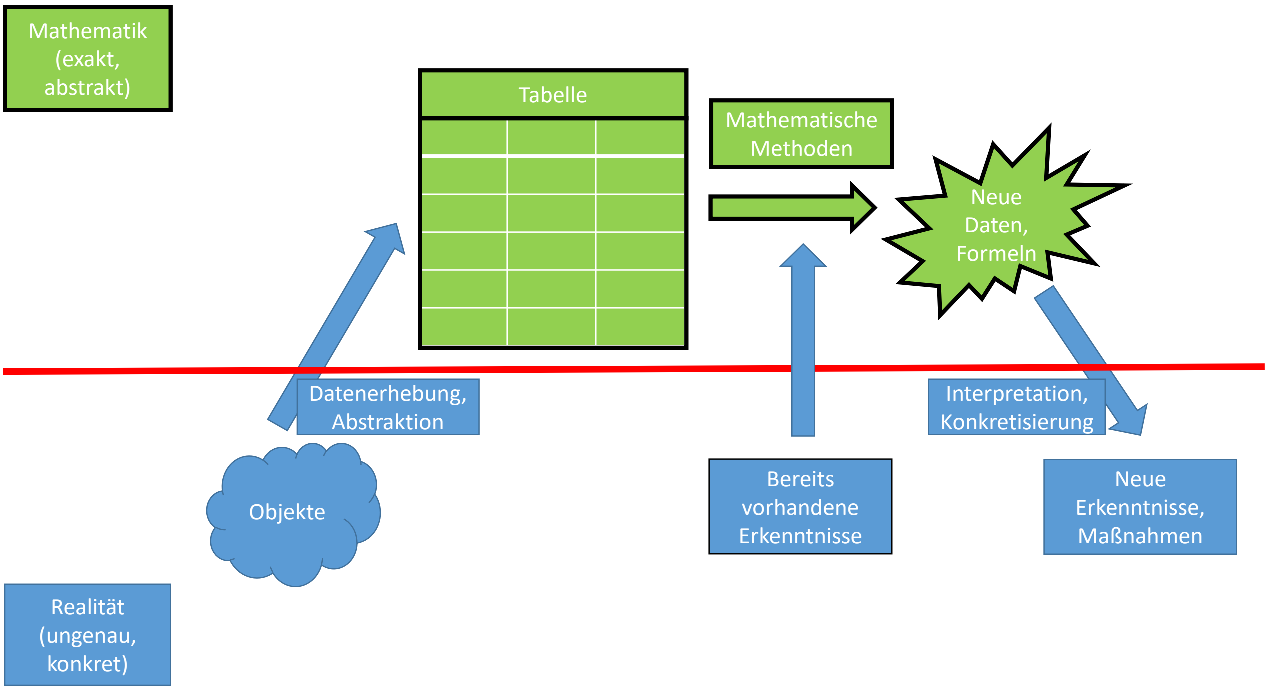
(1.2-02) ***Definition***
Stochastik ist der Oberbegriff von „Wahrscheinlichkeitsrechnung“ und „schließender Statistik“.
* * *

Der Begriff Wahrscheinlichkeitsrechnung wurde gewählt, um den praktischen Aspekt zu betonen, während Wahrscheinlichkeitstheorie innerhalb des mathematischen Gebietes Maßtheorie behandelt wird.

Im vorliegenden Buch beschäftigen wird uns mit Grundlagen der beschreibenden Statistik und der Wahrscheinlichkeitsrechnung (ab Kapitel 5), aber nicht mit der schließenden Statistik.

1.3 Allgemeine Vorgehensweise

Will man wie in den Beispielen in Kapitel 1.1. mit statistischen Methoden neue Erkenntnisse bei praktischen Problemen erlangen, so kann man folgendermaßen vorgehen:



Für die fachliche Planung geht man in der Grafik schrittweise rückwärts, d. h., man definiert erst, welche Fragestellung man beantworten möchte und ermittelt dann, was man dafür tun muss. Dabei geht man in folgenden Schritten vor:

Schritt 1:

Zunächst wird die zu untersuchende Fragestellung möglichst genau festgelegt: zu welcher Fragestellung möchte man neue Erkenntnisse haben, um dann ggf. Maßnahmen einleiten zu können?
Diese Fragestellung muss letztlich mathematisch formuliert werden, d. h., man wünscht bestimmte neue Daten oder Formeln, die man dann für die Praxis interpretieren kann. Typische Fragestellungen sind: „Wie groß ist die Wahrscheinlichkeit, dass …?“ oder „Welcher Zusammenhang besteht zwischen …?“.

Schritt 2:
Im zweiten Schritt wird festgelegt, welche Daten man zur Ableitung der gewünschten neuen Daten oder Formeln benötigt. Diese Daten werden meistens in Form von Tabellen notiert, in denen alle relevanten Informationen enthalten sind. Aus diesen Tabellen werden die gewünschten neuen Daten oder Formeln mit Hilfe mathematischer Methoden abgeleitet, wobei ggf. bereits vorhandene Erkenntnisse genutzt werden.
Die bereits vorhandenen Erkenntnisse müssen mathematisch formuliert werden, damit sie für die Ableitung neuer Daten oder Formeln genutzt werden können. Beispielsweise kann man in manchen Fällen aufgrund von entsprechenden Erfahrungen ausgehen, dass der Zusammenhang zwischen zwei Größen einem bestimmten Formeltyp wie z. B. einer Normalverteilung entspricht. Die Tabelle dient dann dazu, die Parameter des Formeltyps zu bestimmen.

Schritt 3:
Im dritten Schritt wird festgelegt, wie die Datenerhebung zu erfolgen hat. Dazu definiert man die Grundgesamtheit, also die Menge aller Objekte, die man in die Untersuchung einbeziehen will, und die zu erhebenden Eigenschaften dieser Objekte.

Da sich Untersuchungsobjekte mit der Zeit verändern können, muss bei der Untersuchung auch ein Zeitrahmen festgelegt und/oder der Untersuchungszeitpunkt für jedes Untersuchungsobjekt in der Tabelle gespeichert werden. Wichtig ist, dass auch ein Außenstehender eindeutig erkennen kann, welche Objekte zu welchem Zeitpunkt bezüglich welcher Eigenschaften untersucht wurden.

Je nach Fragestellung und Aufwand gibt es folgende Möglichkeiten, um zu Daten zu kommen:

Eine neue Datenerhebung ist nicht erforderlich, wenn die Daten in ausreichendem Umfang, Detaillierungsgrad, Korrektheit und Aktualität bereits vorliegen („desk research“). Eine wichtige externe Quelle für Daten, die Deutschland betreffen, ist das statistische Bundesamt. Die wichtigste unternehmensinterne Quelle sind die sowieso vorhandenen Datenbestände über Personal, Produkte, Verkaufszahlen, Kosten, etc.

Reichen die vorhandenen Daten dagegen nicht aus, so müssen sie neu erhoben werden („field research“). Typische Methoden sind Befragungen von Kunden oder Mitarbeitern, Messungen von physikalischen Größen wie z. B. Länge oder Gewicht von Bauteilen oder Beobachtungen des Verhaltens von Menschen oder Tieren.

Die Genauigkeit der Daten kann sehr unterschiedlich sein:

Bei Befragungen zu Gefühlen oder Meinungen sind Fragen und Antworten eher unscharf. Eine Skala z. B. von 0 bis 10 suggeriert dann oft eine Genauigkeit, die tatsächlich nicht gegeben ist.
Bei Messungen physikalischer Größen wie Länge, Zeit oder Masse kann man dagegen in Abhängigkeit von der Messmethode eine hohe Genauigkeit erzielen.
Und Daten wie z. B. Anzahl Mitarbeiter oder Geldbeträge sind in der Regel absolut exakt.

Wichtig ist auch, dass die Daten nach vergleichbaren Methoden erhoben werden. Mischt man Datenbestände aus verschiedenen Quellen, so kann es zu Unverträglichkeiten kommen, auch wenn dieselben Begriffe benutzt werden. Physikalische Größen können noch ineinander umgerechnet werden, wie km in Meilen, aber eine Arbeitslosenquote kann in unterschiedlichen Staaten oder bezogen auf verschiedene Jahre unterschiedlich berechnet werden. So stiegen die Arbeitslosenzahlen in Deutschland von 2004 auf 2005 unter anderem dadurch, dass die offizielle Definition des Begriffs Arbeitslosigkeit geändert wurde (siehe [BP]). Begriffe müssen also normiert sein, um sie in einem Datenbestand sinnvoll zusammenführen zu können.

Bei der Datenerhebung kann man zwischen Vollerhebung und Stichprobe wählen.
Bei einer Vollerhebung werden alle relevanten Daten der kompletten Grundgesamtheit erhoben. Diese Methode lässt exakte Aussagen über die Grundgesamtheit zu, kann aber sehr aufwendig sein. Viele Untersuchungen in Unternehmen beruhen auf Vollerhebungen, da die Daten sowieso vorhanden sind, weil sie z. B. zur Steuerung der Prozesse oder aufgrund gesetzlicher Vorgaben erforderlich sind. Vollerhebungen sind Grundlage der beschreibenden Statistik.
In manchen Fällen sind Vollerhebungen aber nicht möglich oder nicht sinnvoll, so dass man sich mit Stichproben zufriedengeben muss. Ein Grund kann sein, dass die Durchführung einer Vollerhebung praktisch nicht durchführbar oder nicht finanzierbar ist, wie z. B. die Befragung aller Bundesbürger zu einem Thema innerhalb eines kurzen Zeitraums. Ein anderer Grund kann sein, dass das Objekt bei der Untersuchung zerstört wird, wie z. B. bei Crashtests in der Automobilindustrie, die Untersuchung einer Charge in der Pharmaindustrie oder die „zerstörende Prüfung“ gemäß Fertigpackungsverordnung (siehe [FE]).
Stichproben lassen Aussagen über die Grundgesamtheit nur mit einer gewissen Wahrscheinlichkeit zu. Das ist Thema der schließenden Statistik und wird im vorliegenden Buch nicht weiter behandelt.

Da Daten im Allgemeinen in Computern gespeichert und verarbeitet werden, ist es zweckmäßig, sich vorher mit den erforderlichen Grundlagen der Informatik vertraut zu machen, insbesondere mit Fragen zum Design von Datenstrukturen und Datentypen, sowie von Erfassungs- und Auswerteprogrammen – soweit man nicht Standardsoftware einsetzt.

Die Grafik oben und die Erläuterungen machen deutlich: obwohl ab der Auswertung der Tabelle (hoffentlich) korrekte mathematische Verfahren verwendet werden, bestehen bei der Gewinnung von Basisdaten und bei der Interpretation der mathematisch gewonnenen Ergebnisse Spielräume. Diese subjektiven Spielräume spielen auch in der Kapiteln 4 und 5 eine wichtige Rolle.

Statistische Ergebnisse sind also immer mit Unsicherheiten behaftet. Die Übergänge „Realität → Mathematik“ (Datenerhebung) und „Mathematik → Realität“ (Interpretation) sind kritisch. Diese Beziehung zwischen Mathematik und Realität wird auch in Kapitel 8.1 bei der Betrachtung der Kolmogoroffschen Axiome von grundlegender Bedeutung sein.

2 Beschreibende Statistik

Zusammenfassung
In der beschreibenden Statistik liegen alle relevanten Daten der Grundgesamtheit vor, also nicht nur eine Stichprobe. Wie diese Daten konkret gewonnen wurden, wird hier nicht näher betrachtet – sie sind einfach da.
Aus diesen Datenbeständen werden Lage- und Streuparameter angeleitet, um zunächst einen groben Überblick über die Lage und Verteilung der Daten zu erhalten. Besonders ausführlich werden Klassierungen untersucht, bei denen für die Ausgangsdaten statt exakter Werte nur Intervalle vorliegen. In diesem Fall lassen sich trivialerweise auch für alle daraus abgeleiteten Parameter nur Intervalle angeben – im vorliegenden Kapitel sind das exakte unverbesserbare Intervalle. Die in manchen Büchern benutzten spekulativen Annahmen über die Lage der einzelnen Werte werden kritisch untersucht.
* * *

2.1 Grundbegriffe der beschreibenden Statistik

2.1.1 Merkmale, Skalen

Wie im Kapitel 1.3 bereits dargestellt, muss man zur Speicherung von Daten geeignete Datenstrukturen vorsehen. In der folgenden Tabelle werden dazu die wesentlichen Begriffe aus Statistik und Informatik gegenübergestellt. Trotz unterschiedlicher Begriffe werden in vielen Fällen in Statistik und Informatik ganz ähnliche Sachverhalte beschrieben.

(2.1.1-01) ***Tabelle***

Statistik	Informatik	Grafik in Kapitel 1.3	Beispiel
Eine Statistik	Datei, Tabelle	Tabelle	Verkaufsstatistik
Statistische Masse (Grundgesamtheit)	Menge von Objekten	Ausschnitt aus der Realität (Objekte)	Menge der verkauften Produkte
statistische Einheit (Merkmalsträger)	Einzelnes Objekt, ggf. identifiziert durch einen Schlüssel	Einzelne Zeile in der Tabelle	Einzelnes verkauftes Produkt, formal identifiziert durch eine Produktnummer
(statistisches) Merkmal	Attribut	Spaltenüberschrift in der Tabelle	Produktbezeichnung
Merkmalsausprägungen	Mögliche gültige Feldinhalte (bei der Erfassung zu prüfen)		Alle gültigen Produktbezeichnungen oder Verkäufernamen
Beobachtungswerte, Messwerte	Tatsächliche Feldinhalte		Produkt „ABC“, Verkäufer „XYZ“

Attribute müssen beschrieben werden, da sie je nach Kontext etwas anderes bedeuten können. So kann „Hamburg“ den Geburtsort oder den Wohnort einer Person bezeichnen oder eine Zahl wie 38100 kann ein Geldbetrag, eine Anzahl oder die Postleitzahl von Braunschweig sein.

Zur Beschreibung der Attribute zählen auch angemessene Datentypen.
Analog zu den Datentypen in der Informatik gibt es auch unterschiedliche Datentypen in der Statistik. Unterschiedliche Datentypen erlauben unterschiedliche Operationen und werden auf unterschiedlichen Skalen dargestellt. Mit Ortsnamen oder Farben kann man nicht rechnen, man kann sie höchstens aufzählen und angeben, wie häufig sie in einem bestimmten Kontext vorkommen. Weitaus mehr Möglichkeiten gibt es natürlich bei Geldbeträgen: sie können positiv oder negativ sein, man kann sie addieren, subtrahieren, mit Zahlen multiplizieren oder einen Mittelwert bilden.

In der Informatik haben Felder in Dateien ein bestimmtes Format, da davon abhängt, wie die gespeicherten Bits zu interpretieren sind und welche Operationen man damit durchführen kann. Ein bestimmtes Bitmuster kann je nach Kontext z. B. ein ASCII-Zeichen, eine Binärzahl oder Teil einer Instruktion sein. Entsprechend gehört in der Statistik zu einem Feldinhalt einer Tabelle auch eine Beschreibung, um den Inhalt richtig interpretieren zu können. Durch die Datentypen wird dann festgelegt, welche Operationen sinnvoll sind.

Manche Datentypen ergeben sich unmittelbar aus dem Inhalt: „Hamburg“ ist qualitativ und mit einem solchen Wert kann nicht gerechnet werden. Anders sieht es z. B. bei Zahlen aus: „38100“ kann ein Geldbetrag oder eine Anzahl sein – dann kann man damit rechnen – oder eine Postleitzahl – dann kann man damit nicht sinnvoll rechnen.

In der folgenden Tabelle wird dargestellt, welche Datentypen es in der Statistik gibt und welche Operationen sinnvoll sind. Damit diese Operationen sinnvoll sind, müssen die Daten nach einem einheitlichen Verfahren bestimmt worden sein. Es wäre nicht sinnvoll, wenn man z. B. Farben bei unterschiedlichen Beleuchtungsverhältnissen feststellt oder Entfernungen mal in km und mal in Meilen misst.

Die Begriffsbildungen sind in der Literatur nicht einheitlich, siehe z. B. [OM], Kapitel 2.2.2. oder [SJ], Seite 22/23.

(2.1.1-02) ***Tabelle***

Klassifikation des Merkmals	**Ausprägungen werden dargestellt durch**	**Rangordnung**	**Abstände von Ausprägungen sinnvoll**	**Skala**
Qualitativ	Begriffe, Nummernsysteme	Nein	Nein	Nominalskala
Komparativ	Begriffe, Nummernsysteme	Ja	Nein	Ordinalskala
Quantitativ	Zahlen	Ja	Ja	Metrische Skalen: Intervallskala, Verhältnisskala,

				Absolutskala

(2.1.1-03) ***Erläuterungen***

„Qualitativ“ (qualis (lat.) = wie beschaffen):

Bei qualitativen Merkmalen kann man zwei Feldinhalte auf Gleichheit prüfen und die Häufigkeit einer bestimmten Ausprägung feststellen, andere Operationen sind nicht möglich.
Beispiele: Farben, Geschlecht, Beruf, Religion, Familienstand, Wohnort, Blutgruppe

„Komparativ“ (comparare (lat.) = vergleichen).

Komparative Merkmale sind qualitative Merkmale, bei denen man zusätzlich die Ausprägungen in eine sinnvolle Reihenfolge bringen. Grundlage für diese Reihenfolge ist eine qualitative Bewertung (siehe auch Kapitel 4), nicht aber z. B. einfach nur eine alphabetische Reihenfolge.
Beispiele: Firmenrating AAA, AA+, …; Abschnitte eines zeitlichen Ablaufes, wie z. B. Reifegrad von Obst; T-Shirt-Größen S, M, L, XL, XXL, …; Sprachniveaustufen A1, A2, B1, … (siehe [ER]).

„Quantitativ“ (quantus (lat.) = wie groß):

Quantitative Merkmale kann man zahlenmäßig in einer bestimmten Maßeinheit messen. Man kann Ausprägungen sinnvoll addieren und subtrahieren und mit Zahlen multiplizieren. Quantitative Merkmale lassen sich auch immer vergleichen.
Beispiele: Physikalische Größen wie Länge, Fläche, Volumen, Gewicht, Zeit; kaufmännische Größen wie Preis, Menge, Kosten, Zinssatz; relative Anteile wie relativer Flächenanteil, relativer Zeitanteil, relativer Kostenanteil.

„Nummernsysteme“:

Nummernsysteme sind Zahlensysteme, die zur Identifikation eines Objektes benutzt werden, mit denen aber nicht sinnvoll gerechnet werden kann. So lassen sich z. B. zwei Bankleitzahlen nicht sinnvoll addieren.
Beispiele: Bankleitzahlen, Postleitzahlen, Telefonnummern, Personalnummern.

„Rangordnung“:

Eine Rangordnung von Merkmalsausprägungen liegt vor, wenn man die Ausprägungen nach einem bestimmten Kriterium sinnvoll sortieren kann.
Beispiele: siehe Erläuterungen zu „Komparativ“

„Abstände von Ausprägungen sinnvoll“:

Es gibt drei mögliche Fälle:
Im ersten Fall gibt es keinen Abstand zwischen Ausprägungen. Dies ist zum Beispiel bei Namen, Steuerklassen oder Sprachniveaustufen so.
Im zweiten Fall kann man zwar einen Abstand ausrechnen, aber er ist nicht sinnvoll interpretierbar. Das ist z. B. bei den Plätzen in einer Bundesligatabelle so. Der Abstand von Tabellenplatz 1 zu Tabellenplatz 5 ist rechnerisch genauso groß wie der Abstand von Tabellenplatz 14 zu Tabellenplatz 18, aber das ist nicht sinnvoll interpretierbar.
Im dritten Fall kann man problemlos Abstände ermitteln und interpretieren, wie zum Beispiel bei physikalischen Größen wie Länge, Zeit oder Masse oder bei kaufmännischen Größen wie Kosten oder Gewinn.

„Intervallskala“:

Eine Intervallskala ist eine Skala für quantitative Größen, bei denen der Quotient zweier Werte nicht sinnvoll interpretierbar ist.
Beispiele:
Jahreszahlen sind quantitative Größen, der Quotient zweier Jahreszahlen ist aber nicht sinnvoll, auch wenn man ihn rechnerisch bilden kann. Formulierungen wie „doppelt so großes Geburtsjahr" sind nicht sinnvoll interpretierbar. Ähnliches gilt für Temperaturen: „heute (10°C) ist es doppelt so warm wie gestern (5°C)" wird man kaum sagen.

„Verhältnisskala":
Im Gegensatz zur Intervallskala ist bei einer Verhältnisskala der Quotient zweier Werte sinnvoll interpretierbar.
Beispiele:
Aussagen wie „Die Birke ist halb so hoch wie die Eiche" oder „ich habe doppelt so viel auf dem Bankkonto wie du" sind sinnvoll interpretierbar.

„Absolutskala":
Eine Absolutskala ist ein Sonderfall der Verhältnisskala mit der Zusatzeigenschaft, dass die Ausprägungen keine Dimension (also nicht m, sec, kg, …) tragen.
Beispiel: absolute und relative Häufigkeiten
* * *

Ausgangspunkt aller statistischen Betrachtungen ist die ursprüngliche Tabelle der erhobenen Daten.

(2.1.1-04) ***Definition***
Gegeben sei eine Menge von Merkmalsträgern, bei denen bestimmte Merkmale interessieren, zusammen mit den jeweiligen Beobachtungswerten. Die zugehörige Datenstruktur ist

(Merkmalsträger, Beobachtungswert Merkmal 1, …, Beobachtungswert Merkmal n)

Die Datei mit dieser Datenstruktur heißt **Urliste** – oft enthält sie sogar nur die Beobachtungswerte ohne die Merkmalsträger. Merkmalsträger ist der identifizierende Begriff.
* * *

(2.1.1-05) ***Beispiel***
Im einfachsten Fall hat man nur den Merkmalsträger und ein Merkmal, wie z. B. folgender Ausschnitt aus einer Personaldatei:

Personalnummer	Gehalt (€)
01	2.400
02	3.600
03	3.000
04	5.100
05	4.300
06	3.000

Wenn sogar nur die Gehälter interessieren, so kann man diese Tabelle anonymisieren und die Spalte „Personalnummer" weglassen.

(2.1.1-04) ***Definition***
Liegt die Urliste für ein komparatives oder quantitatives Merkmal vor, so kann man die Beobachtungswerte gemäß der Rangordnung sortieren. Eine solche sortierte Reihe von Beobachtungswerten heißt **Variationsreihe**.
* * *

Eine solche Umsortierung sollte aber nur vorgenommen werden, wenn die ursprüngliche Reihenfolge für die weitere Betrachtung unwesentlich ist. Will man das arithmetische Mittel oder den Median (siehe Kapitel 2.2) der Umsätze der letzten zehn Jahre darstellen, so kommt es auf die Reihenfolge der einzelnen Umsätze nicht an. In diesem Fall kann man problemlos eine Variationsreihe bilden und daraus das arithmetische Mittel und den Median ableiten. Will man aber die Umsatzentwicklung der letzten 10 Jahre darstellen, so ist die Reihenfolge wichtig, die Werte dürfen nicht umsortiert werden.

2.1.2 Diskret / stetig

Da man in der Praxis immer nur mit einer bestimmten Genauigkeit, innerhalb bestimmter Grenzen und mit einer beschränkten Häufigkeit messen kann, liegen die einzelnen Ausprägungen isoliert voneinander. Das gilt für alle Arten von Merkmalen. Insbesondere haben Ausprägungen quantitativer Merkmale in der Realität stets einen bestimmten von der Messmethode abhängigen Mindestabstand.
Nimmt man z. B. einen Zollstock, so kann man Längen nur auf einen Millimeter genau messen. Alle Messwerte haben einen Mindestabstand von 1 mm, sofern sie nicht gleich sind. 123,4 mm kann als Messergebnis mit dieser Methode nicht vorkommen. Ähnliches gilt für beliebige andere Messungen.
Und da die Anzahl der Messwerte in der Realität beschränkt ist, gibt es auch für beliebige abgeleitete Merkmale nur endlich viele verschiedene Werte, die dann auch einen Mindestabstand haben.
Oder mathematisch: Die Menge der von 0 verschiedenen Abstände von Merkmalswerten ist endlich und hat damit ein Minimum > 0.

(2.1.2-01) ***Definition***
Ein quantitatives Merkmal, bei dem die Ausprägungen einen Mindestabstand voneinander haben, heißt **diskret**.
* * *

Hat man viele dicht beieinanderliegende Ausprägungen eines quantitativen Merkmals, so kann es mathematisch einfacher sein, dieses Merkmal näherungsweise wie ein kontinuierliches Merkmal zu behandeln. Dann sind also rechnerisch auch alle Zwischenwerte von tatsächlich messbaren Werten möglich und man kann Methoden der Analysis, wie die Differential- und Integralrechnung, anwenden.
Typische Beispiele für stetige Merkmale sind physikalische Größen wie Temperatur, Länge oder Zeitdauer. Aber z. B. auch die zeitliche Entwicklung von Aktienkursen wird oft als kontinuierliche Kurve dargestellt, obwohl es sich tatsächlich um einzelne Werte mit einem bestimmten zeitlichen Mindestabstand handelt.

(2.1.2-02) ***Definition***
Ein quantitatives Merkmal mit kontinuierlichen Werten heißt **stetig**.

* * *

In der Realität gibt es keine stetigen Merkmale, man benutzt sie nur, wenn man damit rechnerisch einfacher umgehen kann und die Näherung ausreichend gut ist.
Die Entscheidung, ob es zweckmäßig ist, ein Merkmal durch eine stetige Größe anzunähern, hängt davon ab, wie viele Messwerte man hat, wie dicht sie liegen, wie hoch die geforderte Genauigkeit bei der Analyse der Werte ist und was mathematisch einfacher zu behandeln ist. Durch den Einsatz von Computern ist die Notwendigkeit, stetige Merkmale zu nutzen, nicht mehr so hoch wie früher. Mit dem Computer kann man große Mengen diskreter Daten untersuchen, wo man früher einfacher ein stetiges Merkmal als Näherung benutzt hätte.

Im Folgenden wird wie allgemein üblich unter einem stetigen Merkmal eine ausreichend gute Annäherung an ein reales und damit diskretes Merkmal verstanden.

(2.1.2-03) ***Bemerkung***
In manchen Büchern wird eine andere Definition der Begriffe verwendet (siehe z. B. [BG], Seite 11/12), die innerhalb der Mathematik von Interesse sein kann, aber keine praktische Bedeutung hat, da es in der Praxis sowieso nur endlich viele mögliche Messwerte gibt:
Eine Menge ist **abzählbar**, wenn die Elemente abgezählt, also vollständig mit natürlichen Zahlen durchnummeriert werden können (x_1, x_2, x_3, …), ansonsten heißt die Menge nicht abzählbar oder **überabzählbar**.
Wichtigste Beispiele:

- Die Menge der rationalen Zahlen $\mathbb{Q}$ und jede ihrer endlichen oder unendlichen Teilmengen wie z. B. die Menge der natürlichen Zahlen $\mathbb{N}$ sind abzählbar.
- Beliebige Intervalle reeller Zahlen mit Intervalllänge > 0 sind nicht abzählbar, ebenso jede Obermenge eines solchen Intervalls, wie z. B. die gesamte Menge der reellen Zahlen $\mathbb{R}$.

Ein Merkmal heißt **diskret**, wenn es abzählbar viele Ausprägungen gibt.
Ein Merkmal heißt **stetig**, wenn es überabzählbar viele Ausprägungen gibt.
* * *

Hinweis: diese Definition von Stetigkeit darf nicht mit dem aus der Analysis bekannten Begriff der Stetigkeit einer Funktion verwechselt werden.

2.1.3 Aussagen

Bei Häufigkeiten und Wahrscheinlichkeiten (Siehe Kapitel 5) spielt der Begriff „Aussage“ eine zentrale Rolle.

Eine Aussage ist ein sprachliches Gebilde, für das es sinnvoll ist, es als wahr oder falsch zu bezeichnen. Eine Aussage muss genau genug sein, damit man sie als wahr oder falsch bezeichnen kann. Ein Hinweis auf eine unzureichende Genauigkeit ist, wenn die Aussage und ihr Gegenteil gleichzeitig wahr sein könnten.

(2.1.3-01) ***Beispiel***
Die Aussage „es regnet“ ist sehr ungenau. Die Aussage „es regnet und es regnet nicht“ ist immer wahr, da beides vorkommen kann, solange man nichts über Ort und Zeit aussagt. Dagegen ist „hier und jetzt regnet es und regnet es nicht“ eine falsche Aussage, da beides nicht gleichzeitig am selben Ort passieren kann. Umgangssprachlich sagt man auch „jein“ oder „kommt darauf an“, wenn eine Fragestellung nicht präzise genug ist.
* * *

Man kann Aussagen kombinieren und erhält dann wieder Aussagen. Die wichtigsten logischen Operationen bei Aussagen sind „und“, „oder“ und „nicht“. Sind A (z. B.: „ich habe ein Auto“) und B (z. B.: „ich habe ein Fahrrad“) zwei Aussagen, so gilt für die logischen Operationen:

- „und“ ist gleichbedeutend mit „sowohl – als auch“.
  Beispiel: „ich habe ein Auto und ein Fahrrad“ bedeutet: ich habe sowohl ein Auto, als auch ein Fahrrad.
- „oder“ ist das „nicht ausschließende oder“, d. h., es können auch A und B gleichzeitig wahr sein. Das „ausschließende oder“ entspricht „entweder – oder“ und ist hier nicht gemeint.
  Beispiel: „ich habe ein Auto oder ein Fahrrad“ bedeutet also: ich habe entweder nur ein Auto (aber kein Fahrrad) oder nur ein Fahrrad (aber kein Auto) oder ein Auto und ein Fahrrad (also beides).
- „nicht“ ist die Verneinung oder das Gegenteil; statt „nicht A“ schreibt man auch kurz $\overline{A}$; $\overline{A}$ nennt man auch die Gegenaussage von A.
  Beispiel: $\overline{A}$ bedeutet so viel wie „ich habe kein Auto“.

(2.1.3-02) ***Regeln***
Für beliebige Aussagen A, B, C gelten die Regeln der Aussagenlogik:

A und B	= B und A
A oder B	= B oder A
A und (B oder C)	= (A und B) oder (A und C)
A und (B und C)	= (A und B) und C = A und B und C
A oder (B oder C)	= (A oder B) oder C= A oder B oder C
A und $\overline{A}$	ist falsch
A oder $\overline{A}$	ist wahr

* * *

Wählt man als Beispiel
A = „ich habe ein Auto“
B = „ich habe ein Fahrrad“
C = „ich habe ein Motorrad“
so sieht man leicht, dass die Regeln dem umgangssprachlichen Gebrauch entsprechen.

2.1.4 Häufigkeiten und relative Anteile

(2.1.4-01) ***Definition***

Hat man eine Aussage A, so gibt die **absolute Häufigkeit** h(A) an, für wie viele Fälle diese Aussage zutrifft. Dabei sind der Gültigkeitsbereich der Untersuchung und das genaue Verfahren, wie man die Richtigkeit der Aussage feststellt, anzugeben. Wurden im Rahmen dieser Untersuchung insgesamt n Fälle betrachtet, so bezeichnet $f(A) = \frac{h(A)}{n}$ die **relative Häufigkeit** der Aussage A.
* * *

(2.1.4-02) ***Hinweise***

- Der Gültigkeitsbereich beinhaltet oft eine zeitliche und räumliche Eingrenzung.
- h(A) und f(A) sind dimensionslose Größen. Sie können auf einer Absolutskala dargestellt werden.
- Bei der relativen Häufigkeit muss mehr bekannt sein, als bei der absoluten Häufigkeit. Wenn man z. B. die Anzahl der roten Autos, die in einem bestimmten Zeitraum an einer bestimmten Stelle vorbeifahren, ermitteln will, so muss man die anderen Autos nicht explizit registrieren. Die Menge aller Autos, die in dem bestimmten Zeitraum an der bestimmten Stelle vorbeifahren, kann unbekannt bleiben, da nur die roten Autos interessieren. Will man dagegen den relativen Anteil der roten Autos feststellen, so muss man zusätzlich die gesamte Anzahl Autos kennen.

(2.1.4-03) ***Beispiel***

Für die Veröffentlichung im Geschäftsbericht soll die Anzahl der Frauen in einem Unternehmen angegeben werden.

- Aussage: eine Person ist eine Frau
- Gültigkeitsbereich: Alle Mitarbeiter/innen des Unternehmens, deren Daten zu einem bestimmten Zeitpunkt (z. B. zum Jahresende) im Personalstamm enthalten sind
- Verfahren: Prüfen des Attributs „Geschlecht“ im Personalstamm
- Absolute Häufigkeit: Anzahl Frauen
- Relative Häufigkeit: Anzahl Frauen im Verhältnis zu n = Anzahl Datensätze im Personalstamm

* * *

(2.1.4-04) ***Beispiel***

Es wird 100-mal mit einem Würfel gewürfelt und die Anzahl Sechsen notiert.

- Aussage: die gewürfelte Augenzahl ist eine „6“
- Gültigkeitsbereich: die vorgegebene Versuchsreihe
- Verfahren: Würfeln, wobei nur Ergebnisse berücksichtigt werden, bei denen der Würfel eindeutig ablesbar auf der Unterlage liegen bleibt, in allen anderen Fällen wird der Wurf wiederholt.
- Absolute Häufigkeit: Anzahl Sechsen.
- Relative Häufigkeit: Anzahl Sechsen geteilt durch n = 100.

* * *

(2.1.4-05) ***Beispiel***

Es wird gezählt, wie oft ein Redner bei einer Ansprache „äh“ gesagt hat. Die absolute Häufigkeit ist einfach die Anzahl der Ähs, aber für eine sinnvolle relative Häufigkeit muss eine entsprechende Bezugsgröße ermittelt werden. Wenn man (was sicherlich realitätsfern ist) alle Worte der Ansprache mitzählt und „äh“ als Wort definiert, so kann man daraus die relative Häufigkeit der Ähs ableiten.
* * *

(2.1.4-06) ***Beispiel***
„Du hast heute schon dreimal gesagt, dass das Wetter schlecht ist". Die absolute Häufigkeit ist 3, aber eine sinnvolle relative Häufigkeit ergibt sich nicht unmittelbar.
* * *

Häufig betrachtet man den Sonderfall, bei dem sich die Aussage A auf die Ausprägungen eines bestimmten Merkmals von Objekten bezieht.
Gegeben ist eine endliche Menge G von Objekten mit $|G| = n$.
A ist die Teilmenge von G, deren Elemente bestimmte Ausprägungen haben, alle anderen Elemente von G haben diese Ausprägungen nicht.
Dann ist
$h(A) = |A|$
$f(A) = \frac{|A|}{|G|} = \frac{|A|}{n}$
Die Beispiele (2.1.4.-03) und (2.1.4.-04) sind solche Fälle, bei den Beispielen (2.1.4.-05) und (2.1.4.-06) ist G aber nicht ohne Weiteres sinnvoll bestimmbar.

Bei manchen Meinungsumfragen besteht die Möglichkeit, zu den Fragen mehrere Antworten zu geben. Wird zum Beispiel gefragt „Was ist Ihnen wichtig?", so kann man aus einer vorgegebenen Liste mehrere Antworten auswählen. Die Anzahl der Antworten ist dann höher als die Anzahl der Befragten. Dagegen kann beim Würfeln bei jedem Wurf nur genau ein Ergebnis entstehen.

Im Folgenden wird vorausgesetzt, dass es bei jeder Untersuchung jeweils nur genau eine Merkmalsausprägung als Ergebnis gibt. Im Falle der Umfrage müsste bei jeder Frage genau ein Kreuz gesetzt werden – nicht mehr und nicht weniger.

(2.1.4-07) ***Bezeichnungen***
Gegeben sei eine Menge von n Objekten, bei denen ein bestimmtes Merkmal interessiert, es gibt also **n Beobachtungswerte** $\mathbf{x_i}$ ($i = 1, \ldots, n$). Ferner habe das Merkmal **m Ausprägungen** $\mathbf{a_j}$ ($j = 1, \ldots, m$).
Dann ist

$h_j = h($„die Ausprägung des Merkmals ist a_j") die absolute Häufigkeit der Ausprägung a_j

$f_j = \frac{h_j}{n}$ die relative Häufigkeit der Ausprägung a_j

* * *

Offenbar gilt:

$$\sum_{j=1}^{m} h_j = n$$

$$\sum_{j=1}^{m} f_j = \sum_{j=1}^{m} \frac{h_j}{n} = \frac{1}{n}\sum_{j=1}^{m} h_j = \frac{1}{n} \cdot n = 1$$

(2.1.4-08) ***Definition***

Gegeben sei eine endliche Menge von Objekten, bei denen ein bestimmtes Merkmal interessiert. Ordnet man allen Ausprägungen eines Merkmals jeweils die absolute Häufigkeit zu, so erhält man die **Häufigkeitsverteilung** des Merkmals. Eine solche Zusammenfassung heißt **Gruppierung**.
* * *

Bei dieser Zusammenfassung bleiben die einzelnen Beobachtungswerte erhalten, der Informationsverlust besteht höchstens darin, dass die Reihenfolge der Beobachtungen nicht mehr erkennbar ist.

(2.1.4-09) ***Beispiel***
In einem Unternehmen sind zu einem bestimmten Zeitpunkt 618 Personen beschäftigt. Die absolute und relative Häufigkeitsverteilung bezüglich des Geschlechtes ist:

Geschlecht	Absolute Häufigkeit	Relative Häufigkeit
m	273	$\frac{273}{618} = 44{,}175\%$
w	345	$\frac{345}{618} = 55{,}825\%$

Da es auf irgendwelche Reihenfolgen – z. B. nach Personalnummer – nicht ankommt, entsteht durch die Gruppierung kein praktisch bedeutsamer Informationsverlust.
* * *

(2.1.4-10) ***Beispiel***
Mit einem Würfel wird 100-mal gewürfelt, wobei wieder nur die gültigen Ergebnisse berücksichtigt werden. Die absolute Häufigkeitsverteilung ist in folgender Tabelle dargestellt.

Augenzahl	Anzahl = absolute Häufigkeit
1	15
2	18
3	12
4	16
5	18
6	21

Die Reihenfolge der Würfelergebnisse geht durch die Gruppierung verloren. Insbesondere kann man nach der Gruppierung nicht mehr eine Tendenz, wie sie z. B. durch Abnutzung des Würfels entstehen könnte, erkennen.
* * *

Zwei Beobachtungswerte können gleich sein, während sich die einzelnen Ausprägungen alle voneinander unterscheiden. Zu jedem x_i gibt es also genau ein a_j, aber zu einem a_j kann es kein, ein oder mehrere x_i geben.

In manchen Veröffentlichungen wird x_i sowohl für die Ausprägungen, als auch für die Beobachtungswerte benutzt (siehe z. B. [WM]). Das kann zu Missverständnissen führen, denn für die beiden Beobachtungswerte x_1 und x_2 kann $x_1 = x_2$ sein, während für die beiden Ausprägungen x_1 und x_2 sicher x_1

$\neq x_2$ ist. Solche Missverständnisse werden durch die unterschiedlichen Bezeichnungen a_j und x_i vermieden.

Da alle a_j unterschiedlich sind, gibt es die Menge der Ausprägungen: $A = \{a_j \mid j = 1, \ldots, m\}$.
Da die x_i nicht alle unterschiedlich sein müssen, gibt es die Menge der Beobachtungswerte im Allgemeinen nicht, denn Elemente einer Menge müssen wohlunterschieden sein. Als Unterscheidungsmerkmal kann dann zusätzlich der Index hinzugezogen werden.

(2.1.4-11) ***Beispiel***
Es werden folgende Augenzahlen gewürfelt: 1, 5, 4, 3, 3, 3, 5, 4, 5, 6. Dann ist die zugehörige …

- Menge der Ausprägungen:
  $A = \{1, 2, 3, 4, 5, 6\}$
- Menge der Beobachtungswerte mit der zugehörigen Wurfnummer:
  $B_1 = \{(i, x_i) \mid \text{Wurf } i = 1, \ldots, 10; \text{ zugehöriges Wurfergebnis } x_i\} = \{(1, 1), (2, 5), \ldots, (9, 5), (10, 6)\}$
- Menge der unterschiedlichen Beobachtungswerte:
  $B_2 = \{1, 3, 4, 5, 6\} \subset A$

* * *

(2.1.4-12) ***Definition***
Ist das Merkmal komparativ oder quantitativ, so können die Beobachtungswerte aufsteigend als Variationsreihe sortiert werden:

$$x_1 \leq x_2 \ldots \leq x_n$$

Setzt man diese Sortierung voraus, so gilt (i = 1, …, m):
Die **absolute Summenhäufigkeit**

$$H_i = \sum_{j=1}^{i} h_j$$

gibt an, mit welcher absoluten Häufigkeit Beobachtungswerte $\leq x_i$ vorkommen.
Die **relative Summenhäufigkeit**

$$F_i = \sum_{j=1}^{i} f_j$$

gibt an, mit welcher relativen Häufigkeit Beobachtungswerte $\leq x_i$ vorkommen.
Die Tabelle

(Ausprägung, absolute bzw. relative Summenhäufigkeit)

ist die **absolute** bzw. **relative Summenhäufigkeitsverteilung**.

* * *

(2.1.4-13) ***Beispiel***
Es werden folgende Zahlen gewürfelt: 1, 5, 4, 3, 3, 3, 5, 4, 5, 6.

- **Merkmalsträger** = die Würfe, identifiziert durch die Wurfnummer.
- n = 10 (Anzahl Würfe)
- **Beobachtungswerte** = die gewürfelten Augenzahlen x_i (i = 1, …, n=10)
- m = 6 (Anzahl Ausprägungen = mögliche Augenzahlen)
- **Ausprägungen**: $a_j = j$ (j = 1, …, m=6)
- **Urliste**:

Wurfnr. i	Gewürfelte Zahl
1	$x_1 = 1 = a_1$
2	$x_2 = 5 = a_5$
3	$x_3 = 4 = a_4$
4	$x_4 = 3 = a_3$
5	$x_5 = 3 = a_3$
6	$x_6 = 3 = a_3$
7	$x_7 = 5 = a_5$
8	$x_8 = 4 = a_4$
9	$x_9 = 5 = a_5$
10	$x_{10} = 6 = a_6$

Für die **Summenhäufigkeiten** werden die Beobachtungswerte aufsteigend sortiert und eine **Gruppierung** vorgenommen; die Information über die Reihenfolge der Würfe geht dadurch verloren. Dann ist:

j	**Ausprägung a_j**	**Absolute Häufigkeit h_j**	**Relative Häufigkeit f_j**	**Absolute Summenhäufigkeit H_j**	**Relative Summenhäufigkeit F_j**
1	1	1	$\frac{1}{10} = 0{,}1$	1	0,1
2	2	0	$\frac{0}{10} = 0{,}0$	1+0=1	0,1+0=0,1
3	3	3	$\frac{3}{10} = 0{,}3$	1+0+3=4	0,1+0+0,3=0,4
4	4	2	$\frac{2}{10} = 0{,}2$	1+0+3+2=6	0,1+0+0,3+0,2=0,6
5	5	3	$\frac{3}{10} = 0{,}3$	1+0+3+2+3=9	0,1+0+0,3+0,2+0,3=0,9
6	6	1	$\frac{1}{10} = 0{,}1$	1+0+3+2+3+1=10	0,1+0+0,3+0,2+0,3+0,1=1
		Summe = 10	Summe = 1		

Die Tabelle mit den Spalten „Ausprägung a_j“ und „Absolute Häufigkeit h_j“ bilden die Häufigkeitsverteilung des Merkmals „Augenzahl“.

Die absolute Summenhäufigkeit $H_3 = 4$ bedeutet, dass es 4 Beobachtungswerte $\leq a_3 = 3$ gibt.
Die relative Summenhäufigkeit $F_4 = 0{,}6$ bedeutet, dass 60% der Beobachtungswerte $\leq a_4 = 4$ sind.
* * *

Um die Häufigkeit von kombinierten Aussagen bestimmen zu können, gibt es eine Reihe von Rechenregeln.

(2.1.4-14) ***Satz***
A und B seien zwei Aussagen. Dann gelten für die absoluten Häufigkeiten folgende Rechenregeln:

1. $h(A) = h(A \text{ und } B) + h(A \text{ und } \overline{B})$
2. $h(A \text{ oder } B) = h(A \text{ und } B) + h(A \text{ und } \overline{B}) + h(\overline{A} \text{ und } B)$
3. $h(A \text{ oder } B) = h(A) + h(B) - h(A \text{ und } B)$
4. $h(\overline{A}) = n - h(A)$

Wurden bei insgesamt n Fällen betrachtet, ob die Aussage A zutrifft, so sind die relativen Häufigkeiten sinnvoll. Durch Division durch n erhält man dann die analogen Gleichungen für die relativen Häufigkeiten:

5. f(A) = f(A und B) + f(A und $\overline{B}$)

6. f(A oder B) = f(A und B) + f(A und $\overline{B}$) + f($\overline{A}$ und B)

7. f(A oder B) = f(A) + f(B) – f(A und B)

8. f($\overline{A}$) = 1 – f(A)

<u>Beweis:</u>
Die 1. und 2. Gleichung werden anhand des folgenden Beispiels plausibel hergeleitet.
Die 3. Gleichung kann aus den beiden vorangegangenen Gleichungen abgeleitet werden:

$$\begin{aligned} h(A \text{ oder } B) &= h(A \text{ und } \overline{B}) + h(\overline{A} \text{ und } B) + h(A \text{ und } B) \\ &= (h(A \text{ und } \overline{B}) + h(A \text{ und } B)) + (h(\overline{A} \text{ und } B) + h(A \text{ und } B)) - h(A \text{ und } B) \\ &= h(A) + h(B) - h(A \text{ und } B) \end{aligned}$$

Die 4. Gleichung kann aus der 3. Gleichung abgeleitet werden, indem man B = $\overline{A}$ setzt:

$$h(A \text{ oder } \overline{A}) = h(A) + h(\overline{A}) - h(A \text{ und } \overline{A}) \iff 1 = h(A) + h(\overline{A}) - 0 \iff h(\overline{A}) = 1 - h(A)$$

* * *

(2.1.4-15) ***Beispiel***
Anhand eines Beispiels werden die Rechenregeln für absolute und relative Häufigkeiten plausibel hergeleitet.
100 Personen einer Gruppe werden befragt, ob sie ein Auto oder ein Fahrrad besitzen. Es ergibt sich folgendes Bild:

Auto	**Fahrrad**	**Anzahl**
Ja	Ja	30
Nein	Ja	20
Ja	Nein	40
Nein	Nein	10
SUMME		100

A und F bzw. $\overline{A}$ und $\overline{F}$ seien die folgenden Aussagen:

A = „Person besitzt ein Auto“; $\overline{A}$ = „Person besitzt kein Auto“
F = „Person besitzt ein Fahrrad“; $\overline{F}$ = „Person besitzt kein Fahrrad“

Für die absoluten Häufigkeiten gilt dann:

1. h(A) = h(A und F) + h(A und $\overline{F}$) = 30 + 40 = 70
   im Klartext:
   Anzahl Personen, die ein Auto haben = Anzahl Personen, die ein Auto und ein Fahrrad haben plus Anzahl Personen, die ein Auto und kein Fahrrad haben

   h(F) = h(A und F) + h($\overline{A}$ und F) = 30 + 20 = 50

im Klartext:
Anzahl Personen, die ein Fahrrad haben = Anzahl Personen, die ein Auto und ein Fahrrad haben
plus Anzahl Personen, die kein Auto und ein Fahrrad haben

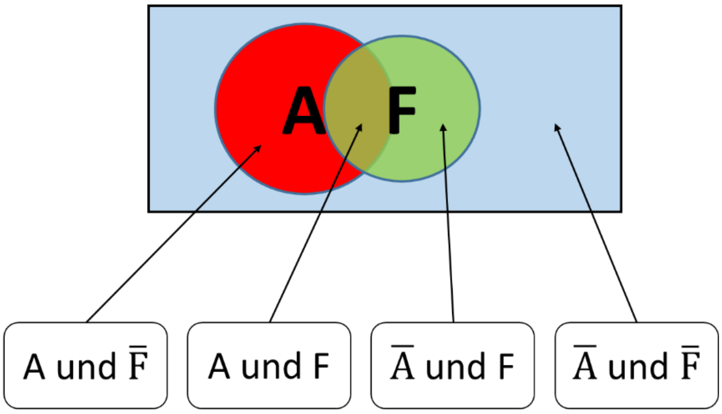
2. $h(A \text{ oder } F) = h(A \text{ und } F) + h(A \text{ und } \overline{F}) + h(\overline{A} \text{ und } F) = 30 + 40 + 20 = 90$

3. $h(A \text{ oder } F) = h(A) + h(F) - h(A \text{ und } F) = 70 + 50 - 30 = 90$

4. $h(\overline{A}) = n - h(A) = 100 - 70 = 30$
   $h(\overline{F}) = n - h(F) = 100 - 50 = 50$

* * *

Man kann sich die Regeln gut durch Mengen-Diagramme verdeutlichen:



A und $\overline{F}$	=	A ohne F	=	die linke rote Sichel
A und F	=	sowohl A, als auch F	=	Schnittmenge von A und F in der Mitte
$\overline{A}$ und F	=	F ohne A	=	die grüne rechte Sichel
$\overline{A}$ und $\overline{F}$	=	weder A noch F	=	die blaue Fläche ohne die rote und die grüne Fläche

Die absolute Häufigkeit wird dann durch die Fläche, die relative Häufigkeit durch den relativen Flächenanteil repräsentiert.
Zum Beispiel ist die Formel

$$h(A \text{ oder } F) = h(A) + h(F) - h(A \text{ und } F)$$

dann zu interpretieren als:

Gesamtfläche der beiden Kreisscheiben = Fläche der roten Kreisscheibe
+ Fläche der grünen Kreisscheibe
– Fläche der Überschneidung der Kreisscheiben

(2.1.4-16) ***Die bedingte relative Häufigkeit***
Gegeben ist eine endliche Menge G von Objekten. A und B seien Teilmengen von G, die absolute Häufigkeit h sei wieder die Anzahl der Elemente, die relative Häufigkeit f sei der relative Anteil bezüglich der Gesamtheit G.
Soll f nur auf die Teilmenge B statt auf die gesamte Menge G bezogen werden, so schreibt man für die relative Häufigkeit: f_B. Es ist also insbesondere $f_G = f$.

$f_B(A)$ ist dann die relative Häufigkeit der Elemente von A, die gleichzeitig in B liegen, bezogen auf B. Man nimmt also von allem nur den Ausschnitt, der in B liegt, alles andere wird ausgeblendet. f_B heißt die **bedingte relative Häufigkeit**, $f_B(A)$ ist die relative Häufigkeit von A unter der Bedingung B. Statt $f_B(A)$ schreibt man häufig auch f(A|B).
Es gilt:

$$f(A|B) = f_B(A) = \frac{h(A \text{ und } B)}{h(B)} = \frac{\frac{h(A \text{ und } B)}{h(G)}}{\frac{h(B)}{h(G)}} = \frac{f(A \text{ und } B)}{f(B)}$$

Die Formel wird durch Beispiel (2.1.4-15) plausibel hergeleitet:
$f_A(F)$ ist die relative Häufigkeit der Fahrradbesitzer, wenn als Basis nur die Autobesitzer genommen werden, also:

$$f(F|A) = f_A(F) = \frac{h(F \text{ und } A)}{h(A)} = \frac{30}{70} = \frac{3}{7}$$

$f_F(A)$ ist die relative Häufigkeit der Autobesitzer, wenn als Basis nur die Fahrradbesitzer genommen werden, also:

$$f(A|F) = f_F(A) = \frac{h(A \text{ und } F)}{h(F)} = \frac{30}{50} = \frac{3}{5}$$

Sind in der Ausgangstabelle nur relative statt absolute Häufigkeiten genannt, so ergibt sich:

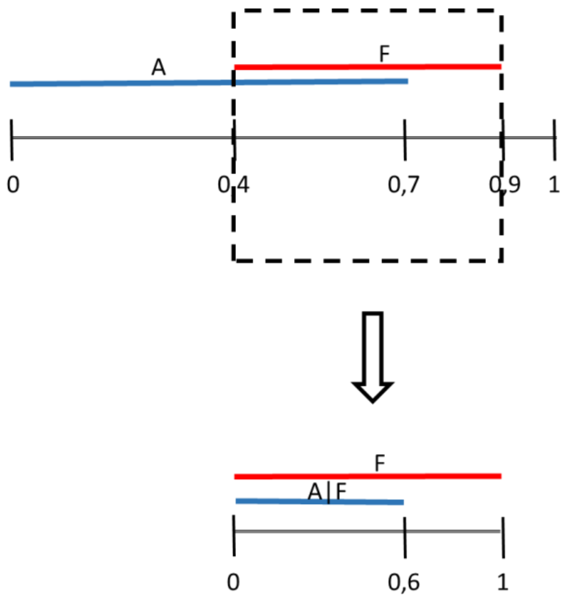
Auto	**Fahrrad**	**Relative Häufigkeit**
Ja	Ja	0,3
Nein	Ja	0,2
Ja	Nein	0,4
Nein	Nein	0,1
SUMME		1,0

und damit wie oben

$$f(A|F) = \frac{f(A \text{ und } F)}{f(F)} = \frac{0{,}3}{0{,}5} = \frac{3}{5}$$

$$f(F|A) = \frac{f(F \text{ und } A)}{f(A)} = \frac{0{,}3}{0{,}7} = \frac{3}{7}$$

Grafisch kann das folgendermaßen veranschaulicht werden:



Der obere Teil der Grafik zeigt die Ausgangssituation mit den auf die Gesamtheit G bezogenen relativen Häufigkeiten, während der untere Teil die Situation mit den auf F bezogenen relativen Häufigkeiten zeigt.

(2.1.4-17) ***Relative Anteile***
Die Gleichungen für die relativen Häufigkeiten gelten ganz analog auch für andere relative Anteile, wie relativer Flächen-, Zeit-, Gewichts- oder Geld-Anteil. Das kann man sich verdeutlichen, indem man die einzelnen Größen in gleichgroße Teile zerlegt, gewissermaßen die „Atome“ der Größen, und dann abzählt. Auf diese Weise hat man die relativen Größen auf die relativen Häufigkeiten zurückgeführt.
Ist z. B. die Messgenauigkeit bei einer Fläche 1 mm², so zerlegt man die Gesamtfläche in 1 mm² große Teilflächen. Dabei muss die Zerlegung so gewählt werden, dass bei jeder Teilfläche eindeutig feststellbar ist, ob sie die betrachtete Eigenschaft hat oder nicht.
Anschließend zählt dann ab, wie viele dieser Teilflächen die untersuchte Eigenschaft haben, und bildet daraus die relative Häufigkeit. Diese relative Häufigkeit hat denselben Wert wie der relative Flächenanteil.

Die Gleichungen für relative Häufigkeiten gelten auch analog für Wahrscheinlichkeiten (siehe Kapitel 5.11).

2.2 Lagemaße

Hat man eine große Menge von Daten, so dienen Lagemaße zum einen dazu, einen ersten Eindruck zu haben, wo die Daten liegen. Zum anderen helfen sie aber auch dabei, mehrere große Datenmengen grob vergleichbar zu machen, indem man die jeweiligen Lagemaße miteinander vergleicht.

Wichtig ist aber auch, dass die Verdichtung einer großen Datenmenge auf die wenigen Lagemaße mit einem erheblichen Informationsverlust verbunden ist – also Vorsicht bei der Interpretation der Ergebnisse! Hat man z. B. die Einkünfte aus Kapitalvermögen von ca. 2,6 Millionen Bürgern (siehe [SB]), so ist das arithmetische Mittel dieser Werte eine Verdichtung von 2,6 Millionen Daten auf eine einzige Zahl und damit gehen fast alle Informationen verloren.

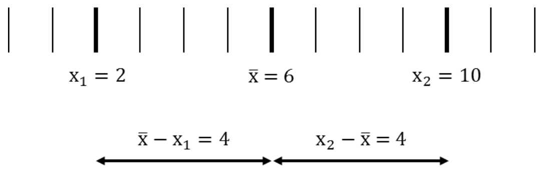
Statt Lagemaße wird auch manchmal der Begriff Lageparameter verwendet.

2.2.1 Arithmetisches Mittel

Hinweis: Wird nur das Wort „Mittel“ oder „Durchschnitt“ ohne nähere Spezifikation benutzt, so ist stets das **arithmetische** Mittel gemeint.

Zunächst bedeutet das arithmetische Mittel den Wert, der in der Mitte ist, also:

Hat man zwei Werte, so hat die Mitte die Eigenschaft, dass der Abstand nach links genauso groß ist wie der Abstand nach rechts, z. B.:



Ist allgemein $x_1 \leq x_2$ und $\bar{x}$ die Mitte dieser beiden Zahlen, so ist:

$$\bar{x} - x_1 = x_2 - \bar{x} \quad \Leftrightarrow \quad \bar{x} - x_1 = -(\bar{x} - x_2) \quad \Leftrightarrow \quad (\bar{x} - x_1) + (\bar{x} - x_2) = 0$$

Hat man mehr als zwei Werte, so ist als Verallgemeinerung naheliegend, dass die Mitte die Eigenschaft hat, dass die Summe der Abstände nach links genauso groß ist, wie die Summe der Abstände nach rechts. Die Mitte liegt zwischen zwei benachbarten Werten, d. h., es gibt einen Index j mit der Eigenschaft

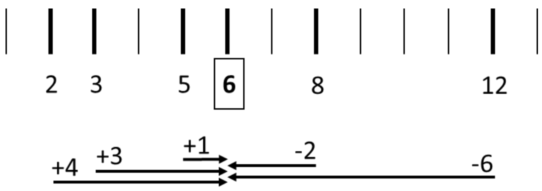
$$x_j \leq \bar{x} \leq x_{j+1}$$

Damit ergibt sich:

$$\begin{aligned} & \text{Summe der Abstände nach links} = \text{Summe der Abstände nach rechts} \\ \Leftrightarrow \quad & (\bar{x} - x_1) + \ldots + (\bar{x} - x_j) = (x_{j+1} - \bar{x}) + \ldots + (x_n - \bar{x}) \\ \Leftrightarrow \quad & (\bar{x} - x_1) + \ldots + (\bar{x} - x_j) + (\bar{x} - x_{j+1}) + \ldots + (\bar{x} - x_n) = 0 \\ \Leftrightarrow \quad & (\bar{x} - x_1) + \ldots + (\bar{x} - x_n) = 0 \end{aligned}$$

(2.2.1-01) ***Beispiel***

Das arithmetische Mittel der Zahlen 2, 3, 5, 8, 12 ist die Zahl 6, denn die Summe der Abstände nach links ist +8 und die Summe der Abstände nach rechts ist -8, insgesamt ist also die Summe der Abstände = 0:



$(\mathbf{6} - 2) + (\mathbf{6} - 3) + (\mathbf{6} - 5) + (\mathbf{6} - 8) + (\mathbf{6} - 12) = 0$

* * *

Löst man die allgemeine Gleichung nach $\bar{x}$ auf, so ergibt sich:

$$
\begin{aligned}
& (\bar{x} - x_1) + (\bar{x} - x_2) + \ldots + (\bar{x} - x_n) = 0 \\
\Leftrightarrow\quad & (\bar{x} + \bar{x} + \ldots + \bar{x}) - (x_1 + x_2 + \ldots + x_n) = 0 \\
\Leftrightarrow\quad & n \cdot \bar{x} - \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad n \cdot \bar{x} = \sum_{i=1}^{n} x_i \quad \Leftrightarrow \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\end{aligned}
$$

Folgende Definition ist also sinnvoll:

(2.2.1-02) ***Definition***
Gegeben sei ein quantitatives Merkmal mit den Beobachtungswerten $x_1, \ldots, x_n$. Dann ist das **arithmetische Mittel** der Beobachtungswerte definiert als:

$$\bar{\mathbf{x}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i}$$

$\bar{x}$ hat dieselbe Maßeinheit wie die Beobachtungswerte.
* * *

Man kann das arithmetische Mittel auch folgendermaßen interpretieren:
Statt die n Werte $x_1, \ldots, x_n$ zu addieren, kann man einfacher n-mal das arithmetische Mittel $\bar{x}$ addieren, also $\bar{x}$ mit n multiplizieren. Diese Interpretation ist in der Praxis aber nur dann sinnvoll, wenn die Summe der Beobachtungswerte eine praktisch sinnvolle Größe ist.

(2.2.1-03) ***Beispiele***

1. Haben 100 Personen im Durchschnitt 73,49€ gespendet, so ist die Gesamtspendensumme 73,49€·100 = 7.349€. Die einzelnen tatsächlichen Beträge sind unerheblich, entscheidend ist nur die Summe.

2. Eine Durchschnittsschulnote errechnet sich zwar gemäß der Formel für das arithmetische Mittel, aber die Summe der Noten (ohne die Division durch n) ist keine sinnvolle Größe.

3. Hat man einen annähernd masselosen Stab und verteilt darauf n Gewichte gleicher Masse, so ist der Schwerpunkt dieser Figur das arithmetische Mittel der Abstände der Massen vom Rand des Stabes. In diesem Fall nennt man das arithmetische Mittel den Massenmittelpunkt.

4. In einem Dreieck gilt: Bildet man das arithmetische Mittel der Koordinaten der Eckpunkte, so erhält man die Koordinaten des geometrischen Schwerpunktes.

* * *

Ist h_j die absolute Häufigkeit und f_j die relative Häufigkeit der Ausprägungen a_j, so gilt für das arithmetische Mittel der Beobachtungswerte:

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{m} h_j \cdot a_j = \sum_{j=1}^{m} f_j \cdot a_j$$

(2.2.1-04) ***Beispiel***
Es wurde 10-mal mit folgendem Ergebnis gewürfelt:

i	Augenzahl a_j	Abs. Häufigkeit h_j	Rel. Häufigkeit f_j
1	1	1	$\frac{1}{10}$
2	2	0	0
3	3	3	$\frac{3}{10}$
4	4	2	$\frac{2}{10}$
5	5	3	$\frac{3}{10}$
6	6	1	$\frac{1}{10}$

Dann ist die Anzahl der Beobachtungswerte n = 10 und die Anzahl der Ausprägungen m = 6.

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}\cdot(1+3+3+3+4+4+5+5+5+6) = \frac{39}{10} = 3{,}9$$

$$\bar{x} = \frac{1}{10}\sum_{j=1}^{6} h_j \cdot a_j = \frac{1}{10}\cdot(1\cdot 1+0\cdot 2+3\cdot 3+2\cdot 4+3\cdot 5+1\cdot 6) = \frac{39}{10} = 3{,}9$$

$$\bar{x} = \sum_{j=1}^{6} f_j \cdot a_j = \frac{1}{10}\cdot 1+0\cdot 2+\frac{3}{10}\cdot 3+\frac{2}{10}\cdot 4+\frac{3}{10}\cdot 5+\frac{1}{10}\cdot 6 = \frac{39}{10} = 3{,}9$$

* * *

(2.2.1-05) ***Beispiel***
Das arithmetische Mittel ist auch juristisch wichtig: gemäß Fertigpackungsverordnung (siehe [FE]) darf der Mittelwert einer Charge die angegebene Füllmenge nicht unterschreiten, einzelne Packungen dürfen dagegen eine geringere Füllmenge beinhalten.

* * *

2.2.2 Geometrisches Mittel

Analog zur Addition beim arithmetischen Mittel kann man auch bei der Multiplikation ein Mittel definieren.

Anstelle des Produktes von n einzelnen Werten q_i (i = 1, …, n) kann auch n-mal derselbe Wert $\bar{q}_G$ als Faktor genommen werden:

$$q_1 \cdot q_2 \cdot \ldots \cdot q_n = \bar{q}_G \cdot \bar{q}_G \cdot \ldots \cdot \bar{q}_G = \bar{q}_G^{\,n}$$

(2.2.2-01) ***Definition***

Für Zahlen $q_i > 0$ (i = 1, …, n) heißt

$$\bar{q}_G = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots \cdot q_n} = \sqrt[n]{\prod_{i=1}^{n} q_i}$$

das **geometrische Mittel**. $\bar{q}_G$ hat dieselbe Dimension = Maßeinheit wie die q_i-Werte.

Stellt man die Gleichung um, so ergibt sich:

$$\frac{\bar{q}_G}{q_1} \cdot \frac{\bar{q}_G}{q_2} \cdot \ldots \cdot \frac{\bar{q}_G}{q_n} = 1 \qquad \Leftrightarrow \qquad \frac{q_1}{\bar{q}_G} \cdot \frac{q_2}{\bar{q}_G} \cdot \ldots \cdot \frac{q_n}{\bar{q}_G} = 1$$

(2.2.2-02) ***Anwendung***

Das geometrische Mittel wird insbesondere bei Wachstumsprozessen benötigt.
Gegeben ist ein Merkmalsträger mit einem quantitativen Merkmal, das sich im Laufe der Zeit ändert.
Das Merkmal wird in festen Zeitabständen gemessen.
Es sei $x_i > 0$ der Messwert nach i festen Zeitabständen (i = 0, 1, …, n), x_0 ist der erste Messwert und x_n ist der letzte Messwert.
$q_i = \frac{x_i}{x_{i-1}}$ ist der Wachstumsfaktor des Merkmals in der i-ten Zeitperiode gegenüber der Vorperiode.
Der durchschnittliche Wachstumsfaktor pro Zeiteinheit wird dann durch das geometrische Mittel

$$\bar{q}_G = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots \cdot q_n} = \sqrt[n]{\prod_{i=1}^{n} q_i}$$

oder gleichwertig

$$\bar{q}_G = \sqrt[n]{\frac{x_n}{x_0}}$$

beschrieben.
Die zweite Gleichung ergibt sich aus der ersten, da sich die Werte x_1 bis x_{n-1} herauskürzen:

$$\bar{q}_G = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots \cdot q_n} = \sqrt[n]{\frac{x_1}{\mathbf{x_0}} \cdot \frac{x_2}{x_1} \cdot \ldots \cdot \frac{x_{n-1}}{x_{n-2}} \cdot \frac{\mathbf{x_n}}{x_{n-1}}} = \sqrt[n]{\frac{x_n}{x_0}}$$

In diesem Fall sind die q_i-Werte und $\bar{q}_G$ dimensionslos, da sich die identische Dimension der x_i-Werte herauskürzt.

(2.2.2.-03) ***<u>Beispiel</u>***

Es wird Geld zu folgenden Zinssätzen angelegt:

1. Jahr: 1 % ; 2. Jahr: 1 % ; 3. Jahr: 1,5 % ; 4. Jahr: 1,8 % ; 5.Jahr: 2 %

Die Zinsen werden am jeweiligen Jahresende dem Konto gutgeschrieben, werden also weiter verzinst (Zinseszins).
Wie hoch ist der durchschnittliche Zinssatz?

Es ist n = 5, die jährlichen Veränderungsfaktoren des Kapitals („Aufzinsfaktor") sind q_i = 1 + Zinssatz, also:

$q_1 = 1{,}01;\ q_2 = 1{,}01;\ q_3 = 1{,}015;\ q_4 = 1{,}018;\ q_5 = 1{,}02$

$$q_G = \sqrt[5]{q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot q_5} = \sqrt[5]{1{,}01 \cdot 1{,}01 \cdot 1{,}015 \cdot 1{,}018 \cdot 1{,}02} \approx \sqrt[5]{1{,}07512} \approx 1{,}0146$$

Der durchschnittliche Steigerungsfaktor pro Jahr ist 1,0146, der durchschnittliche Zinssatz also 1,0146 – 1 = 0,0146 = 1,46 %. Man erhält also dasselbe Endkapital, wenn man das Ausgangskapital mit einheitlich jährlich 1,46 % anstelle der 5 vorgegebenen Zinssätze verzinst.

(2.2.2.-04) ***Beispiel***

Zu Beginn der Betrachtung beträgt der Aktienkurs 50 € pro Aktie, nach 6 Jahren beträgt der Aktienkurs 70 € pro Aktie. Wie hoch war die durchschnittliche Wertsteigerung pro Jahr?
n = 6, x_i ist der Aktienkurs nach i Jahren (i = 0, 1, …, 6). Dann ist

$$\bar{q}_G = \sqrt[6]{\frac{x_6}{x_0}} = \sqrt[6]{\frac{70}{50}} \approx 1{,}0577$$

Im Durchschnitt ist der Aktienkurs jedes Jahr 1,0577-mal so hoch wie im Vorjahr, der Kurs stieg also um durchschnittlich 5,77 % pro Jahr.

2.2.3 Gewichtetes arithmetisches Mittel

Manchmal ist es sinnvoll, einzelne Beobachtungswerte zu gewichten, um den Werten einen unterschiedlichen Einfluss auf den Mittelwert zu geben. Das wird z. B. bei der Durchschnittsnote des Abiturs so gemacht, um Leistungskurse stärker in die Durchschnittsnote einfließen zu lassen.

Statt n Beobachtungswerte x_i (i = 1, …, n) mit n Größen $c_i > 0$ (i = 1, …, n) zu gewichten und dann zu addieren kann man auch jedes Mal den Wert $\bar{x}_H$ gewichten und dann addieren:

$$c_1 \cdot x_1 + c_2 \cdot x_2 + \ldots + c_n \cdot x_n = c_1 \cdot \bar{x}_H + c_2 \cdot \bar{x}_H + \ldots + c_n \cdot \bar{x}_H$$

Daraus folgt:

$$c_1 \cdot x_1 + c_2 \cdot x_2 + \ldots + c_n \cdot x_n = (c_1 + c_2 + \ldots + c_n) \cdot \bar{x}_H \quad \Leftrightarrow$$

$$\bar{x}_H = \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i}$$

(2.2.3.-01) ***Definition***

Gegeben sind die Beobachtungswerte x_i (i = 1, …, n) und die Gewichte $c_i > 0$ (i = 1, …, n).
Dann heißt

$$\bar{x}_H = \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i}$$

das **gewichtete arithmetische Mittel** der Beobachtungswerte x_i mit den Gewichten $c_i > 0$.

$\bar{x}_H$ hat dieselbe Dimension = Maßeinheit wie die x_i-Werte.

Analog zum arithmetischen und zum geometrischen Mittel kann man die Gleichung umstellen:

$$c_1 \cdot (\bar{x}_H - x_1) + c_2 \cdot (\bar{x}_H - x_2) + \ldots + c_n \cdot (\bar{x}_H - x_n) = 0$$

$$\frac{c_1 \cdot \bar{x}_H + c_2 \cdot \bar{x}_H + \cdots + c_n \cdot \bar{x}_H}{c_1 \cdot x_1 + c_2 \cdot x_2 + \cdots + c_n \cdot x_n} = 1$$

(2.2.3.-02) ***Anwendung***
Das gewichtete arithmetische Mittel wird auch beim Durchschnitt von Verhältnisgrößen benötigt.
Es werden Objekte untersucht, die zwei quantitative Merkmale mit den Beobachtungswerten b_i und c_i ($c_i > 0$) haben, der Quotient sei

$$x_i = \frac{b_i}{c_i} \quad (i = 1, ..., n)$$

Dann ist

$$\bar{x}_H = \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i} = \frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} c_i} = \frac{\text{Summe der Zähler}}{\text{Summe der Nenner}}$$

kurz: Durchschnitts-x = Durchschnitts-b pro c $= \frac{\text{Gesamt-b}}{\text{Gesamt-c}}$

Hat man nur die b_i und x_i (bzw. nur die c_i und x_i), so errechnet man erst die fehlenden c_i (bzw. b_i) und ermittelt dann das gewichtete arithmetische Mittel.

(2.2.3.-03) ***Hinweis***
Es ist wichtig, dass x_i aus b_i und c_i ermittelt wird und nicht umgekehrt:
Es sei $x_1 = \frac{1}{2}$ und $x_2 = \frac{1}{3}$. b_i und c_i sind daraus nicht eindeutig ableitbar, so dass sich verschiedene Werte für $\bar{x}_H$ daraus ableiten lassen.
Möglichkeit 1: $x_1 = \frac{1}{2}$ und $x_2 = \frac{1}{3}$; $b_1 = 1, c_1 = 2, b_2 = 1, c_2 = 3 \Rightarrow \bar{x}_H = \frac{1+1}{2+3} = \frac{2}{5} = 0{,}4$
Möglichkeit 2: $x_1 = \frac{1}{2}$ und $x_2 = \frac{2}{6}$; $b_1 = 1, c_1 = 2, b_2 = 2, c_2 = 6 \Rightarrow \bar{x}_H = \frac{1+2}{2+6} = \frac{3}{8} = 0{,}375$

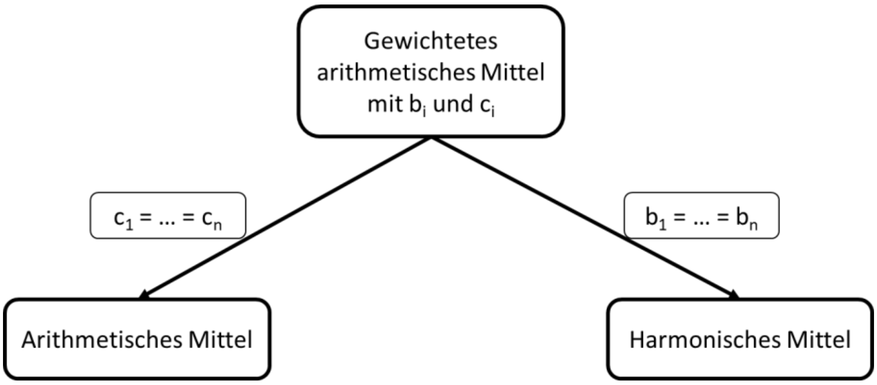
(2.2.3.-04) ***Sonderfall***
Sind alle b_i gleich, also $b_1 = \ldots = b_n = b \neq 0$, so ist $c_i = \frac{b}{x_i}$ und $\sum_{i=1}^{n} b_i = n \cdot b$ und man erhält:

$$\bar{x}_H = \frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} \frac{b_i}{x_i}} = \frac{n \cdot b}{\sum_{i=1}^{n} \frac{b}{x_i}} = \frac{b \cdot n}{b \cdot \sum_{i=1}^{n} \frac{1}{x_i}} = \frac{1}{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{x_i}}$$

Dieser Sonderfall wird als das **harmonische Mittel** der x_i bezeichnet. An den Beispielen unten erkennt man aber, dass das gewichtete arithmetische Mittel einfacher zu verstehen und zu handhaben ist als dieser Sonderfall.

(2.2.3.-05) ***Sonderfall***
Sind alle c_i gleich (d. h., jeder Messwert wird gleich gewichtet), also $c_1 = \ldots = c_n = c \neq 0$, so ist $b_i = c \cdot x_i$ und $\sum_{i=1}^{n} c_i = n \cdot c$ und man erhält das (ungewichtete) arithmetische Mittel $\bar{x}$ der Zahlen x_i (i = 1, ..., n):

$$\bar{x}_H = \frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} c_i} = \frac{\sum_{i=1}^{n} c \cdot x_i}{n \cdot c} = \frac{c \cdot \sum_{i=1}^{n} x_i}{c \cdot n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$



(2.2.3.-06) ***Sonderfall***
Sind die c_i prozentuale Anteile, so ist

$$\sum_{i=1}^{n} c_i = 100\,\% = 1$$

d. h., der Nenner kann in diesem Fall weggelassen werden:

$$\bar{x}_H = \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i} = \sum_{i=1}^{n} c_i \cdot x_i$$

(2.2.3.-07) ***Sonderfall***
Bei manchen Fragestellungen ist es wichtig, die Entwicklung des Mittels zu analysieren, wenn man immer mehr Beobachtungswerte berücksichtigt. Das „n“ in der Formel für $\bar{x}_H$ ist dann variabel.

n	1	2	3	4	…
$\bar{x}_H$	$\frac{b_1}{c_1}$	$\frac{b_1+b_2}{c_1+c_2}$	$\frac{b_1+b_2+b_3}{c_1+c_2+c_3}$	$\frac{b_1+b_2+b_3+b_4}{c_1+c_2+c_3+c_4}$	…

Da die $c_i > 0$ sind, wächst der Nenner mit wachsendem n, es können also nicht zwei gleiche Nenner auftreten. Zu jedem dieser Nenner gibt es genau einen Wert im Zähler, so dass der Nenner durch eine Variable

$r_n = c_1 + \ldots + c_n$

und der Zähler als eine Funktion dieser Variablen r_n, also

$$f(r_n) = b_1 + \ldots + b_n$$
beschrieben werden kann, also:
$$\overline{x}_H(r_n) = \frac{f(r_n)}{r_n}$$
kurz:
$$\overline{x}_H(r) = \frac{f(r)}{r}$$

In diesem Fall ist $\frac{f(r)}{r}$ das Durchschnitts-x oder auch „Durchschnitts-f(r) pro r“.
Diese Darstellung wird häufig bei **ökonomischen Funktionen** benutzt, wie z. B. bei den Durchschnittskosten: r ist die produzierte Menge (gemessen in Anzahl oder Gewicht oder Volumen oder …), f(r) sind die zugehörigen Gesamtkosten. Dann sind $\overline{x}_H(r) = \frac{f(r)}{r}$ die Kosten pro Mengeneinheit = Durchschnittskosten.

(2.2.3.-08) ***Beispiel***
Ein Auto fährt eine Strecke gemäß der folgenden Tabelle:

Größe	Wert 1	Wert 2
Strecke s_i [km]	100	80
Zeit t_i [h]	2	1
Geschwindigkeit $v_i = \frac{s_i}{t_i}$ [km/h]	50	80

Wie hoch ist die Durchschnittsgeschwindigkeit für die Gesamtstrecke / Gesamtzeit?
Da die s_i und t_i bereits vorliegen, ergibt sich als Durchschnittsgeschwindigkeit:
$$\overline{v}_H = \frac{s_1+s_2}{t_1+t_2} = \frac{100+80}{2+1} = 60 \text{ [km/h]}$$
kurz:
Durchschnitts-Geschwindigkeit = Durchschnittsstrecke pro Stunde = $\frac{\text{Gesamt-Strecke}}{\text{Gesamt-Zeit}}$
Die Formel für das harmonische Mittel wird nicht benötigt.

Man darf nicht einfach das arithmetische Mittel der Geschwindigkeiten nehmen (das wäre 65 [km/h]), da die Zeitintervalle unterschiedlich lang sind. Normiert man aber die Zeiten (in der Zeile „Zeit“ muss immer derselbe Wert stehen, z. B. der ggT („größter gemeinsamer Teiler“) der Einzelzeiten), so kann einfach das arithmetische Mittel der Geschwindigkeiten genommen werden (siehe oben Sonderfall 2):

Größe	Wert 1	Wert 2	Wert 3
Strecke s_i [km]	50	50	80
Zeit t_i [h]	1	1	1
Geschwindigkeit v_i [km/h]	50	50	80

$$\overline{v} = \frac{50+50+80}{3} = 60 \text{ [km/h]}$$

Die Formel für das harmonische Mittel wird nicht benötigt.

Hinweis:
Das arithmetische Mittel von Geschwindigkeiten kann auch sinnvoll sein: wird eine Radarkontrolle zur Messung der Geschwindigkeiten von Fahrzeugen durchgeführt, so ist das arithmetische Mittel die

durchschnittliche gemessene Geschwindigkeit. Das gewichtete arithmetische Mittel ist in diesem Falle nicht sinnvoll.

(2.2.3.-09) ***Beispiel***
Man tankt dreimal:

Größe	Wert 1	Wert 2	Wert 3
Preis g_i [€]	40,50	49,00	33,00
Menge m_i in Litern [l]	30	35	25
Preis pro Liter $p_i = \frac{g_i}{m_i}$ [€/l]	1,35	1,40	1,32

Wie hoch ist der durchschnittliche Preis pro Liter?
Der durchschnittliche Preis pro Liter ist:

$$\bar{p}_H = \frac{g_1+g_2+g_3}{m_1+m_2+m_3} = \frac{40{,}5+49+33}{30+35+25} = 1{,}3611 \quad [€/l]$$

kurz: Durchschnitts-Preis pro Liter = $\frac{\text{Gesamt-Preis}}{\text{Gesamt-Menge}}$
Die Formel für das harmonische Mittel wird nicht benötigt.

(2.2.3.-10) ***Beispiel***
Kauft man Aktien eines Unternehmens zu unterschiedlichen Zeitpunkten, so beträgt der durchschnittliche Kurs einer Aktie = $\frac{\text{Gesamt-Kaufpreis}}{\text{Gesamt-Anzahl Aktien}}$.

(2.2.3.-11) ***Beispiel***
Ein Safthersteller hat 600 kg Äpfel von verschiedenen Erzeugern gesammelt und dabei je nach Menge und Qualität der Äpfel unterschiedliche Zahlungen geleistet, insgesamt aber 480 €. Wieviel hat der Safthersteller durchschnittlich pro kg bezahlt?

$$\bar{p}_H = \frac{\text{Gesamtzahlung}}{\text{Gesamtmenge}} = \frac{480€}{600kg} = 0{,}80 \text{ €/kg}$$

(2.2.3.-12) ***Beispiel***
Der Benzinverbrauch eines Autos kann als eine Funktion der gefahrenen Strecke ab einem bestimmten Startpunkt dargestellt werden: je mehr man fährt, umso mehr verbraucht man. Der Durchschnittsverbrauch pro 100 km beim Auto ist dann eine Funktion in Abhängigkeit von der gefahrenen Strecke, wie man in der Anzeige im Auto mitverfolgen kann.
Zunächst werden nur Teilstrecken betrachtet: erst wird die erste Teilstrecke gefahren, dann die zweite Teilstrecke, etc.

Größe	Wert 1	Wert 2	Wert 3
Teilstrecke [km]	10	15	15
Menge pro Teilstrecke in Litern [l]	0,6	1,0	0,8
Verbrauch pro Teilstrecke [l/km]	$\frac{0,6}{10} = 0{,}060$	$\frac{1,0}{15} = 0{,}067$	$\frac{0,8}{15} = 0{,}053$
Verbrauch pro Teilstrecke [l/100km]	6,0	6,7	5,3

Dann werden die Teilstrecken und Verbräuche ab Startpunkt kumuliert:

Größe	Wert 1	Wert 2	Wert 3
Gesamtstrecke s ab Startpunkt [km]	10	10+15 = 25	10+15+15 = 40
Gesamtverbrauch m ab Startpunkt [l]	0,6	0,6+1,0 = 1,6	0,6+1,0+0,8 = 2,4

Die Funktion f beschreibt die verbrauchte Menge m in Abhängigkeit von der gefahrenen Strecke s, also m = f(s). Dann ist

Durchschnittsverbrauch pro km = $\frac{\text{Gesamtverbrauch}}{\text{Gesamtstrecke}} = \frac{f(s)}{s}$, also:

Durchschnittsverbrauch nach 10 km = $\frac{f(10)}{10} = \frac{0,6}{10}$ = 0,060 [l/km] = 6,0 [l/100km]

Durchschnittsverbrauch nach 25 km = $\frac{f(25)}{25} = \frac{1,6}{25}$ = 0,064 [l/km] = 6,4 [l/100km]

Durchschnittsverbrauch nach 40 km = $\frac{f(40)}{40} = \frac{2,4}{40}$ = 0,060 [l/km] = 6,0 [l/100km]

Diese Berechnung wird von der Autoelektronik alle paar Sekunden neu durchgeführt und demzufolge kann sich die Anzeige alle paar Sekunden ändern.

2.2.4 Allgemeine Eigenschaften der Mittelwerte

Es liegt nahe, dass der Mittelwert innerhalb des Bereiches der Messwerte liegen muss, genauer:
Wählt man die Indices so, dass $x_1 \leq x_2 \ldots \leq x_n$ ist, so sollte

$x_1 \leq$ Mittelwert $\leq x_n$

sein. Also Sonderfall ergibt sich: der Mittelwert von lauter gleichen Werten ist auch wieder gleich diesem Wert. Das ist für alle genannten Mittelwerte erfüllt.

Für die obere Grenze des Mittelwertes ergibt sich (analog für die untere Grenze):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \leq \frac{1}{n}\sum_{i=1}^{n} x_n = \frac{1}{n}\cdot(n\cdot x_n) = x_n$$

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} \leq \sqrt[n]{\prod_{i=1}^{n} x_n} = \sqrt[n]{(x_n)^n} = x_n$$

$$\bar{x}_H = \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i} \leq \frac{\sum_{i=1}^{n} c_i \cdot x_n}{\sum_{i=1}^{n} c_i} = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} c_i}\cdot x_n = x_n$$

Ferner sollte ein Mittelwert von der Maßeinheit unabhängig sein, genauer: multipliziert man alle Messwerte mit derselben Zahl a, so ist auch der daraus gebildete Mittelwert das a-fache des ursprünglichen Mittelwertes. Misst man also z. B. Entfernungen in km statt in Meilen, so sollte auch der neue Mittelwert einfach das 1,609-fache des ursprünglichen Mittelwertes sein.
Werden alle Beobachtungswerte mit einer Konstanten a multipliziert und ist $y_i = a\cdot x_i$, so folgt:

$$a\cdot\bar{x} = a\cdot\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} a\cdot x_i = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

$$a \cdot \bar{x}_G = a \cdot \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{a^n \cdot \prod_{i=1}^{n} x_i} = \sqrt[n]{\prod_{i=1}^{n} (a \cdot x_i)} = \sqrt[n]{\prod_{i=1}^{n} y_i} = \bar{y}_G \quad , \qquad a \geq 0$$

$$a \cdot \bar{x}_H = a \cdot \frac{\sum_{i=1}^{n} c_i \cdot x_i}{\sum_{i=1}^{n} c_i} = \frac{\sum_{i=1}^{n} a \cdot c_i \cdot x_i}{\sum_{i=1}^{n} c_i} = \frac{\sum_{i=1}^{n} c_i \cdot y_i}{\sum_{i=1}^{n} c_i} = \bar{y}_H$$

Diese beiden naheliegenden Eigenschaften sind weitere Argumente dafür, dass

- beim arithmetischen Mittel die Summe durch n geteilt
- beim geometrischen Mittel aus dem Produkt die n-te Wurzel gezogen
- beim gewichteten arithmetischen Mittel durch die Summe der Gewichte geteilt

werden muss.

2.2.5 Modus

(2.2.5.-01) ***Definition***
Der **Modalwert** oder **Modus** oder **dichteste Wert** x_D ist die Merkmalsausprägung, die am häufigsten vorkommt. Es kann mehrere Modalwerte geben.
Bei qualitativen Merkmalen ist der Modus das einzige mögliche Lagemaß unter den hier aufgeführten Lagemaßen.

(2.2.5.-02) ***Beispiele***
Beispiele für den Modus bei qualitativen Merkmalen sind:

- Die Verteilung der Geschlechter bei Neugeborenen
- Die Verteilung der Blutgruppen
- Die Verteilung der Farben bei Gummibärchen

2.2.6 Quantile

Bei komparativen und quantitativen Merkmalen kann man ermitteln, wieviel Prozent einer Grundgesamtheit unterhalb eines bestimmten Wertes liegen. Dieser Wert heißt dann ein Quantil. Quantile dienen dazu, die Menge der Beobachtungswerte in annähernd gleichgroße Abschnitte zu unterteilen.

Im Folgenden bedeutet $\lceil n \cdot p \rceil$ die Zahl, die man erhält, wenn man $n \cdot p$ auf die nächste ganze Zahl aufrundet.

(2.2.6.-01) ***Definition***
Die n quantitativen Beobachtungswerte seien geordnet in Form einer Rangliste gegeben, also eine Variationsreihe: $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$.
Es sei $0 < p < 1$. Ein p-**Quantil** (oder $p \cdot 100$ %-Quantil) ist dann definiert als:

$\tilde{x}_p = \frac{1}{2}\cdot(x_{n\cdot p} + x_{n\cdot p+1})$,   falls n·p ganzzahlig

$\tilde{x}_p = x_{\lceil n\cdot p\rceil}$,   falls n·p nicht ganzzahlig

Sonderfälle:

p = 0,5: $\tilde{x}_{0,5} = \tilde{x}$ heißt **Median** (= 50 %-Quantil = 2. Quartil = 5. Dezil = 5. Dezentil)

$\tilde{x} = \frac{1}{2}(x_{0,5\cdot n} + x_{0,5\cdot n+1})$, falls 0,5·n ganzzahlig (also n gerade)

$\tilde{x} = x_{\lceil 0,5\cdot n\rceil} = x_{0,5\cdot n+0,5}$, falls 0,5·n nicht ganzzahlig (also n ungerade)

$\tilde{x}_{0,25}$, $\tilde{x}_{0,50}$, $\tilde{x}_{0,75}$ heißen 1., 2., 3. **Quartil** (= 25 %-, 50 %-, 75 %-Quantil)

$\tilde{x}_{0,1}$, $\tilde{x}_{0,2}$, … , $\tilde{x}_{0,9}$ heißen 1., 2., … 9. **Dezentil** oder **Dezil** (= 10 %-, 20 %-, .. , 90 %-Quantil)

Es gilt näherungsweise: p·100 % der Daten sind kleiner oder gleich $\tilde{x}_p$.

(2.2.6.-02) ***Beispiel***

Würfelergebnisse wie im Beispiel (2.2.1.-04), also sortiert: 1,3,3,3,4,4,5,5,5,6. Dann ist:

p = 0,25 (1. Quartil): 10·0,25 = 2,5 ist nicht ganzzahlig, also ist

$\tilde{x}_{0,25} = x_{\lceil 10\cdot 0,25\rceil} = x_{\lceil 2,5\rceil} = x_3 = 3$.

p = 0,5 (Median = 2. Quartil = 5. Dezentil): 10·0,5 = 5 ist ganzzahlig, also ist

$\tilde{x}_{0,5} = \tilde{x} = \frac{1}{2}\cdot(x_5 + x_6) = \frac{1}{2}\cdot(4 + 4) = 4$.

2.3 Streumaße

Hinweis: Statt Streumaße werden auch die Begriffe Streuungsmaße bzw. Streu-/Streuungsparameter verwendet.

Hat man eine große Menge von Daten, so dienen Streumaße zum einen dazu, einen ersten Eindruck zu haben, wie dicht die Daten zusammenliegen. Zum anderen helfen sie aber wie auch bei den Lagemaßen, mehrere große Datenmengen grob vergleichbar zu machen, indem man die jeweiligen Streumaße miteinander vergleicht.

Die Streuungsmaße geben an, wie stark die Beobachtungswerte verteilt sind. Ein kleiner Wert gibt an, dass die Beobachtungswerte dicht beieinanderliegen, ein großer Wert gibt an, dass die Beobachtungswerte weit verteilt sind.

Beispiele:

- Bei der Produktion von Kolben für Motoren sind nur sehr geringe Streuungen um den vorgegebenen Wert für den Durchmesser zulässig.
- Bei mehreren Schüssen treffen gute Schützen die Scheibe mit wenig Streuung um den Scheibenmittelpunkt, bei schlechten Schützen ist die Streuung groß.

Es gibt eine Vielzahl von Streumaßen. Einige davon werden im Folgenden mit ihren Eigenschaften dargestellt.

(2.3.-01) ***Definition***
Spannweite = Differenz zwischen größtem und kleinstem Beobachtungswert

$R = x_{max} - x_{min}$

- Dimension (Maßeinheit): dieselbe, wie bei den Beobachtungswerten
- Die Spannweite ist zwar sehr einfach zu berechnen, hat aber den Nachteil, dass sie nur von 2 Werten abhängig ist, d.h., alle anderen Werte können beliebig verteilt sein, ohne dass dies die Spannweite beeinflusst.

(2.3.-02) ***Definition***
Mittlere absolute Abweichung = mittlere lineare Abweichung:

$$s_M = \frac{1}{n}\sum_{i=1}^{n}|x_i - M| \qquad \text{mit } M = \bar{x} \text{ oder } M = \tilde{x} \text{ oder } M = x_D$$

- Dimension (Maßeinheit): dieselbe, wie bei den Beobachtungswerten
- Es werden alle Beobachtungswerte berücksichtigt
- s_M ist das arithmetische Mittel der absoluten Abweichungen der Beobachtungswerte von M

<u>Vereinfachte Formeln im Falle $M = \bar{x}$:</u>

$$s_{\bar{x}} = \frac{2}{n}\sum_{x_i<\bar{x}}(\bar{x} - x_i) \qquad \text{nur die Werte } x_i < \bar{x} \text{ werden in der Summe berücksichtigt}$$

$$s_{\bar{x}} = \frac{2}{n}\sum_{x_i>\bar{x}}(x_i - \bar{x}) \qquad \text{nur die Werte } x_i > \bar{x} \text{ werden in der Summe berücksichtigt}$$

Beide Formeln führen zu demselben Ergebnis.

Beweis der vereinfachten Formeln:
Bei der Herleitung des arithmetischen Mittels (Kapitel 2.2.1) wurde gezeigt:
Summe der Abstände nach links = minus Summe der Abstände nach rechts.
In Formeln bedeutet das:

$$\sum_{x_i<\bar{x}}(\bar{x} - x_i) = -\sum_{x_i>\bar{x}}(\bar{x} - x_i)$$

$$\Leftrightarrow \quad \sum_{x_i<\bar{x}}(\bar{x} - x_i) = \sum_{x_i>\bar{x}}(x_i - \bar{x})$$

Damit folgt

$$s_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{n}\cdot\left(\sum_{x_i<\bar{x}}(\bar{x} - x_i) + \sum_{x_i>\bar{x}}(x_i - \bar{x})\right) = \frac{2}{n}\cdot\sum_{x_i<\bar{x}}(\bar{x} - x_i)$$

Und analog für die zweite vereinfachte Formel.
Durch die vereinfachten Formeln verringert sich der Rechenaufwand um mindestens 50 %, wenn man die Formel mit der geringeren Zahl an Summanden wählt.

<u>Beispiel:</u>

In einer Abteilung mit 7 Mitarbeitern beträgt die Betriebszugehörigkeit in Jahren:

i (Mitarbeiternr.)	1	2	3	4	5	6	7
Jahre	2	5	4	3	6	1	28

Die durchschnittliche Betriebszugehörigkeit ist:

$\bar{x} = \frac{2+5+4+3+6+1+28}{7} = 7$ [Jahre]

Dann ergibt sich mit der …

… ursprünglichen nicht vereinfachten Formel:

$s_{\bar{x}} = \frac{1}{7} \cdot (|2–7| + |5–7| + |4–7| + |3–7| + |6–7| + |1–7| + |28–7|) = \frac{1}{7} \cdot 42 = 6$,

… der ersten vereinfachten Formel:

$s_{\bar{x}} = \frac{2}{7} \cdot ((7–2) + (7–5) + (7–4) + (7–3) + (7–6) + (7–1)) = \frac{2}{7} \cdot 21 = 6$

… der zweiten vereinfachten Formel:

$s_{\bar{x}} = \frac{2}{7} \cdot (28–7) = \frac{2}{7} \cdot 21 = 6$

Benutzt man die zweite vereinfachte Formel, so hat man den mit Abstand geringsten Aufwand zur Berechnung von $s_{\bar{x}}$.

(2.3.-03) ***<u>Definition</u>***

Varianz = **mittlere quadratische Abweichung** vom arithmetischen Mittel:

$$\mathrm{Var}(X) = \sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i{}^2 - \bar{x}^2 = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i{}^2 \right) - \bar{x}^2$$

- Dimension (Maßeinheit): Quadrat der Dimension der Beobachtungswerte, also im Allgemeinen nicht sinnvoll interpretierbar
- Es werden alle Beobachtungswerte berücksichtigt. Wegen der Quadrierung werden Ausreißer besonders stark berücksichtigt, während Beobachtungswerte, die einen Abstand < 1 von $\bar{x}$ haben, besonders schwach berücksichtigt werden.
- Die Varianz ist das arithmetische Mittel der quadrierten Abweichungen der Beobachtungswerte von $\bar{x}$.
- Beweis der Umformung:

$$\begin{aligned}
\mathrm{Var}(X) = \sigma^2 &= \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2 \right) \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \cdot 2\bar{x} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot \sum_{i=1}^{n} \bar{x}^2 \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - 2\bar{x} \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot n\bar{x}^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - 2\bar{x} \cdot \bar{x} + \bar{x}^2 \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \bar{x}^2
\end{aligned}$$

(2.3.-04) ***<u>Definition</u>***

Standardabweichung vom arithmetischen Mittel:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var(X)}}$$

- Dimension (Maßeinheit): dieselbe wie bei den Beobachtungswerten
- Es werden alle Beobachtungswerte berücksichtigt. Der Vorteil gegenüber der Varianz ist, dass dieselbe Dimension wie bei den Beobachtungswerten vorliegt und dass die Empfindlichkeit gegenüber Ausreißern geringer ist.
- Die Standardabweichung ist das am häufigsten benutzte Streuungsmaß (daher der Name Standardabweichung), da sie mathematisch einfacher zu handhaben ist als die mittlere absolute Abweichung.

(2.3.-05) ***Vergleich*** **mittlere absolute Abweichung <-> Standardabweichung**

- Betrag der Abweichung:
  Bei beiden Streumaßen ist nur der Betrag, aber nicht das Vorzeichen der Abweichung wichtig:
  - Mittlere absolute Abweichung: Vorzeichen werden durch die Betragsstriche neutralisiert
  - Standardabweichung: Vorzeichen werden durch die Quadrierung neutralisiert

- Größenvergleich bei Abweichung vom arithmetischen Mittel $\bar{x}$:
  - Sind alle Messwerte gleich, so ist $s_{\bar{x}} = \sigma = 0$
  - Gibt es unterschiedliche Messwerte, so ist $s_{\bar{x}} < \sigma$.
    (ohne Beweis)

- Minimaleigenschaft:
  Die mittlere absolute Abweichung ist bezüglich des Medians minimal, die Standardabweichung ist bezüglich des arithmetischen Mittels minimal. Damit ist gemeint:
  - Minimaleigenschaft der mittleren absoluten Abweichung: Ersetzt man den Median $\tilde{x}$ durch einen Wert z, so bleibt der Ausdruck gleich oder er wird größer, d. h., für $z = \tilde{x}$ ist der Ausdruck minimal:

$$s_{\tilde{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}| \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n}|x_i - z| \qquad \text{für beliebige } z$$

Beweis:

Es sei sign die Funktion, die einer Zahl ihr Vorzeichen zuordnet. Diese Funktion ist nur im Nullpunkt nicht differenzierbar, ansonsten ist die Ableitung = 0. Im Folgenden sei z ungleich jedem x_i, anderenfalls kann der Summand $x_i - z$ einfach weggelassen werden.

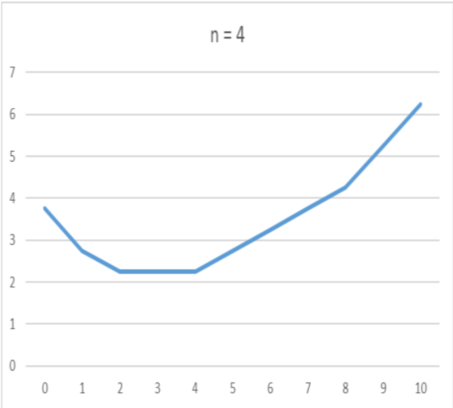
$$f(z) = \frac{1}{n}\sum_{i=1}^{n}|x_i - z| = \frac{1}{n}\sum_{i=1}^{n}\mathrm{sign}(x_i - z)\cdot(x_i - z)$$

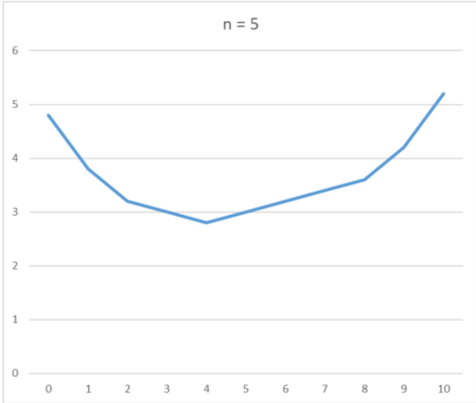
Dann folgt nach der Produktregel:

$$f'(z) = \frac{1}{n}\sum_{i=1}^{n}(0\cdot(x_i - z) + \big(\mathrm{sign}(x_i - z)\cdot(-1)\big) = 0$$

$$f'(z) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{sign}(z - x_i) = 0$$

Ist n gerade, so ist die Gleichung nur dann erfüllt, wenn die Hälfte der Beobachtungswerte kleiner und die andere Hälfte größer als z ist, ähnlich bei ungeradem n, z muss also der Median sein.
Beispiele für
n = 4 (linkes Diagramm): $x_i = 1, 2, 4, 8$; $\tilde{x} = 3$
n = 5 (rechtes Diagramm): $x_i = 1, 2, 4, 8, 9$; $\tilde{x} = 4$





- Minimaleigenschaft der Standardabweichung: Ersetzt man das arithmetische Mittel $\bar{x}$ durch einen Wert $z \neq \bar{x}$, so wird der Ausdruck größer, d. h., für $z = \bar{x}$ ist der Ausdruck minimal:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} < \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - z)^2} \quad \text{für } z \neq \bar{x}$$

Beweis:

$$f(z) = \frac{1}{n}\sum_{i=1}^{n}(x_i - z)^2 = z^2 - \left(\frac{2}{n}\sum_{i=1}^{n} x_i\right)\cdot z + \left(\frac{1}{n}\sum_{i=1}^{n} x_i{}^2\right)$$

$$f'(z) = 2z - \left(\frac{2}{n}\sum_{i=1}^{n} x_i\right) = 0 \quad \Leftrightarrow \quad z = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Da der Graph von f eine nach oben offene Parabel ist, liegt nur an der Stelle $z = \bar{x}$ ein absolutes Minimum vor und das war zu zeigen.

(2.3.-06) ***Definition***
Variationskoeffizient = Standardabweichung im Verhältnis zum arithmetischen Mittel:

$v_\sigma = \frac{\sigma}{\bar{x}}$, $\bar{x} \neq 0$

- Dimension (Maßeinheit): keine, da Zähler und Nenner dieselbe Dimension haben
- Der Vorteil des Variationskoeffizienten ist, dass die Streuung im Verhältnis zum Mittelwert berücksichtigt wird, d. h., es wird die prozentuale Streuung gemessen.
- Beispiel:
  - Mittelwert_1 = 10, Standardabweichung_1 = 1; v_σ = 10 %
  - Mittelwert_2 = 1000, Standardabweichung_2 = 100; v_σ = 10 %

Die absoluten Streuungen sind in den beiden Fällen stark unterschiedlich, die relativen Streuungen sind dagegen gleich. Hat man z. B. die Preise von Äpfeln und von Autos, so ist die absolute Streuung deutlich unterschiedlich, die relative Streuung bezogen auf das arithmetische Mittel kann dagegen ähnlich sein.
Insbesondere ist es für den Variationskoeffizienten unerheblich, in welcher Einheit gemessen wird (also z. B., ob in kg oder Tonnen gemessen wird).
Die prozentuale Schwankung ist z. B. auch bei Aktienkursen interessant.

2.4 Klassierungen

Klassierungen werden z. B. in [BG], [SJ], [WM] und natürlich auch im Internet ausführlich behandelt. Die Aussagen im vorliegenden Kapitel weichen zum Teil erheblich von den Aussagen in diesen Quellen ab.

2.4.1 Grundlagen

Oft wird bei Umfragen nur abgefragt, in welchem Intervall sich ein bestimmter Wert befindet.
Wird z. B. gefragt „Wie hoch ist Ihr Jahreseinkommen?“, so soll man nicht den exakten Wert angeben, sondern nur ein Intervall ankreuzen, also z. B.: „zwischen 10.000 und 15.000 €“. Der Bereich der Einkommen wird also in einzelne Intervalle oder Klassen eingeteilt.

Es ist also manchmal zweckmäßig, den gesamten Bereich der Beobachtungswerte überschneidungsfrei und lückenlos in Intervalle einzuteilen, jeder Beobachtungswert liegt dann in genau einem Intervall.

(2.4.1.-01) ***Definition***
Eine Einteilung eines Intervalls in mehrere Intervalle heißt eine **Klassierung** oder **Klasseneinteilung**, wenn die Einteilung vollständig und überschneidungsfrei ist. Die Einzelintervalle heißen auch **Klassen**. Jeder Klasse wird dann die Anzahl der darin befindlichen Beobachtungswerte zugeordnet.
Eine **klassierte Häufigkeitsverteilung** besitzt die Datenstruktur
(Klasse, absolute Häufigkeit der Beobachtungswerte in der Klasse)
Die **Klassenbreite** ist die Länge des Intervalls.
Die **Klassenmitte** ist die Mitte des Intervalls.
Die **relative Häufigkeit** einer Klasse ist $\frac{\text{Anzahl Beobachtungswerte in der Klasse}}{\text{Gesamtzahl Beobachtungswerte}}$.
Die **Klassendichte** (oder **Dichte der Klasse**) ist $\frac{\text{Anzahl Beobachtungswerte in der Klasse}}{\text{Klassenbreite}}$.
Die **Modalklasse ist** die Klasse mit der größten Dichte.
Die **absolute Summenhäufigkeit** einer bestimmten Klasse ist die Anzahl der Elemente aller Klassen $\leq$ der bestimmten Klasse.

(2.4.1.-02) ***Hinweis***
Die Modalklasse kann mehrfach vorkommen. In manchen Veröffentlichungen wird als Modus einer Klassierung die Mitte der Modalklasse definiert. Falls die Urliste noch vorliegt, hätte man dann aber

zwei verschiedene Arten von Modalwerten: einmal den häufigsten Wert aus der Urliste und zum anderen die Mitte der Modalklasse. Um Widersprüche zu vermeiden, sollte man also nur die übliche Definition des Modus als häufigster Wert benutzen.

(2.4.1.-03) ***Hinweis***
Der Unterschied zur Gruppierung ist:
Bei der Gruppierung bleibt jeder einzelne Beobachtungswert exakt erhalten, nur die ursprüngliche Beobachtungsreihenfolge geht verloren. Bei der Klassierung geht zusätzlich der exakte Beobachtungswert verloren, es bleibt nur noch die Information über das Intervall, in dem der Beobachtungswert liegt, erhalten.

(2.4.1.-04) ***Beispiel***

Klasse	**Häufigkeit**	**Klassenbreite**	**Klassenmitte**	**Relative Häufigkeit**	**Klassendichte**	**Absolute Summenhäufigkeit**
[0; 10[	5	10 – 0 = 10	$\frac{0+10}{2} = 5$	$\frac{5}{24} =$ 0,2083	$\frac{5}{10} = 0{,}50$	5
[10; 100[	9	100 – 10 = 90	$\frac{10+100}{2} = 55$	$\frac{9}{24} =$ 0,3750	$\frac{9}{90} = 0{,}10$	5+9 = 14
[100; 300[	10	300 – 100 = 200	$\frac{100+300}{2} = 200$	$\frac{10}{24} =$ 0,4167	$\frac{10}{200} = 0{,}05$	5+9+10 = 24

Z. B. liegen im Intervall [0; 10[insgesamt 5 Beobachtungswerte und analog in den anderen Klassen.
Die Modalklasse ist die Klasse [0; 10[, da sie die höchste Dichte hat.

Für die Analyse von Klassierungen ist wichtig, ob Zusatzinformationen vorliegen.

(2.4.1.-05) ***Fallunterscheidung***

<u>Fall 1:</u> Klassierung ohne Zusatzinformationen
Über die Verteilung der Werte innerhalb der Klassen ist in diesem Fall nichts bekannt, insbesondere liegt keine Urliste mit exakten Werten vor. Dieser Fall tritt z. B. ein, wenn für das betrachtete Merkmal nur Bandbreiten erhoben werden, wie z. B. bei Gehaltsumfragen. Da für die Beobachtungswerte nur Intervalle bekannt sind, können trivialerweise auch für alle daraus abgeleiteten Werte, wie z. B. das arithmetische Mittel, nur Intervalle angegeben werden.

<u>Fall 2:</u> Klassierung mit Zusatzinformationen
Diese Zusatzinformationen haben zwar einen geringeren Informationsgehalt als die Urliste, aber in manchen Fällen lassen sie sich bei der Bestimmung der abgeleiteten Werte nutzen, z. B. wenn die arithmetischen Mittel pro Klasse vorliegen.

<u>Fall 3:</u> Klassierung mit Urliste
Die Klassierung wurde aus einer Urliste abgeleitet und diese Urliste liegt vor. Die Klassierung dient dabei nur zur besseren Übersicht. In diesem Fall ermittelt man alle abgeleiteten Werte wie z. B. Lage- und Streumaße direkt aus der Urliste. Dieser Fall muss daher nicht weiter untersucht werden.

2.4.2 Klassierung ohne Zusatzinformationen

In diesem Abschnitt wird von „Fall 1“ aus (2.4.1.-05) ausgegangen. Es gilt die triviale Feststellung: Wenn man von den Beobachtungswerten nur Intervalle kennt, dann kann man daraus für die Lage- und Streuparameter auch nur Intervalle ableiten.

Die Verfahren werden anhand des Beispiels (2.4.1.-04) erläutert. Die 24 Beobachtungswerte sind aufsteigend nummeriert, also: $x_1 \leq x_2 \leq \ldots \leq x_{23} \leq x_{24}$.

(2.4.2.-01) ***Arithmetisches Mittel***

Die Untergrenze erhält man, indem man das arithmetische Mittel der Untergrenzen der Intervalle berechnet, also: $\bar{x}_u = \frac{5 \cdot 0 + 9 \cdot 10 + 10 \cdot 100}{24} = \frac{1090}{24}$

Die Obergrenze erhält man, indem man das arithmetische Mittel der Obergrenzen der Intervalle berechnet, also: $\bar{x}_o = \frac{5 \cdot 10 + 9 \cdot 100 + 10 \cdot 300}{24} = \frac{3950}{24}$

Also insgesamt (auf „<“ und „≤“ achten!):

$$45{,}4167 \approx \frac{1090}{24} \leq \bar{x} < \frac{3950}{24} \approx 164{,}5833$$

Diese Grenzen für $\bar{x}$ sind unverbesserbar, d.h.:

- Zu jedem Wert für $\bar{x}$ innerhalb dieses Intervalls kann man eine Urliste finden, die zu der Klassierung konform ist und $\bar{x}$ als arithmetisches Mittel hat (tatsächlich gibt es zu jedem $\bar{x}$ sogar unendlich viele passende Urlisten)
- Zu jedem Wert für $\bar{x}$ außerhalb dieses Intervalls gibt es keine Urliste, die zu der Klassierung konform ist und $\bar{x}$ als arithmetisches Mittel hat.

Oder anders formuliert: ermittelt man für jede Urliste, die zu der Klassierung konform ist, das arithmetische Mittel, so ist die Menge dieser arithmetischen Mittel gerade das oben berechnete Intervall.

Analoge Aussagen zur Unverbesserbarkeit gelten auch für die Intervalle der folgenden Lageparameter.

(2.4.2.-02) ***Quartile***

Wegen n = 24 und p = 0,25 ist n·p = 6, also ganzzahlig. Also ist

$$\tilde{x}_{0,25} = \frac{1}{2}(x_6 + x_7),$$

das 1. Quartil liegt also im 2. Intervall:

$$10 \leq \tilde{x}_{0,25} < 100$$

Wegen n = 24 und p = 0,50 ist n·p = 12, also ganzzahlig. Also ist

$$\tilde{x}_{0,50} = \frac{1}{2}(x_{12} + x_{13}),$$

das 2. Quartil = der Median liegt also im 2. Intervall:

$$10 \leq \tilde{x}_{0,50} = \tilde{x} < 100$$

Wegen n = 24 und p = 0,75 ist n·p = 18, also ganzzahlig. Also ist

$$\tilde{x}_{0,75} = \frac{1}{2}(x_{18} + x_{19}),$$

das 3. Quartil liegt also im 3. Intervall:

$$100 \leq \tilde{x}_{0,75} < 300$$

Alle angegebenen Grenzen für die Quartile sind unverbesserbar.

(2.4.2.-03) ***Modus***

Da die konkreten Beobachtungswerte nicht bekannt sind, kann nicht ermittelt werden, welche Werte am häufigsten vorkommen. Es gilt also nur die triviale und unverbesserbare Aussage:

$$0 \leq x_D < 300$$

Es ist im Beispiel leicht, Urlisten anzugeben, die mit der Klassierung konform sind und für die der Modus z. B. 3,4 oder 297,61 ist. Insbesondere kann also grundsätzlich nicht, wie in manchen Veröffentlichungen dargestellt, der Modus approximativ bestimmt werden.

(2.4.2.-04) ***Streumaße***

Da von den Ausgangswerten nur Intervalle bekannt sind, können trivialerweise auch für Streumaße grundsätzlich nur Intervalle angegeben werden.

Die Berechnung von Streumaßen bei Klassierungen ist mühselig. Daher werden hier nur anhand des Standardbeispiels die unverbesserbaren Intervallgrenzen für die mittlere absolute Abweichung vom arithmetischen Mittel abgeleitet.

Die Beobachtungswerte seien wieder aufsteigend nummeriert, also: $x_1 \leq x_2 \leq \ldots \leq x_{23} \leq x_{24}$.

Klasse	Häufigkeit	Beobachtungswerte
[0; 10[	5	$x_1, \ldots, x_5$
[10; 100[	9	$x_6, \ldots, x_{14}$
[100; 300[	10	$x_{15}, \ldots x_{24}$

$$s_{\bar{x}} = \frac{1}{24}\sum_{i=1}^{24}|x_i - \bar{x}|$$

Im Folgenden wird zunächst an manchen Stellen zur Vereinfachung angenommen, dass die Beobachtungswerte die oberen Grenzen in den Intervallen erreichen können. Beim Endergebnis wird korrekt darauf geachtet, ob die Grenzen erreichbar sind oder nicht, beim Endergebnis passt also wieder alles.

Wie bereits in (2.4.2.-01) gezeigt wurde, ist in diesem Beispiel

$$45{,}4166\ldots = \frac{1090}{24} \leq \bar{x} < \frac{3950}{24} = 164{,}5833\ldots,$$

$\bar{x}$ liegt also im 2. oder 3. Intervall.

Ermittlung der *unverbesserbaren unteren Grenze* für $s_{\bar{x}}$:

Im Folgenden sei x definiert als ein $\bar{x}$, das ein minimales $s_{\bar{x}}$ ergibt.
Das minimale $s_{\bar{x}}$ erhält man, indem man annimmt, dass x in einem bestimmten Intervall liegt und dann

- die Werte in den Intervallen links davon maximal,
- die Werte im Intervall selbst = x,
- die Werte in den Intervallen rechts davon minimal

wählt. Auf dieser Basis werden alle Fälle systematisch durchprobiert.

Annahme: x liegt im 1. Intervall, also $0 \le x < 10$

$x = \frac{1}{24} \cdot (5 \cdot x + 9 \cdot 10 + 10 \cdot 100) \iff 24x = 5x + 1090 \iff x = \frac{1090}{19} = 57{,}3\ldots > 10$

Widerspruch zu $x < 10$!

Annahme: x liegt im 2. Intervall, also $10 \le x < 100$

$x = \frac{1}{24} \cdot (5 \cdot 10 + 9 \cdot x + 10 \cdot 100) \iff 24x = 9x + 1050 \iff x = \frac{1050}{15} = 70$ **passt!**

Annahme: x liegt im 3. Intervall, also $100 \le x < 300$

$x = \frac{1}{24} \cdot (5 \cdot 10 + 9 \cdot 100 + 10 \cdot x) \iff 24x = 10x + 950 \iff x = \frac{950}{14} = 67{,}\ldots < 100$

Widerspruch zu $x \ge 100$!

<u>Ergebnis:</u>

$s_{\bar{x}} > \frac{1}{24} \cdot (5 \cdot (70 - 10) + 9 \cdot (70 - 70) + 10 \cdot (100 - 70)) = \frac{600}{24} = 25$

Diese untere Grenze ist unverbesserbar, wie man an folgendem Beispiel sieht:

Klasse	Häufigkeit	Beobachtungswerte
[0; 10[	5	$x_1 = \ldots = x_5 = 9{,}99999$
[10; 100[	9	$x_6 = \ldots = x_{14} = 70$
[100; 300[	10	$x_{15} = \ldots = x_{24} = 100$

ergibt $\bar{x} = 69{,}99999\ldots$ und $s_{\bar{x}} = 25{,}00000\ldots$

Ermittlung der *<u>unverbesserbaren oberen Grenze</u>* für $s_{\bar{x}}$:

Im Folgenden sei x ein $\bar{x}$, das ein maximales $s_{\bar{x}}$ ergibt.
Das maximale $s_{\bar{x}}$ erhält man, indem man annimmt, dass x in einem bestimmten Intervall liegt und dann

- die Werte in den Intervallen links davon minimal,
- die Werte im Intervall selbst = teilweise linke / teilweise rechte Grenze,
- die Werte in den Intervallen rechts davon maximal

wählt. Auf dieser Basis werden alle Fälle systematisch durchprobiert.

Liegt x im betrachteten Intervall, so liegen k Werte am unteren Rand und die restlichen Werte am oberen Rand, denn sonst kann $s_{\bar{x}}$ nicht maximal werden.

Annahme: x liegt im 1. Intervall, also $0 \le x < 10$ und $0 \le k \le 5$:

$x = \frac{1}{24} \cdot ((k \cdot 0 + (5 - k) \cdot 5) + 9 \cdot 100 + 10 \cdot 300) \iff x = \frac{3925 - 5k}{24}$

$0 \le x < 10 \iff 0 \le \frac{3925 - 5k}{24} < 10 \iff 737 < k \le 785$ Widerspruch!

x kann also nicht im 1. Intervall liegen (war wegen $\frac{1090}{24} \le \bar{x}$ sowieso klar)

Annahme: x liegt im 2. Intervall, also $10 \le x < 100$ und $0 \le k \le 9$:

$x = \frac{1}{24} \cdot (5 \cdot 0 + (k \cdot 10 + (9 - k) \cdot 100) + 10 \cdot 300) \iff x = \frac{3900 - 90k}{24}$

$10 \le x < 100 \iff 10 \le \frac{3900 - 90k}{24} < 100 \iff 16{,}66\ldots = \frac{50}{3} < k \le \frac{122}{3} = 40{,}66$ Widerspruch!

x kann also nicht im 2. Intervall liegen

Annahme: x liegt im 3. Intervall, also $100 \le x < 300$ und $0 \le k \le 10$:
$x = \frac{1}{24}\cdot(5\cdot 0 + 9\cdot 10 + (k\cdot 100 + (10-k)\cdot 300)) \iff x = \frac{3090-200k}{24}$
$100 \le x < 300 \iff 100 \le \frac{3000-200k}{24} < 300 \iff -21 < k \le 3$
Es ist also insgesamt $0 \le k \le 3$ und damit $103{,}75 = \frac{415}{4} \le x \le \frac{515}{4} = 128{,}75$
Dann ist

$$\begin{aligned}
s_{\bar{x}} &= \tfrac{1}{24}\cdot(5\cdot(x-0) + 9\cdot(x-10) + (k\cdot(x-100) + (10-k)\cdot(300-x))) \\
&= \tfrac{1}{24}\cdot(5\cdot(\tfrac{3090-200k}{24} - 0) + 9\cdot(\tfrac{3090-200k}{24} - 10) + (k\cdot(\tfrac{3090-200k}{24} - 100) + (10-k)\cdot(300 - \tfrac{3090-200k}{24}))) \\
&= \tfrac{1}{24^2}\cdot(5\cdot(3090-200k) + 9\cdot(2850-200k) + k\cdot(690-200k) + (10-k)\cdot(4110+200k)) \\
&= \tfrac{10}{24^2}\cdot(1545 - 100k + 2565 - 180k + 69k - 20k^2 + 4110 + 200k - 411k - 20k^2) \\
&= \tfrac{10}{24^2}\cdot(-40k^2 - 422k + 8220) \\
&\le \tfrac{10}{24^2}\cdot 8220 \qquad \text{, da } k = 0 \text{ zum Maximum führt} \\
&= \tfrac{3425}{24} = 142{,}7083\ldots
\end{aligned}$$

k = 0 bedeutet: kein Beobachtungswert im 3. Intervall liegt an der unteren Grenze des Intervalls = 100, alle Beobachtungswerte im 3. Intervall liegen an der oberen Grenze des Intervalls = 300.

Diese obere Grenze für $s_{\bar{x}}$ ist unverbesserbar, wie man an folgendem Beispiel sieht:

Klasse	**Häufigkeit**	**Beobachtungswerte**
[0; 10[	5	$x_1 = \ldots = x_5 = 0$
[10; 100[	9	$x_6 = \ldots = x_{14} = 10$
[100; 300[	10	$x_{15} = \ldots = x_{24} = 299{,}99999$

Dann ist $\bar{x} = \frac{3089{,}999\ldots}{24} = 128{,}7499\ldots$ und $s_{\bar{x}} = \frac{3424{,}999\ldots}{24} = 142{,}7083\ldots$

Die <u>unverbesserbaren Schranken für $s_{\bar{x}}$</u> sind also:

$$\mathbf{25 < s_{\bar{x}} < \frac{3425}{24} = 142{,}7083\ldots}$$

2.4.3 Klassierung mit Zusatzinformationen

Der „Fall 2“ aus (2.4.1.-05) wird anhand eines Beispiels erläutert.

(2.4.3.-01) ***<u>Beispiel</u>***
Die folgende Tabelle ist ein Auszug aus einer Tabelle des Statistischen Bundesamtes (siehe [SB]). Die grau unterlegten Bereiche wurden zur Verdeutlichung hinzugefügt.

Klasse	**Anzahl Steuer-pflichtige**	**Aufkommen in 1.000 € = T€**			
		Unter-grenze	***Obergrenze***	***Mittelwert***	***tatsächlich***

[0; 500[	576.123	0	288.062	144.031	141.077
[500; 1.000[	351.328	175.664	351.328	263.496	251.472
[1.000; 2.500[	575.322	575.322	1.438.305	1.006.814	960.781
[2.500; 5.000[	475.358	1.188.395	2.376.790	1.782.593	1.704.907
[5.000; 10.000[	365.604	1.828.020	3.656.040	2.742.030	2.565.832
[10.000; 25.000[	211.516	2.115.160	5.287.900	3.701.530	3.127.746
[25.000; 50.000[	37.696	942.400	1.884.800	1.413.600	1.254.838
[50.000; 100.000[	9.146	457.300	914.600	685.950	609.317
[100.000; 1.000.000[	3.846	384.600	3.846.000	2.115.300	849.334
Summe	2.605.939	7.666.861	20.043.825	13.855.343	11.465.304

Die Spalten „Klasse“ und „Anzahl Steuerpflichtige“ bilden die Klassierung.

Die Spalten „Untergrenze“, „Obergrenze“ und „Mittelwert“ wurden aus den beiden ersten Spalten abgeleitet.
Z. B. in der 2. Zeile:

„Untergrenze“ $= 500 \cdot 351.328$ [€] = 175.664 [T€]
„Obergrenze“ $= 1000 \cdot 351.328$ [€] = 351.328 [T€]

Das Steueraufkommen in der Klasse [500; 1000] liegt also zwischen 175.664 T€ und 351.328 T€, der Mittelwert von Unter- und Obergrenze beträgt 263.496 T€.

Die Spalte „tatsächlich“ ist eine Zusatzinformation aus der Ausgangstabelle des Statistischen Bundesamtes und stellt das tatsächliche Steueraufkommen in diesem Intervall dar, z. B. in der zweiten Zeile 251.472 T€.

<u>Arithmetisches Mittel:</u>
Wenn Fall 1 aus (2.4.1.-05) vorliegen würde, also die Spalte „tatsächlich“ in der Tabelle nicht vorhanden wäre, so könnte nur folgendes unverbesserbare Intervall abgeleitet werden:

$$7.666.861 \leq \text{Gesamtaufkommen} < 20.043.825 \quad [\text{T€}]$$

Daraus ergibt sich für das arithmetische Mittel, also den durchschnittlichen Wert pro Steuerpflichtigem:

$$\frac{7.666.861 \cdot 1000}{2.605.939} \leq \bar{x} < \frac{20.043.825 \cdot 1000}{2.605.939} \quad [\text{€}],$$

gerundet:

$$2.942{,}07 < \bar{x} < 7.691{,}60 \quad [\text{€}]$$

Statt für das arithmetische Mittel ein Intervall anzugeben, kann man auch den Mittelpunkt dieses Intervalls und die maximal mögliche Abweichung angeben, in diesem Fall also:

Mitte des Intervalls für $\bar{x} = 5.316{,}83$ [€],

wobei die Abweichung der Intervallmitte vom tatsächlichen arithmetischen Mittel maximal $\pm 2.374{,}77$ [€] beträgt, also kurz:

$$\bar{x} = 5.316{,}83 \pm 2.374{,}77 \quad [\text{€}]$$

Da aber das Aufkommen pro Intervall bekannt ist, ergibt sich als tatsächlicher Wert:

$$\bar{x} = \frac{11.465.304 \cdot 1000}{2.605.939} = 4.399{,}68 \quad [\text{€}] ,$$

also zusammenfassend (Werte in €):

ohne Zusatzinformationen (ohne Spalte „tatsächlich“): $2.942{,}07 < \bar{x} < 7.691{,}60$
mit Zusatzinformationen (mit Spalte „tatsächlich“): $\bar{x} = 4.399{,}68$

nur die Intervallmitte (siehe nächstes Kapitel): $\bar{x} \approx 5.316,83$

2.4.4 Klassierungen mit Annahmen

Es ist wieder wie in Kapitel 2.4.2 eine Klassierung ohne Zusatzinformationen gegeben. Je nach Quelle werden für die Ermittlung der Lage- und Streuparameter unterschiedliche Annahmen über die mögliche Verteilung der Werte innerhalb der Klassen gemacht.

(2.4.4.-01) ***Annahmen***
Typische Annahmen sind:

- Das arithmetische Mittel pro Klasse ist die jeweilige Klassenmitte, meist mit einer der folgenden zusätzlichen Annahmen
    - Alle Werte pro Klasse sind gleich der Klassenmitte
    - Alle Werte sind in der jeweiligen Klasse gleichmäßig verteilt

  In beiden Fällen wird durch diese zusätzlichen Annahmen die Klassierung durch eine eindeutige Urliste ersetzt, was aber im Regelfall nicht erwähnt wird. Für diese exakte Urliste lassen sich dann leicht exakte Lage- und Streuparameter mit bereits bekannten Methoden ermitteln.
- Eine bestimmte stetige Kurve ist eine gute Approximation der Beobachtungswerte.

Zwar kann man auf Basis solcher Annahmen mit korrekten Methoden Ergebnisse über Lage- und Streuparameter ableiten, diese Ergebnisse sind aber aus folgenden Gründen praktisch nicht nutzbar:

- Die Annahmen sind nicht begründet, sondern stellen nur eine mögliche Verteilung der Werte innerhalb der Klassen dar, sind also nur ein Beispiel unter beliebig vielen anderen Beispielen. Es handelt sich um reine Gedankenexperimente, deren Bezug zur Realität offen bleibt.
- In der Praxis lässt sich in der Regel nicht feststellen, ob die Annahmen zutreffen. Tatsächlich sind die Annahmen in der Realität (fast) immer falsch, z. B. sind Daten (fast) nie gleichmäßig in den jeweiligen Intervallen verteilt.
- Sollten die Zahlen Grundlage für eine Entscheidung sein, so geht man unnötige Risiken ein, wenn man einfach nur die aus den Annahmen ableiteten konkreten Werte anstelle von Intervallen wie in Kapitel 2.4.2 nimmt. Aber nur auf Basis eines konkreten Intervalls bzw. der Intervallmitte mit Angabe des maximal möglichen Fehlers kann eine seriöse Beurteilung gemacht werden.

Wenn man die Klassierung nicht nur mit den exakten Methoden in Kapitel 2.4.2, sondern zusätzlich auch unter (spekulativen) Annahmen untersucht, kann das vielleicht als Gedankenexperiment interessant sein. Wenn man aber Klassierungen nur auf Basis von nicht verifizierbaren Annahmen untersucht, führt das zu Fehleinschätzungen.

(2.4.4.-02) ***Beispiel***
Nimmt man wie in vielen Veröffentlichungen in der Einkommenstabelle (2.4.3.-01) an, dass der Mittelwert pro Intervall ungefähr die Intervallmitte ist, so ergibt sich $\bar{x} \approx 5.316,83$ €. Es wird aber auf dieser Basis vermutlich kaum jemand auf die Idee kommen, dass der tatsächliche korrekte Wert $\bar{x} = 4.399,68$ € ist. Die üblichen Annahmen führen also zu falschen Ergebnissen.

(2.4.4.-03) ***Vergleich***

Man kann den Unterschied in den Methoden auch folgendermaßen darstellen:

- Methode gemäß Kapitel 2.4.2: Schätzwert des Lage- oder Streuparameters mit exakter Angabe des maximal möglichen Fehlers oder gleichwertig: der tatsächliche Lage- oder Streuparameter liegt in einem exakt definierten Intervall um den Schätzwert.
- Methode mit Annahmen: Schätzwert des Lage- oder Streuparameters ohne Angabe des maximal möglichen Fehlers oder gleichwertig: der tatsächliche Lage- oder Streuparameter liegt in einem undefinierten Intervall um den Schätzwert, könnte also beliebige Werte innerhalb der Klassierungsgrenzen annehmen.

Es gibt keinen Grund, die Fehlergrenzen bei den Schätzwerten wegzulassen, aber es gibt gute Gründe, die Fehlergrenzen anzugeben – insbesondere, wenn man bedenkt, dass sie mit einfachen Methoden der Schulmathematik leicht zu ermitteln sind. Warum also sollte man die Fehlergrenzen weglassen?

(2.4.4.-04) ***Annahme Gleichverteilung***

Wie bereits erwähnt wird in manchen Veröffentlichungen die Annahme getroffen, dass die Beobachtungswerte pro Intervall gleichmäßig verteilt sind. Konkret wird dabei angenommen:

- die Werte pro Klasse sind gleichverteilt, d. h., benachbarte Werte haben den Abstand $\frac{\text{Klassenbreite}}{\text{Häufigkeit}}$
- die Werte pro Klasse liegen symmetrisch zur Klassenmitte.

Was in den mir bekannten Veröffentlichungen nicht steht, ist, dass aus diesen Annahmen bereits eine eindeutige Urliste folgt: dazu beginnt man pro Klasse stets mit „linker Rand“ + „halber Abstand“, damit die Symmetrie zur Klassenmitte erzeugt wird und addiert dann den Abstand mehrfach dazu.

(2.4.4.-05) ***Beispiel zur Gleichverteilung***

Klasse	**Häufigkeit**	**Klassenmitte**	**Abstand**	**Eindeutige Urliste**
[0; 10[	5	5	$\frac{10}{5} = 2$	$x_1 = 1$; $x_2 = 3$; $x_3 = 5$; $x_4 = 7$; $x_5 = 9$
[10; 100[	9	55	$\frac{90}{9} = 10$	$x_6 = 15$; $x_7 = 25$; $x_8 = 35$; $x_9 = 45$; $x_{10} = 55$; $x_{11} = 65$; $x_{12} = 75$; $x_{13} = 85$; $x_{14} = 95$
[100; 300[	10	200	$\frac{200}{10} = 20$	$x_{15} = 110$; $x_{16} = 130$; $x_{17} = 150$; $x_{18} = 170$; $x_{19} = 190$; $x_{20} = 210$; $x_{21} = 230$; $x_{22} = 250$; $x_{23} = 270$; $x_{24} = 290$

Da es nur genau diese eine Urliste gibt, die beide Annahmen erfüllt, ersetzt man faktisch die Klassierung durch diese ganz bestimmte sehr einfach gebaute Urliste, es wird also künstlich der Fall 3 aus (2.4.1.-05) erzeugt.

Alle Lage- und Streumaße lassen sich dann natürlich mit den bekannten Methoden leicht aus dieser eindeutigen Urliste ermitteln; z. B. ergibt sich für den Median:

$$\tilde{x} = \tilde{x}_{0,5} = \frac{1}{2}\cdot(x_{12} + x_{13}) = \frac{1}{2}\cdot(75 + 85) = 80$$

Wegen der Annahme der Gleichverteilung lassen sich die einzelnen Werte leicht durch Polynome 1. Grades darstellen:

Klasse	Eindeutige Urliste	Formel
[0; 10[	$x_1 = 1$; $x_2 = 3$; $x_3 = 5$; $x_4 = 7$; $x_5 = 9$	$x_i = 2 \cdot i - 1$; $i = 1, \ldots, 5$
[10; 100[	$x_6 = 15$; $x_7 = 25$; $x_8 = 35$; $x_9 = 45$; $x_{10} = 55$; $x_{11} = 65$; $x_{12} = 75$; $x_{13} = 85$; $x_{14} = 95$	$x_i = 10 \cdot i - 45$; $i = 6, \ldots, 14$
[100; 300[	$x_{15} = 110$; $x_{16} = 130$; $x_{17} = 150$; $x_{18} = 170$; $x_{19} = 190$; $x_{20} = 210$; $x_{21} = 230$; $x_{22} = 250$; $x_{23} = 270$; $x_{24} = 290$	$x_i = 20 \cdot i - 190$; $i = 15, \ldots, 24$

Die Formel erhält man, indem man $x_i = a \cdot i + b$ setzt, wobei a der Abstand der Elemente ist. b erhält man, indem man irgendeinen Wert für i aus dem betreffenden Intervall einsetzt.

Um eine Gesamtformel zu errechnen, setzt man an

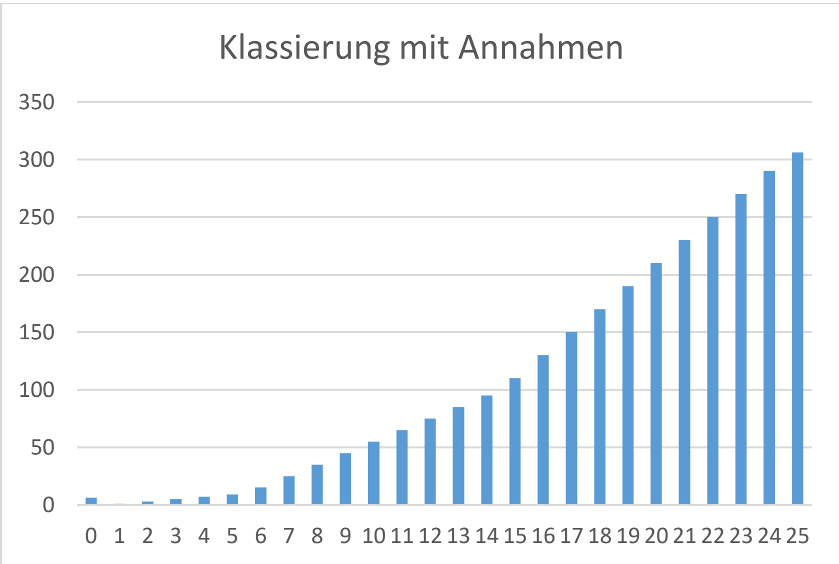
$$x_i = a \cdot |i - 0{,}5| + b \cdot |i - 5{,}5| + c \cdot |i - 14{,}5| + d \cdot |i - 24{,}5|$$

Dabei wurden in den Betragsstrichen von i die Mitte zwischen zwei aufeinander folgenden Klassen eingesetzt. Die Lösung lautet:

$$x_i = 7{,}25 \cdot |i - 0{,}5| + 4 \cdot |i - 5{,}5| + 5 \cdot |i - 14{,}5| - 3{,}75 \cdot |i - 24{,}5|$$

Setzt man in diese Gleichung für i die Werte 1, 2, …, 24 ein, so erhält man genau wieder die Urliste.

Als Balkendiagramm:



Statt einfach die eindeutige Urliste aufzuschreiben, aus der alle Lage- und Streumaße unmittelbar ableitbar sind, werden in vielen Veröffentlichungen Interpolationsmethoden oder Strahlensätze bemüht und dafür der irreführende Begriff „Feinberechnung“ oder „Feinanalyse“ benutzt. Tatsächlich muss aber nur eine exakte Urliste mit Standardmethoden analysiert werden.

In manchen Veröffentlichungen wird sogar ein Modus errechnet (siehe z. B. [BG], [EP]), obwohl das bei Klassierungen grundsätzlich noch nicht einmal näherungsweise möglich ist, denn:

- hat man keine Zusatzannahmen, so kommt jeder beliebige Wert innerhalb der Klassierungsgrenzen für den Modus in Frage
- hat man die Zusatzannahme der Gleichverteilung, so kommt jeder Wert in der Urliste genau einmal vor, ist also auch Modus.

Weitere gelegentlich zu findende Annahmen über die Verteilung der Daten in den Klassen können dann z. B. zur Sheppard Korrektur bei der Berechnung der Varianz klassierter Daten führen.

(2.4.4.-06) ***Das* ≈-Zeichen**

Manchmal wird das ≈-Zeichen benutzt, um zu dokumentieren, dass das Ergebnis nicht exakt den korrekten Wert widerspiegelt. Tatsächlich wird in diesen Fällen aber ausgedrückt, dass jede beliebige Zahl innerhalb der Klassierungsgrenzen möglich ist, man kann also eine solche Aussage in der Praxis nicht nutzen. Was sollte z. B. ≈100 bedeuten? Ist 110 noch abgedeckt? Oder 120? Oder 200? Erst durch die Angabe der Intervalle mit der exakten Methode können solche Fragen beantwortet werden.

Im Gegensatz dazu ist bei Ergebnissen von numerischen Verfahren ein bestimmtes Intervall unterstellt. So bedeutet z. B. $\pi \approx 3{,}14$ implizit, dass $3{,}13 < \pi < 3{,}15$ ist. Eine solche Ungleichung ist in den entsprechenden Veröffentlichungen bei Klassierungen aber nicht gemeint und wäre im Regelfall auch falsch.

2.5 Korrelation

Die Korrelation drückt aus, wie stark die Messwerte zweier Merkmale zusammenhängen. Die mathematischen Methoden der Korrelation werden hier nicht näher erläutert. Sie sind in den meisten Statistik-Büchern ausführlich beschrieben.

Typisch für eine Korrelation ist zum Beispiel, dass sich eine Größe in ähnlicher oder in genau gegenläufiger Weise verändert wie eine andere Größe. Daraus wird manchmal geschlossen, dass ein kausaler Zusammenhang besteht, was aber nicht der Fall sein muss. Eine solche Beobachtung ist lediglich ein Hinweis darauf, die Beziehung der beiden Größen näher zu überprüfen.

Es ist naheliegend und gilt gerade auch für Big Data: je mehr Daten man hat, umso eher kann man Größen finden, die eine ähnliche Entwicklung zeigen, obwohl sie tatsächlich nichts miteinander zu tun haben. Wer Statistik als zu trocken empfindet und unbedingt eine Auffrischung benötigt, dem empfehle ich, sich Seiten im Internet anzusehen, die kuriose Scheinkorrelationen (engl. „spurious correlation“) dokumentieren.

Korrelieren die Daten zweier Merkmale A und B, so kann es dafür grundsätzlich 4 Interpretationen geben. Dies wird an folgendem häufig zitierten Beispiel verdeutlicht, dass in einem bestimmten Zeitraum in einer bestimmten Gegend die Anzahl Störche in gleichem Maße zurückgegangen ist, wie die Anzahl der Geburten beim Menschen.

Interpretation 1: „A ist die Ursache von B“

Das Sinken der Zahl der Störche ist die Ursache für das Sinken der Zahl der Geburten, denn die Störche bringen bekanntlich die Babys.

Interpretation 2: „B ist die Ursache von A“

Das Sinken der Zahl der Geburten ist die Ursache für das Sinken der Zahl der Störche, denn Störche werden erst beim Anblick von Babys zur Fortpflanzung animiert.

Interpretation 3: „A und B haben eine gemeinsame Ursache“
Die Veränderung der gesellschaftlichen Verhältnisse, wie z. B. die Verstädterung, führt sowohl zur Zerstörung der Natur und damit zu weniger Störchen, als auch zu weniger Interesse an Kindern.

Interpretation 4: „die Korrelation ist zufällig, es gibt keinen Zusammenhang“.

3 Kombinatorik

Zusammenfassung
Die Kombinatorik gibt Antworten auf Fragen der Art: „Wie viele Möglichkeiten gibt es, …?“
Theoretisch könnte man bei jeder solchen Frage alle denkbaren Möglichkeiten aufschreiben und dann abzählen, wie viele davon die Bedingung aus der Frage erfüllen. Die Antwort an den Fragesteller lautet also: „Zähl’s halt ab!“ und damit ist das Thema „Kombinatorik“ vollständig und abschließend behandelt.
Aber weil das unnötigerweise viel zu anstrengend und ab einer gewissen Größenordnung auch nicht mehr praktisch durchführbar wäre, bleibt als Aufgabe, möglichst effiziente Verfahren zu entwickeln, um mit möglichst wenig Aufwand zu einer Antwort zu kommen.
* * *

3.1 Fakultät / Binomialkoeffizient

Zunächst ein paar Grundbegriffe.

(3.1.-01) ***Definition***
Eine **Permutation** von Elementen ist eine bestimmte Reihenfolge der Elemente.
Eine **Fakultät n!** (sprich: „n Fakultät“) ist definiert als:

$0! = 1$

$n! = n\cdot(n-1)!$ für $n \in \mathbb{N}$

Daraus ergibt sich für $n \in \mathbb{N}$: $n! = 1\cdot 2\cdot \ldots \cdot(n-1)\cdot n$

Die ersten Fakultäten sind:

$$\begin{aligned}
0! &= & 1 &= 1\\
1! &= 1\cdot 0! = & 1 &= 1\\
2! &= 2\cdot 1! = & 2\cdot 1 &= 2\\
3! &= 3\cdot 2! = & 3\cdot 2\cdot 1 &= 6\\
4! &= 4\cdot 3! = & 4\cdot 3\cdot 2\cdot 1 &= 24\\
5! &= 5\cdot 4! = & 5\cdot 4\cdot 3\cdot 2\cdot 1 &= 120\\
6! &= 6\cdot 5! = & 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 &= 720\\
7! &= 7\cdot 6! = & 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 &= 5040
\end{aligned}$$

<u>Zusammenhang Permutationen/Fakultäten</u>
Hat man n Elemente einer Menge, so gibt es n! Permutationen dieser Elemente.

Herleitung am Beispiel n = 4:
Hat man 4 Elemente a, b, c, d, so gibt es 4! = 24 Permutationen dieser Elemente:

- für die 1. Stelle hat man 4 Möglichkeiten,

- für jede der 4 Möglichkeiten an der 1. Stelle hat man für die 2. Stelle dann noch 3 Möglichkeiten, insgesamt also $4 \cdot 3$ Möglichkeiten
- für jede der $4 \cdot 3$ Möglichkeiten an der 1. und 2. Stelle hat man für die 3. Stelle dann noch 2 Möglichkeiten, insgesamt also $4 \cdot 3 \cdot 2$ Möglichkeiten
- für jede der $4 \cdot 3 \cdot 2$ Möglichkeiten an der 1., 2. und 3. Stelle hat man für die 4. Stelle dann noch 1 Möglichkeit, insgesamt also $4 \cdot 3 \cdot 2 \cdot 1$ Möglichkeiten

Insgesamt sind das dann $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$ Möglichkeiten

Durchführung:

a b c d	b a c d	c a b d	d a b c
a b d c	b a d c	c a d b	d a c b
a c b d	b c a d	c b a d	d b a c
a c d b	b c d a	c b d a	d b c a
a d b c	b d a c	c d a b	d c a b
a d c b	b d c a	c d b a	d c b a

Hat man also 4 Stühle und 4 Personen, so gibt es 4! = 24 Möglichkeiten, wie sich die Personen auf die Stühle verteilen können – natürlich immer genau eine Person pro Stuhl!

Bei einem Statistikkurs mit 30 Studierenden gibt es dann bereits $30! \approx 2{,}65 \cdot 10^{32}$ Sitzanordnungen, also ganz schön viel.

Eine wichtige Anwendung für Permutationen ist das "Travelling Salesman Problem" oder "Problem des Handlungsreisenden": ein Handlungsreisender soll eine Rundreise von seiner Heimadresse (wo er auch einen Kunden hat) aus machen und dabei alle n Städte, in denen er Kunden hat, genau einmal besuchen. Die Frage ist: welcher Weg hat die geringste Gesamtlänge? Das Problem ist deshalb so anspruchsvoll, weil es bei n Städten (n–1)! mögliche Wege gibt, man also nur bei kleinem n alle Weglängen explizit ausrechnen kann.

(3.1.-02) ***<u>Definition</u>***

$\binom{n}{k}$ heißt **Binomialkoeffizient** und ist definiert als:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \qquad k, n \in \mathbb{N}_0$$

Der Name „Binomialkoeffizient" kommt daher, weil er der Koeffizient in der allgemeinen binomischen Formel ist:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

(3.1.-03) ***<u>Beispiel</u>***

Für n = 2 ergibt sich:

$$\binom{2}{0} = \frac{2!}{0! \cdot (2-0)!} = \frac{2}{1 \cdot 2} = 1$$
$$\binom{2}{1} = \frac{2!}{1! \cdot (2-1)!} = \frac{2}{1 \cdot 1} = 2$$
$$\binom{2}{2} = \frac{2!}{2! \cdot (2-2)!} = \frac{2}{2 \cdot 1} = 1$$

Es ergibt sich damit die aus der Schule bekannte 1. binomische Formel:

$$(a+b)^2 = \sum_{k=0}^{2} \binom{2}{k} a^{2-k}\, b^k = \binom{2}{0} a^{2-0}\, b^0 + \binom{2}{1} a^{2-1}\, b^1 + \binom{2}{2} a^{2-2}\, b^2 = a^2 + 2ab + b^2$$

(3.1.-04) ***Eigenschaften der Binomialkoeffizienten***

- Die Binomialkoeffizienten sind zwar als Bruch definiert, dieser Bruch lässt sich aber immer vollständig kürzen, sodass immer eine natürliche Zahl oder 0 entsteht.

- $\binom{n}{0} = \binom{n}{n} = 1$

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

- $\binom{n}{k} = \binom{n}{n-k}$
  Beispiel: $\binom{50}{47} = \binom{50}{3} = \frac{50 \cdot 49 \cdot 48}{3 \cdot 2 \cdot 1} = 19.600$

- $\binom{n}{k} = \frac{\frac{n!}{(n-k)!}}{k!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1}$
  Im Zähler beginnt man mit n, im Nenner beginnt man mit k, und in beiden Fällen werden k Faktoren absteigend hingeschrieben. Dies ist der einfachste Weg zur Berechnung von Binomialkoeffizienten ohne Taschenrechner. Außerdem kann in dieser Darstellung n eine beliebige reelle Zahl sein, was in manchen Formeln genutzt wird.
  Beispiel: $\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13.983.816$

- $\binom{n}{k} = 0$, falls $n < k$
  Dies folgt aus der Formel vorher, da im Zähler der Faktor 0 auftritt.
  Beispiel:
  $\binom{2}{4} = \frac{2 \cdot 1 \cdot 0 \cdot (-1)}{4 \cdot 3 \cdot 2 \cdot 1} = 0$

3.2 Kombination / Variation

Ausgangspunkt ist die Frage: „Wie viele Möglichkeiten gibt es, n Elemente aus einer Menge von N Elementen auszuwählen?“.

Dabei sind folgende Zusatzbedingungen zu unterscheiden:

- die Reihenfolge wird berücksichtigt („**Variation**“) / nicht berücksichtigt („**Kombination**“)
- Elemente können nur einmal (ohne Zurücklegen / ohne Wiederholung) oder mehrmals (mit Zurücklegen / mit Wiederholung) ausgewählt werden

Es gibt also $2 \cdot 2 = 4$ Möglichkeiten für diese Zusatzbedingungen.
Die n ausgewählten Elemente werden zunächst als geordneter Vektor $(e_1, e_2, \ldots, e_n)$ geschrieben und dann die entsprechenden Möglichkeiten der Auswahl und Anordnung untersucht.

Hinweis: der Begriff „Kombination“ bedeutet in diesem Kontext eine ganz bestimmte Zusammenstellung von Elementen einer Menge. Bei den Bewertungen in Kapitel 4 bedeutet „Kombination“ eine Zusammenstellung von Objekten, die je nach Art der Objekte unterschiedlich sein kann.

(3.2.-01) ***Beispiel***
Man untersucht, welche Ergebnisse es beim zweimaligen Würfeln mit einem Würfel geben kann. Das Ergebnis wird kurz geschrieben als Vektor (Ergebnis 1. Wurf, Ergebnis 2. Wurf).
„Die Reihenfolge wird berücksichtigt“ heißt, dass z. B. (2, 5) ein anderer Fall ist als (5, 2).
„Die Reihenfolge wird nicht berücksichtigt“ heißt, dass z. B. (2, 5) und (5, 2) im Ergebnis als gleich angesehen werden, also wie ein Fall gezählt werden.

(3.2.-02) ***Problemstellung***
Wie viele Möglichkeiten gibt es, aus einer Menge mit N Elementen n Elemente mit/ohne Berücksichtigung der Reihenfolge und mit/ohne Wiederholung (=Zurücklegen) auszuwählen?
Im Folgenden werden die Lösungen der 4 Fälle hergeleitet.

(3.2.-03) ***Fall 1:*** Variation (= Reihenfolge wird berücksichtigt) ohne Wiederholung
Anzahl Möglichkeiten: $\frac{\mathbf{N!}}{\mathbf{(N-n)!}}$; $0 \leq n \leq N$

Herleitung der Formel:
In dem Vektor ($e_1, e_2, \ldots, e_n$) gibt es …
für die 1. Stelle N Möglichkeiten
für die 2. Stelle dann noch (N–1) Möglichkeiten
…
für die k. Stelle dann noch (N–n+1) Möglichkeiten
Insgesamt gibt es dann

$$N\cdot(N-1)\cdot\ldots\cdot(N-n+1) = \frac{[N\cdot(N-1)\cdot\ldots\cdot(N-n+1)]\cdot(N-n)\cdot(N-n-1)\cdot\ldots\cdot 1}{(N-n)\cdot(N-n-1)\cdot\ldots\cdot 1} = \frac{N!}{(N-n)!}$$ Möglichkeiten

Beispiel:
Unter 10 Personen werden 4 Gewinne verlost. In einer Lostrommel sind die 10 Lose mit den Namen der Teilnehmer. Auf das 1. gezogene Los fällt der 1. Preis, …, auf das 4. gezogene Los fällt der 4. Preis. Nach der Ziehung eines Loses wird das Los nicht wieder in den Korb zurückgelegt, jedes Los = jede Person gewinnt höchstens einmal.
Dann gibt es

$$10\cdot 9\cdot 8\cdot 7 = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1} = \frac{10!}{6!} = \frac{10!}{(10-4)!} = 5040$$

Möglichkeiten, wie die 4 Gewinne unter die 10 Personen verteilt werden können oder ausführlich:
Für den 1. Preis gibt es 10 mögliche Gewinner. Zu jedem dieser 10 Gewinner gibt es dann für den 2. Preis noch 9 Gewinner, das sind insgesamt 10·9 = 90 Möglichkeiten, den 1. und den 2. Preis unter die 10 Personen zu verteilen. Zu jeder dieser 90 Möglichkeiten gibt es dann 8 Möglichkeiten für den 3. Preis usw.

(3.2.-04) ***Fall 2:*** Variation (= Reihenfolge wird berücksichtigt) mit Wiederholung
Anzahl Möglichkeiten: $\mathbf{N^n}$; $0 \leq n$

Herleitung der Formel:
In dem Vektor (e_1, e_2, …, e_n) gibt es …
für die 1. Stelle N Möglichkeiten
für die 2. Stelle N Möglichkeiten
…
für die n. Stelle N Möglichkeiten
Insgesamt gibt es dann N^n Möglichkeiten

Beispiel:
Bei der Ermittlung einer 7-stelligen Zahl aus einzelnen Ziffern können dieselben Ziffern mehrmals auftreten. Wird eine Ziffer gezogen, so wird sie in den Korb zurückgelegt, damit für die nächste Stelle wieder alle 10 Ziffern zur Verfügung stehen. Ein konkretes Anwendungsbeispiel ist das „Spiel 77".

(3.2.-05) ***Fall 3:*** Kombination (= Reihenfolge wird nicht berücksichtigt) ohne Wiederholung
Anzahl Möglichkeiten: $\binom{\mathbf{N}}{\mathbf{n}}$; $0 \leq n \leq N$

Herleitung der Formel:
Da die Werte unterschiedlich sind, muss einfach nur der Wert aus Fall 1 durch n! geteilt werden, da es n! Permutationen der n ausgewählten Elemente gibt. Auf diese Permutationen kommt es in Fall 1 an, aber in Fall 3 nicht.
Insgesamt gibt es dann $\frac{N!}{(N-n)!\cdot n!} = \binom{N}{n}$ Möglichkeiten

Beispiel:
Bei der Ziehung der Lottozahlen ist die Reihenfolge beim Ziehen egal, ferner kann eine Zahl kann nur einmal gezogen werden. Die Anzahl der möglichen Kombinationen, wenn man 6 Zahlen aus 49 Zahlen (ohne Zurücklegen) zieht, ist: $\binom{49}{6} = 13.983.816$

Hinweis: Beim Lotto gibt es nur die Varianten „n Richtige mit richtiger / falscher Superzahl". Die Superzahl ist eine einstellige Zahl, es gibt also die 10 Superzahlen 0, …, 9. Die höchste Gewinnklasse 1 ist „6 Richtige mit richtiger Superzahl" und das ist genau eine von

$$\binom{49}{6}\cdot 10 = 139.838.160$$

Möglichkeiten. Daher gibt es bei Werbung immer den Hinweis: „Chance 1 : 140 Millionen".
Die nächsthöhere Gewinnklasse 2 ist: „6 Richtige mit falscher Superzahl" und das sind 9 von $\binom{49}{6}\cdot 10 = 139.838.160$ Möglichkeiten, also eine Gewinnchance 1 : 15.537.573,3.

(3.2.-06) ***Fall 4:*** Kombination (= Reihenfolge wird nicht berücksichtigt) mit Wiederholung
Anzahl Möglichkeiten: $\binom{\mathbf{N+n-1}}{\mathbf{n}}$; $0 \leq n$

Herleitung der Formel:
Die Formel wird anhand eines Beispiels hergeleitet.

Es wird viermal gewürfelt. Dann ist N = 6, n = 4. Die Reihenfolge der Werte wird nicht berücksichtigt, mehrfache Werte können vorkommen.

Die möglichen Ergebnisse werden nun in eine gymnastische Übung kodiert:
Gegeben sei eine Treppe mit 6 Stufen, wobei Stufe 1 ebenerdig ist. Für jede 1 wird nun eine Kugel auf die 1. Treppenstufe gelegt, …, für jede 6 wird eine Kugel auf die 6. Treppenstufe gelegt. Ferner gibt es 2 Arten von Anstrengungen: B = „Bücken und eine Kugel aufheben“ und S = „Eine Treppenstufe hochsteigen“.
Geht man die Treppe bis zur 6. Stufe hoch und sammelt dabei alle Kugeln ein, so sind die Anstrengungen unter Berücksichtigung ihrer Reihenfolge eine eindeutige Kodierung der 4 gewürfelten Zahlen und umgekehrt:
Aus 1, 3, 3, 4 folgt eindeutig B, S, S, B, B, S, B, S, S
Aus S, B, B, B, S, S, S, S, B folgt eindeutig 2, 2, 2, 6
Es gibt also insbesondere genauso viele Ziffernkombinationen wie gymnastische Übungen.

Das Problem lautet dann anders formuliert: wie viele verschiedene gymnastische Übungen gibt es?

Jede gymnastische Übung besteht aus 5 S (nämlich von 1. Stufe zu 6. Stufe) und 4 B (da es 4 gewürfelte Zahlen, also 4 Kugeln sind), also 9 Anstrengungen (allgemein: (N–1) + n Anstrengungen). Bei 9 Anstrengungen gibt es $\binom{9}{4}$ Stellen, an denen man die 4 B platzieren kann oder $\binom{9}{5}$ Stellen, an denen man die 5 S platzieren kann, was wegen $\binom{9}{4} = \binom{9}{5}$ dasselbe ist.
Allgemein gibt es also $\binom{N-1+n}{n} = \binom{N-1+n}{N-1}$ Möglichkeiten.

<u>Beispiel:</u>
Mit einem Würfel werde fünfmal gewürfelt. Die Reihenfolge der Ergebnisse der einzelnen Würfe ist egal, dieselbe Augenzahl kann mehrfach vorkommen. Alternativ kann auch wie beim Spiel „Kniffel“ mit fünf nicht unterscheidbaren Würfeln auf einmal gewürfelt werden.
Z. B. ist (1, 2, 3, 4, 5) derselbe Fall wie (5, 4, 3, 2, 1), da es nicht auf die Reihenfolge ankommt, bzw. da die Würfel nicht unterscheidbar sind.
Dann ist N = 6, n = 5 und die Anzahl der möglichen Ergebnisse ist $\binom{N-1+n}{n} = \binom{10}{5} = 252$.

(3.2.-07) ***<u>Hinweis</u>***
Bei kombinatorischen Fragestellungen, die sich nicht unmittelbar mit einem der 4 oben genannten Fälle identifizieren lassen, geht man oft folgendermaßen vor (siehe auch Fälle 3 und 4):
- Schritt 1:
  Unabhängig von der konkreten Aufgabenstellung werden zunächst werden alle Objekte als unterschiedlich angesehen, man kann sich z. B. vorstellen, dass sie durchnummeriert sind oder unterschiedliche Farben haben. Ferner wird auf die Reihenfolge Rücksicht genommen.
  Auf dieser Basis wird die Anzahl Möglichkeiten errechnet.
- Schritt 2:
  Erst danach werden Bedingungen wie „die Objekte … sind nicht unterscheidbar“ oder „es kommt nicht auf die Reihenfolge an“ rechnerisch berücksichtigt.

(3.2.-08) ***<u>Hinweis</u>***
Es ist egal, ob man einen Versuch mehrfach parallel oder hintereinander durchführt, sofern sich die Versuche nicht gegenseitig beeinflussen. Insbesondere ist es egal, ob man z. B. mit fünf Würfeln gleichzeitig würfelt oder ob man mit einem Würfel fünfmal hintereinander würfelt:

- Fall 1: wie bei Hinweis 1 beschrieben, wählt man fünf unterscheidbare Würfel. Die Unterscheidung kann man z. B. durch Farben darstellen: der erste Würfel ist rot, der zweite grün, usw. Die Ergebnisse schreibt man als Vektor: $(w_1, w_2, w_3, w_4, w_5)$, wobei w_1 das Ergebnis des ersten (roten) Würfels, w_2 das Ergebnis des zweiten (grünen) Würfels etc. ist. Man erhält dann $6{\cdot}6{\cdot}6{\cdot}6{\cdot}6 = 6^5$ mögliche Ergebnisse.
- Fall 2: würfelt man mit einem Würfel fünfmal, so lässt sich das Ergebnis ebenfalls als $(w_1, w_2, w_3, w_4, w_5)$ schreiben, wobei w_1 das Ergebnis des ersten Wurfes, w_2 das Ergebnis des zweiten Wurfes, etc. ist. Auch in diesem Fall erhält man $6{\cdot}6{\cdot}6{\cdot}6{\cdot}6 = 6^5$ mögliche Ergebnisse.

Zum Beispiel ist (2, 4, 5, 3, 2) interpretierbar als:

- Fall 1: mit dem ersten (roten) Würfel wurde eine 2, mit dem zweiten (grünen) Würfel eine 4, … gewürfelt
- Fall 2: beim ersten Wurf wurde eine 2, beim zweiten Wurf eine 4, … gewürfelt.

Beide Fälle sind vollkommen gleichwertig.

3.3 Kombination/Variation mit Zusatzeigenschaft

Als Verallgemeinerung zu Kapitel 3.2 wird in diesem Kapitel zusätzlich davon ausgegangen, dass manche Elemente der Grundgesamtheit eine Zusatzeigenschaft E haben und andere nicht, also z. B.: einige Objekte sind rot und die anderen sind nicht rot. Man erhält wieder die Formeln aus Kapitel 3.2, wenn man die Anzahl der Elemente mit der Eigenschaft E gleich 0 oder gleich N setzt. Die Grundgesamtheit wird durch die Zusatzeigenschaft in zwei disjunkte Teilmengen aufgeteilt.

Die hier dargestellten Methoden lassen sich ohne Weiteres auf den Fall erweitern, dass in der Grundgesamtheit mehrere Eigenschaften unterschieden werden, sofern jedes Objekt nur genau eine dieser Eigenschaften hat. Die Grundgesamtheit wird dadurch in mehrere disjunkte Teilmengen aufgeteilt.

(3.3.-01) ***Hinweis***

Den Inhalt dieses Kapitels findet man in den Büchern normalerweise in den Kapiteln über Wahrscheinlichkeiten. Tatsächlich haben die Fragestellungen und ihre Lösungen rechnerisch nichts mit Wahrscheinlichkeiten zu tun, sondern sind einfach relative Häufigkeiten. Ähnliches gilt auch bei manchen anderen Fragestellungen, wie z. B. bei der klassischen Wahrscheinlichkeit nach Laplace, die rechnerisch einfach nur eine relative Häufigkeit ist (siehe Kapitel 6.3).

(3.3.-02) ***Problemstellung (Ansatz)***

Aus einer Grundgesamtheit, in der einige Elemente die Eigenschaft E haben, wird eine bestimmte Anzahl von Elementen ausgewählt. Wie viele Möglichkeiten gibt es, dass eine vorgegebene Anzahl dieser ausgewählten Elemente die Eigenschaft E hat?

(3.3.-03) ***Definition***

Um diese Frage konkret beantworten zu können, wird definiert:

G   Grundgesamtheit
E   eine Eigenschaft von gewissen Elementen der Grundgesamtheit
N = Anzahl der Elemente von G, $N = |G|$
M = Anzahl der Elemente von G mit der Eigenschaft E $(0 \leq M \leq N)$

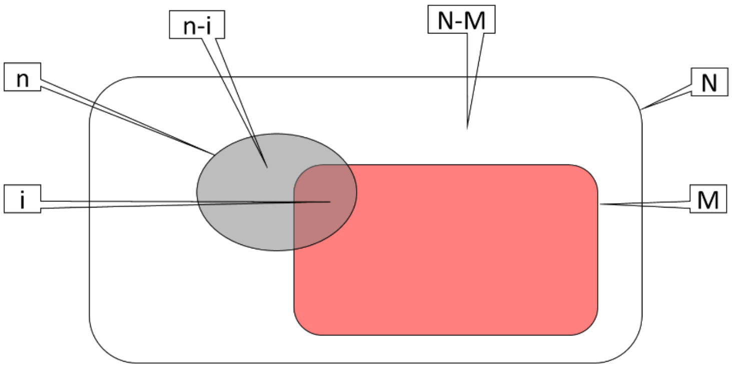
n = Anzahl der ausgewählten Elemente
i = Anzahl der ausgewählten Elemente mit der Eigenschaft E ($0 \leq i \leq n$)

Sind Wiederholungen bei der Auswahl erlaubt („Auswahl mit Zurücklegen“) so gilt:

- n kann beliebig groß werden, insbesondere kann n > N, i > M oder i > N sein.

Sind Wiederholungen bei der Auswahl nicht erlaubt („Auswahl ohne Zurücklegen“), so gilt:

- $0 \leq n \leq N$
- $0 \leq i \leq M$ und $0 \leq n-i \leq N-M$



(3.3.-04) ***<u>Erweiterte Problemstellung</u>***

1. Wie groß ist die <u>Anzahl Möglichkeiten</u>, aus N Elementen, von denen M die Eigenschaft E haben, n Elemente so auszuwählen, dass i von ihnen die Eigenschaft E haben?

2. Darüber hinaus wird bei jedem Fall die <u>relative Häufigkeit</u> dieser Anzahl bezogen auf die Gesamtanzahl von Variationen/Kombinationen mit n Elementen ermittelt. Dazu wird definiert:
   S = Menge der Variationen/Kombinationen von n beliebigen Elementen aus der Grundgesamtheit G. S ist die neue Grundgesamtheit für die Berechnung der relativen Häufigkeit.
   e = $(e_1, \ldots, e_n) \in S$; $e_i \in G$ sind die Komponenten von e
   K(i) = Teilmenge von S, die alle Elemente e enthält, bei denen genau i Komponenten die Eigenschaft E haben.
   f(i) = $\frac{|K(i)|}{|S|}$ relative Häufigkeit
   Die Anzahl der Elemente von S erhält man aus Kapitel 3.2 oder indem man M = N (dann folgt i = n) oder M = 0 (dann folgt i = 0) in die Formel für die Anzahl der Elemente von K(i) einsetzt.

3. Ferner wird der <u>Grenzwert der relativen Häufigkeit</u> betrachtet, wobei der relative Anteil der Elemente mit der Eigenschaft E als (annähernd) konstant vorausgesetzt wird. Genauer:
   Unter den Voraussetzungen:
   - $\lim_{N \to \infty} \frac{M}{N} = p$, d. h., M entwickelt sich proportional zu N, also z. B.: $M = [p \cdot N]$
   - n und i bleiben fest, variieren also nicht mit N

   wird $\lim_{N \to \infty} f(i)$ berechnet. In allen 4 Fällen kommt derselbe Ausdruck als Grenzwert heraus, d. h., unter den genannten Voraussetzungen gilt: je größer N wird, umso geringer sind die Unterschiede der 4 Fälle.

4. Ferner wird in allen Fällen folgendes <u>Beispiel</u> betrachtet:

In einer Urne seien 7 Kugeln, davon genau 3 mit der Eigenschaft „weiß“. Wie viele Möglichkeiten gibt es, 3 Kugeln aus der Grundgesamtheit auszuwählen, so dass genau 2 weiße Kugeln dabei sind?

$N = 7$	Anzahl der Kugeln der Grundgesamtheit
$M = 3$	Anzahl der weißen Kugeln in der Grundgesamtheit
$n = 3$	Anzahl der ausgewählten Kugeln
$i = 2$	Anzahl der ausgewählten weißen Kugeln

w_1, w_2, w_3 sind die 3 weißen Kugeln und n_1, n_2, n_3, n_4 die 4 nicht-weißen Kugeln.

(3.3.-05) ***Fall 1:*** Variation (= Reihenfolge wird berücksichtigt) ohne Wiederholung

Ergebnisse:

- Anzahl Möglichkeiten: $\binom{M}{i} \cdot \binom{N-M}{n-i} \cdot n!$
- Relative Häufigkeit: $f_1(i) = \frac{|K_1|}{|S_1|} = \frac{\binom{M}{i} \cdot \binom{N-M}{n-i}}{\binom{N}{n}}$
- $\lim_{N \to \infty} f_1(i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$; mit $\lim_{N \to \infty} \frac{M}{N} = p$ und n, i konstant

Herleitung:

Die ausgewählten Elemente werden in einen Vektor der Länge n sortiert: $(e_1, e_2, \ldots, e_n)$

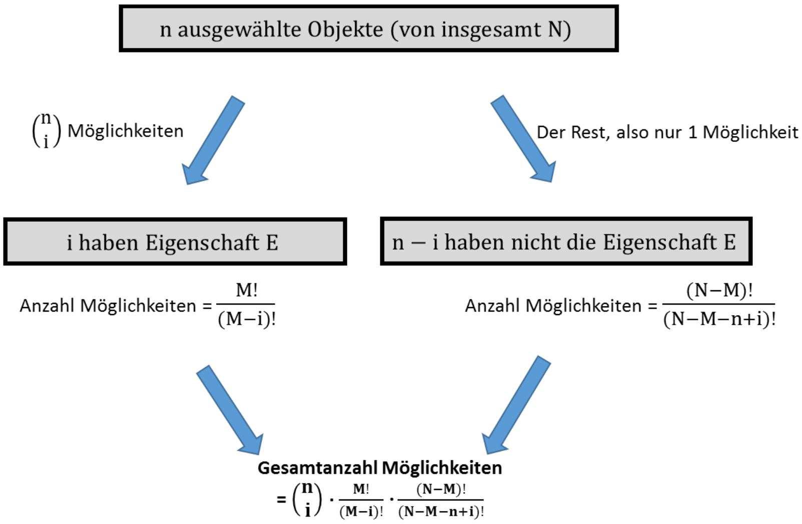
a. Es gibt dann $\binom{n}{i}$ mögliche Positionen für die i Elemente, die die Eigenschaft E haben.

b. Für die i Elemente mit der Eigenschaft E gibt es insgesamt $M \cdot (M-1) \cdot \ldots \cdot (M-i+1) = \frac{M!}{(M-i)!}$ Möglichkeiten der Auswahl

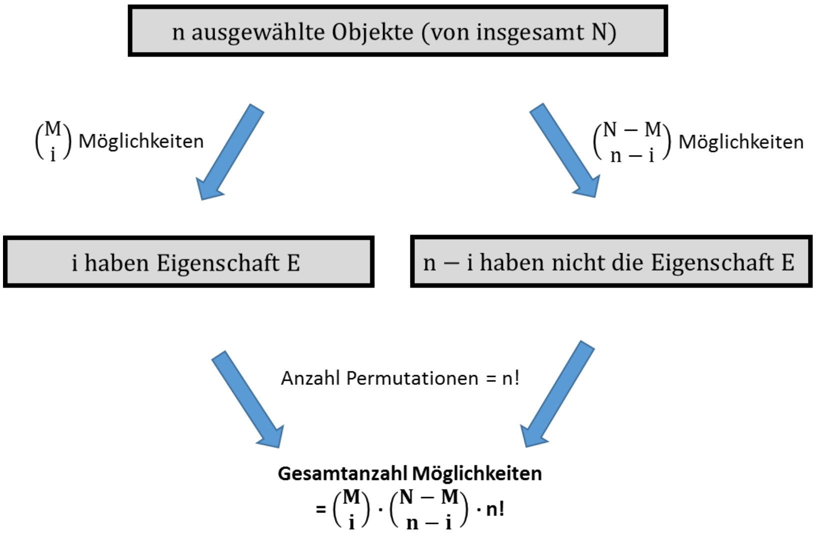
c. Für die verbleibenden n-i Elemente gibt es insgesamt $(N-M) \cdot (N-M-1) \cdot \ldots \cdot (N-M-n+i+1) = \frac{(N-M)!}{(N-M-n+i)!}$ Möglichkeiten, Elemente auszuwählen, die nicht die Eigenschaft E haben.

Insgesamt sind das $\binom{n}{i} \cdot \frac{M!}{(M-i)!} \cdot \frac{(N-M)!}{(N-M-n+i)!} = \binom{M}{i} \cdot \binom{N-M}{n-i} \cdot n!$ Möglichkeiten. Den zweiten Ausdruck erhält man durch Umsortierung der Fakultäten.

Grafische Darstellung:



oder



<u>Relative Häufigkeit:</u>

$K_1(i)$ = Menge der Variationen von n Elementen, von denen i die Eigenschaft E haben

S_1 = Menge der Variationen von beliebigen n Elementen

$$|S_1| = \binom{0}{0}\cdot\binom{N-0}{n-0}\cdot n! = \binom{N}{n}\cdot n!$$

$$f_1(i) = \frac{|K_1(i)|}{|S_1|} = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}\cdot n!}{\binom{N}{n}\cdot n!} = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}}$$

<u>Grenzwert der relativen Häufigkeit:</u>

In den folgenden Umformungen wird ausgenutzt, dass für jedes konstante a gilt:

$$\lim_{N\to\infty} \frac{N-a}{N} = 1.$$

Durch n-malige Anwendung folgt daraus dann z. B. (n konstant):

$$\lim_{N\to\infty} \frac{(N-n)!\cdot N^n}{N!} = \lim_{N\to\infty} \frac{N^n}{\frac{N!}{(N-n)!}} = \lim_{N\to\infty} \left(\frac{N}{N}\cdot\frac{N}{N-1}\cdot \ldots \cdot\frac{N}{N-n+1}\right) = 1$$

Damit ergibt sich:

$$\begin{aligned}
\lim_{N\to\infty} f_1(i) &= \lim_{N\to\infty} \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}} = \lim_{N\to\infty} \frac{M!\cdot(N-M)!\cdot n!\cdot(N-n)!}{i!\cdot(M-i)!\cdot(N-M-n+i)!\cdot(n-i)!\cdot N!} \\
&= \frac{n!}{i!\cdot(n-i)!}\cdot\lim_{N\to\infty}\left(\frac{M!}{(M-i)!}\cdot\frac{(N-M)!}{(N-M-n+i)!}\cdot\frac{(N-n)!}{N!}\right) \\
&= \binom{n}{i}\cdot\lim_{N\to\infty}\left(\frac{M!}{(M-i)!\cdot N^i}\cdot\frac{(N-M)!}{(N-M-n+i)!\cdot N^{n-i}}\cdot\frac{(N-n)!\cdot N^n}{N!}\right) \\
&= \binom{n}{i}\cdot\lim_{N\to\infty}\left(\frac{M!}{(M-i)!\cdot N^i}\right)\cdot\lim_{N\to\infty}\left(\frac{(N-M)!}{(N-M-n+i)!\cdot N^{n-i}}\right)\cdot\lim_{N\to\infty}\left(\frac{(N-n)!\cdot N^n}{N!}\right) \\
&= \binom{n}{i}\cdot\lim_{N\to\infty}\left(\frac{M^i}{N^i}\right)\cdot\lim_{N\to\infty}\left(\frac{(N-M)^{n-i}}{N^{n-i}}\right)\cdot\lim_{N\to\infty}\left(\frac{N^n}{N^n}\right) \\
&= \binom{n}{i}\cdot\left(\lim_{N\to\infty}\left(\frac{M}{N}\right)\right)^i\cdot\left(\lim_{N\to\infty}\left(\frac{N-M}{N}\right)\right)^{n-i}\cdot 1 \\
&= \binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}
\end{aligned}$$

<u>Beispiel:</u>

Die gesuchten Fälle sind für j = 1, 2, 3, 4:

(w_1, w_2, n_j)	(w_1, n_j, w_2)	(n_j, w_1, w_2)
(w_1, w_3, n_j)	(w_1, n_j, w_3)	(n_j, w_1, w_3)

(w_2, w_1, n_j) (w_2, n_j, w_1) (n_j, w_2, w_1)
(w_2, w_3, n_j) (w_2, n_j, w_3) (n_j, w_2, w_3)
(w_3, w_1, n_j) (w_3, n_j, w_1) (n_j, w_3, w_1)
(w_3, w_2, n_j) (w_3, n_j, w_2) (n_j, w_3, w_2)

Diese explizite Aufstellung und die Formel ergeben denselben Wert:

$$18 \cdot 4 = 72 = \binom{3}{2} \cdot \binom{7-3}{3-2} \cdot 3!$$

Die relative Häufigkeit ist:

$$f_1(2) = \frac{\binom{3}{2} \cdot \binom{7-3}{3-2}}{\binom{7}{3}} = \frac{12}{35} \approx 34{,}3\ \%.$$

Betrachtet man also unter den genannten Voraussetzungen alle Möglichkeiten, aus den 7 Kugeln 3 auszuwählen, so haben ca. 34,3 % dieser Möglichkeiten die Eigenschaft, dass genau 2 weiße Kugeln dabei sind.

(3.3.-06) ***Fall 2:*** Variation (= Reihenfolge wird berücksichtigt) mit Wiederholung

Ergebnisse:

- Anzahl Möglichkeiten: $\binom{n}{i} \cdot M^i \cdot (N-M)^{n-i}$
- Relative Häufigkeit: $f_2(i) = \frac{|K_2(i)|}{|S_2|} = \binom{n}{i} \cdot \left(\frac{M}{N}\right)^i \cdot \left(1 - \frac{M}{N}\right)^{n-i}$
- $\lim\limits_{N \to \infty} f_2(i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$; mit $\lim\limits_{N \to \infty} \frac{M}{N} = p$ und n, i konstant

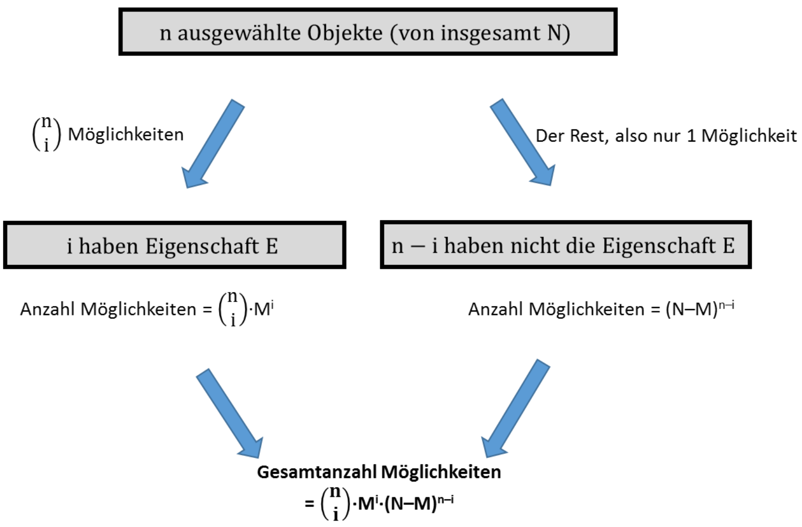
Herleitung:

Die ausgewählten Elemente werden in einen Vektor der Länge n sortiert: $(e_1, e_2, \ldots, e_n)$

a. Es gibt dann $\binom{n}{i}$ mögliche Positionen für die i Elemente, die die Eigenschaft E haben.
b. Bei jedem der i Elemente gibt es M Möglichkeiten, ein Element mit der Eigenschaft E zu nehmen, also M^i Möglichkeiten
c. Bei jedem der verbleibenden n-i Elemente gibt es N-M Möglichkeiten, ein Element zu nehmen, das nicht die Eigenschaft E hat, also $(M-N)^{n-i}$ Möglichkeiten

Insgesamt sind das $\binom{n}{i} \cdot M^i \cdot (N-M)^{n-i}$ Möglichkeiten.

Grafische Darstellung:



Relative Häufigkeit:

$K_2(i)$ = Menge der Variationen von n Elementen, von denen i die Eigenschaft E haben

S_2 = Menge der Variationen von beliebigen n Elementen

$|S_2| = \binom{n}{0}\cdot 0^0\cdot(N-0)^{n-0} = N^n$, wobei $0^0 = 1$ gesetzt wurde.

Relative Häufigkeit: $f_2(i) = \frac{|K_2(i)|}{|S_2|} = \frac{\binom{n}{i}\cdot M^i\cdot(N-M)^{n-i}}{N^n} = \binom{n}{i}\cdot\left(\frac{M}{N}\right)^i\cdot\left(1-\frac{M}{N}\right)^{n-i}$

Grenzwert der relativen Häufigkeit:

$$\lim_{N\to\infty} f_2(i) = \lim_{N\to\infty}\left(\binom{n}{i}\cdot\left(\frac{M}{N}\right)^i\cdot\left(1-\frac{M}{N}\right)^{n-i}\right) = \binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}$$

Beispiel:

Die gesuchten Fälle sind für j = 1, 2, 3, 4:

(w_1, w_1, n_j)	(w_1, n_j, w_1)	(n_j, w_1, w_1)
(w_1, w_2, n_j)	(w_1, n_j, w_2)	(n_j, w_1, w_2)
(w_1, w_3, n_j)	(w_1, n_j, w_3)	(n_j, w_1, w_3)
(w_2, w_1, n_j)	(w_2, n_j, w_1)	(n_j, w_2, w_1)
(w_2, w_2, n_j)	(w_2, n_j, w_2)	(n_j, w_2, w_2)
(w_2, w_3, n_j)	(w_2, n_j, w_3)	(n_j, w_2, w_3)
(w_3, w_1, n_j)	(w_3, n_j, w_1)	(n_j, w_3, w_1)
(w_3, w_2, n_j)	(w_3, n_j, w_2)	(n_j, w_3, w_2)
(w_3, w_3, n_j)	(w_3, n_j, w_3)	(n_j, w_3, w_3)

Diese explizite Aufstellung und die Formel ergeben denselben Wert:

$$27\cdot 4 = 108 = \binom{3}{2}\cdot 3^2\cdot(7-3)^{3-2}$$

Die relative Häufigkeit ist:

$$f_2(2) = \binom{3}{2}\cdot\left(\frac{3}{7}\right)^2\cdot\left(1-\frac{3}{7}\right)^{3-2} = \frac{\binom{3}{2}\cdot 3^2\cdot(7-3)^{3-2}}{7^3} = \frac{3\cdot 9\cdot 4}{7^3} = \frac{108}{343} \approx 31{,}5\ \%$$

Betrachtet man also unter den genannten Voraussetzungen alle Möglichkeiten, aus den 7 Kugeln 3 auszuwählen, so haben ca. 31,5 % dieser Möglichkeiten die Eigenschaft, dass genau 2 weiße Kugeln dabei sind.

(3.3.-07) ***Fall 3:*** Kombination (= Reihenfolge wird nicht berücksichtigt) ohne Wiederholung

Ergebnisse:

- Anzahl Möglichkeiten: $\binom{M}{i}\cdot\binom{N-M}{n-i}$
- Relative Häufigkeit: $f_3(i) = \frac{|K_3(i)|}{|S_3|} = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}}$
- $\lim_{N\to\infty} f_3(i) = \binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}$; mit $\lim_{N\to\infty}\frac{M}{N} = p$ und n, i konstant

Herleitung:

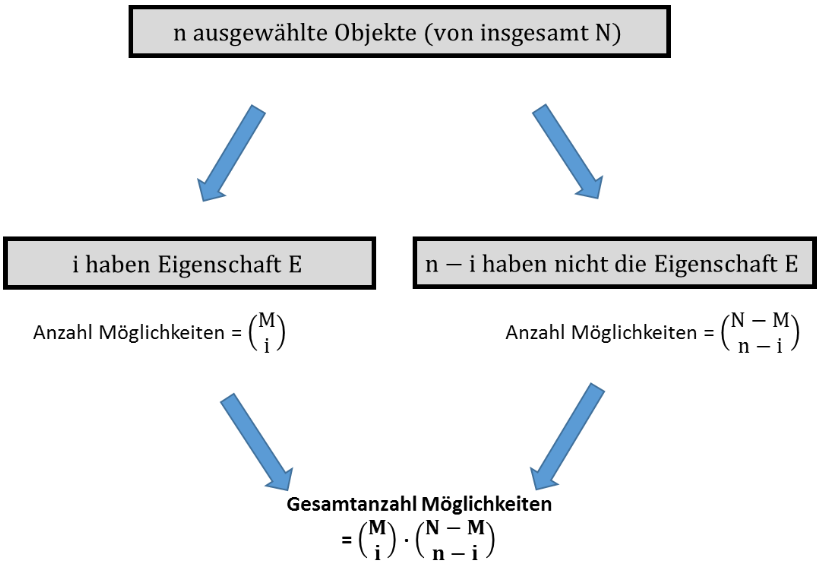
(1) Mit Fall 1:

Da die Elemente alle unterschiedlich sind, gibt es n! Möglichkeiten, die n ausgewählten Elemente anzuordnen. Also muss nur die Formel aus Fall 1 durch n! geteilt werden.

(2) Ohne Fall 1:

Es gibt $\binom{M}{i}$ Möglichkeiten, aus den M Elementen mit der Eigenschaft E genau i Elemente auszuwählen. Es gibt N-M Elemente, die nicht die Eigenschaft E haben. Hat man schon i Elemente mit der Eigenschaft E ausgewählt, so kann man noch (n-i) Elemente aus den N-M Elementen, die nicht die Eigenschaft E haben, auswählen, das sind $\binom{N-M}{n-i}$ Möglichkeiten.
Insgesamt gibt es also $\binom{M}{i} \cdot \binom{N-M}{n-i}$ Möglichkeiten genau n Elemente auszuwählen, von denen genau i die Eigenschaft E haben (und damit n-i nicht die Eigenschaft E haben).

<u>Grafische Darstellung:</u>



<u>Relative Häufigkeit:</u>

$K_3(i)$ = Menge der Kombinationen von n Elementen, von denen i die Eigenschaft E haben
S_3 = Menge der Kombinationen von beliebigen n Elementen
$|S_3| = \binom{0}{0} \cdot \binom{N-0}{n-0} = \binom{N}{n}$

Relative Häufigkeit: $f_3(i) = \frac{|K_3(i)|}{|S_3|} = \frac{\binom{M}{i} \cdot \binom{N-M}{n-i}}{\binom{N}{n}} = f_1(i)$

<u>Grenzwert der relativen Häufigkeit:</u>

$$\lim_{N \to \infty} f_3(i) = \lim_{N \to \infty} f_1(i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

<u>Beispiel:</u>

Die gesuchten Fälle sind für j = 1, 2, 3, 4:

(w_1, w_2, n_j) (w_1, w_3, n_j) (w_2, w_3, n_j)

Diese explizite Aufstellung und die Formel ergeben denselben Wert:

$$3 \cdot 4 = 12 = \binom{3}{2} \cdot \binom{7-3}{3-2}$$

Die relative Häufigkeit ist:

$$f_3(2) = \frac{\binom{3}{2} \cdot \binom{7-3}{3-2}}{\binom{7}{3}} = \frac{12}{35} \approx 34{,}3\ \%.$$

Betrachtet man also unter den genannten Voraussetzungen alle Möglichkeiten, aus den 7 Kugeln 3 auszuwählen, so haben ca. 34,3 % dieser Möglichkeiten die Eigenschaft, dass genau 2 weiße Kugeln dabei sind.

(3.3.-08) ***Fall 4:*** Kombination (= Reihenfolge wird nicht berücksichtigt) mit Wiederholung

Ergebnisse:

- Anzahl Möglichkeiten: $\binom{M+i-1}{i} \cdot \binom{N-M+n-i-1}{n-i}$
- Relative Häufigkeit: $f_4(i) = \frac{|K_4(i)|}{|S_4|} = \frac{\binom{M+i-1}{i} \cdot \binom{N-M+n-i-1}{n-i}}{\binom{N+n-1}{n}}$
- $\lim\limits_{N \to \infty} f_4(i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$; mit $\lim\limits_{N \to \infty} \frac{M}{N} = p$ und n, i konstant

Herleitung:

Die N Elemente der Grundgesamtheit G seien folgendermaßen durchnummeriert:

$e_1, \ldots, e_M$ haben die Eigenschaft E

$e_{M+1}, \ldots, e_N$ haben nicht die Eigenschaft E

Daraus werden n Elemente mit Wiederholung ausgewählt

Im folgenden Beispiel ist N=7, M=4, n=8 und i=4.

Gegeben sei eine Treppe mit N Stufen, wobei Stufe 1 ebenerdig ist, also in der folgenden Tabelle nicht aufgeführt ist. Für jedes ausgewählte Objekt e_j (j = 1, …, n) wird eine Kugel auf die j. Treppenstufe gelegt. Ferner gibt es 2 Arten von Anstrengungen:

B = „Bücken und eine Kugel aufheben“

S = „Eine Treppenstufe hochsteigen“. Dabei werden die ersten M-1 Aktivitäten mit S1 und die restlichen N-M Aktivitäten mit S2 bezeichnet.

Die Auswahl der Elemente führt dann z. B. zu folgender Tabelle

1	**2**	**3**	**4**	**5**	**6**	**7**
	S1	S1	S1	S2	S2	S2
B		B	B		B	B
		B			B	
		B				

Diese Tabelle wird linearisiert, indem man die Inhalte der Spalten ohne die Spaltennummer hintereinander schreibt und berücksichtigt, dass durch ein „S“ eine neue Spalte beginnt. Dann ergibt sich

B, S1, S1, B, B, B, S1, B – S2, S2, B, B, S2, B

Das Problem lautet dann anders formuliert: wie viele Anstrengungen gibt es?

- Im 1. Abschnitt (betrifft die Elemente mit der Eigenschaft E) gibt es (M–1)+i Anstrengungen,
- im 2. Abschnitt (betrifft die Elemente ohne die Eigenschaft E) gibt es (N–M)+(n–i) Anstrengungen.

Damit ergibt sich:

- Im 1. Abschnitt gibt es i Kugeln und M–1 Stufen, also $\binom{M+i-1}{i}$ Möglichkeiten der Anordnung ohne Berücksichtigung der Reihenfolge.
- Der 2. Abschnitt beginnt immer mit S2, an dieser Stelle kann in der linearisierten Liste kein B stehen, so dass nur (N–M)+(n–i)–1 Stellen für die Kugeln ohne die Eigenschaft E übrigbleiben. Dafür gibt es dann $\binom{N-M+n-i-1}{n-i}$ Möglichkeiten der Anordnung ohne Berücksichtigung der Reihenfolge.

Da die beiden Anordnungen unabhängig voneinander sind, gibt es insgesamt

$$\binom{M+i-1}{i}\cdot\binom{N-M+n-i-1}{n-i}$$

Möglichkeiten der Anordnung.

Relative Häufigkeit:

$K_4(i)$ = Menge der Kombinationen von n Elementen, von denen i die Eigenschaft E haben

S_4 = Menge der Kombinationen von beliebigen n Elementen

$$|S_4| = \binom{0+0-1}{0}\cdot\binom{N-0+n-0-1}{n-0} = \binom{N+n-1}{n}$$

Relative Häufigkeit: $f_4(i) = \frac{|K_4(i)|}{|S_4|} = \frac{\binom{M+i-1}{i}\cdot\binom{N-M+n-i-1}{n-i}}{\binom{N+n-1}{n}}$

Grenzwert der relativen Häufigkeit:

$$\lim_{N\to\infty} f_4(i) = \lim_{N\to\infty} \frac{\binom{M+i-1}{i}\cdot\binom{N-M+n-i-1}{n-i}}{\binom{N+n-1}{n}} = \lim_{N\to\infty} \frac{(M+i-1)!\cdot(N-M+n-i-1)!\cdot n!\cdot(N-1)!}{i!\cdot(M-1)!\cdot(N-M-1)!\cdot(n-i)!\cdot(N+n-1)!}$$

$$= \frac{n!}{i!\cdot(n-i)!}\cdot\lim_{N\to\infty}\left(\frac{(M+i-1)!}{(M-1)!}\cdot\frac{(N-M+n-i-1)!}{(N-M-1)!}\cdot\frac{(N-1)!}{(N+n-1)!}\right) = \binom{n}{i}\cdot\lim_{N\to\infty}\left(M^i\cdot(N-M)^{n-i}\cdot\frac{1}{N^n}\right)$$

$$= \binom{n}{i}\cdot\lim_{N\to\infty}\left(\left(\frac{M}{N}\right)^i\cdot\left(1-\frac{M}{N}\right)^{n-i}\right) = \binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}$$

Beispiel:

Die gesuchten Fälle sind für j = 1, 2, 3, 4:

(w_1, w_1, n_j) (w_1, w_2, n_j) (w_1, w_3, n_j)

(w_2, w_2, n_j) (w_2, w_3, n_j)

(w_3, w_3, n_j)

Diese explizite Aufstellung und die Formel ergeben denselben Wert:

$$6\cdot 4 = 24 = \binom{3+2-1}{2}\cdot\binom{7-3+3-2-1}{3-2}$$

Die relative Häufigkeit ist:

$$f_4(2) = \frac{\binom{3+2-1}{2}\cdot\binom{7-3+3-2-1}{2-i}}{\binom{7+3-1}{3}} = \frac{24}{84} = \frac{2}{7} \approx 28{,}6\ \%.$$

Betrachtet man also unter den genannten Voraussetzungen alle Möglichkeiten, aus den 7 Kugeln 3 auszuwählen, so haben ca. 28,6 % dieser Möglichkeiten die Eigenschaft, dass genau 2 weiße Kugeln dabei sind.

4 Bewertungen

Zusammenfassung
Fast täglich bewerten wir etwas oder nehmen Bewertungen anderer wahr. Diese Bewertungen werden mit Worten wie „gut“ oder „schlecht“ oder auch mit Noten oder Punkten ausgedrückt – wie z. B. bei Tests der Stiftung Warentest. Damit sollten möglichst gut und für andere nachvollziehbar Rangordnungen oder Entwicklungen beschrieben werden. Da auch bei Wahrscheinlichkeiten häufig eine Zahl genannt wird, lohnt es sich, zunächst einen näheren Blick auf allgemeine Bewertungen zu werfen und ihre Eigenschaften kennenzulernen. Die Wahrscheinlichkeit erscheint dann nur noch als ein Beispiel einer bestimmten Bewertung (siehe Kapitel 5).
* * *

Dazu muss zunächst der Begriff „Bewertung“ definiert werden.

Um etwas zu bewerten kann man bestimmte Worte, Zeichenkombinationen oder Zahlen verwenden. Bewertende Worte sind z. B. „sehr gut“ oder „nicht empfehlenswert“. Zeichenkombinationen wie „AAA“ (triple A) oder A^+ werden zum Beispiel beim Rating von Staaten oder Unternehmen verwendet, um ihre Kreditwürdigkeit auszudrücken. Oder man verwendet Zahlen, wie Noten oder Punkte, um eine Bewertung auszudrücken.

<u>Definition</u>
Bei einer **Bewertung** werden bestimmten Objekten bezüglich einer Eigenschaft bestimmte Worte, eine Zeichenkombination oder eine Zahl zugeordnet. Damit können mehrere Objekte bezüglich dieser Eigenschaft verglichen werden. Eine Bewertung ist komparativ oder quantitativ.
* * *

Im Folgenden werden nur quantitative Bewertungen betrachtet.

Misst man „etwas“, so ordnet man diesem „etwas“ nach einem bestimmten Verfahren eine Zahl zu, man führt also eine quantitative Bewertung durch. Bei vielen Messungen gibt es eindeutige allgemein anerkannte Verfahren, zum Beispiel

- beim Ermitteln von physikalischen Größen wie z. B. Länge, Zeit, Gewicht
- bei der Bestimmung von kaufmännischen Größen wie z. B. Kontostand oder Zinsen
- bei der Bestimmung einer absoluten oder relativen Häufigkeit, wie z. B. der Anzahl der Einwohner in einer Stadt oder des Anteils der Kinder in einer Bevölkerungsgruppe

Einige dieser Verfahren werden durch Gesetze oder Normen eindeutig festgelegt.
Das Ergebnis ist grundsätzlich eindeutig, exakt und objektiv – bis auf eventuelle kleine Ungenauigkeiten bei Messungen physikalischer Größen. Die Messergebnisse können auf einer metrischen Skala dargestellt werden.

Manchmal kann es aber sinnvoll sein, auch dann etwas zahlenmäßig zu bewerten, wenn kein eindeutiges und allgemein akzeptiertes Messverfahren besteht, damit man …

- … verschiedene Objekte bezüglich bestimmter Eigenschaften vergleichen und eine Rangordnung erstellen
- … die bisherige und künftige Entwicklung eines bestimmten Merkmals detailliert darstellen
- … einen Durchschnittswert über mehrere Objekte ermitteln

kann. Natürlich sind solche Verfahren mit großen Vorsicht und einer gehörigen Portion Zweifel anzuwenden, denn sie suggerieren eine Genauigkeit und Objektivität, die sie tatsächlich nicht haben.

Die Bewertungen können dann Grundlage für Entscheidungen sein, sofern die einzelnen Objekte nach einem einheitlichen Verfahren bewertet wurden. Typische Maßeinheiten solcher Bewertungen sind „Punkte“, „Note“, „Prozent“ oder „Geldeinheit (€)“. Man nennt solche Messgrößen auch Indices, Indikatoren oder Kennzahlen.

Weitere Details und Anforderungen an Messverfahren werden im Rahmen der Messtheorie behandelt, aber im vorliegenden Buch nicht weiter betrachtet.

4.1 Beispiele

Die folgenden Beispiele demonstrieren, wie vielfältig Bewertungen sein können. Tatsächlich gibt es aber eine weitaus größere Anzahl und Vielfalt von Bewertungen. In alle Bewertungen fließen sowohl exakt messbare oder allgemein anerkannte, als auch eher vage oder subjektive Faktoren ein.

(4.1.-01) ***Volkswirtschaftliche Indices***

- Volkswirtschaftliche Indices dienen dazu, die zeitliche Entwicklung von volkswirtschaftlichen Größen darzustellen, wie z. B. die Mengen- und Preisindices nach Paasche oder Laspayres. Am bekanntesten ist der Verbraucherpreisindex, der vom statistischen Bundesamt festgelegt wird und umgangssprachlich als Inflationsrate bezeichnet wird. Ein praktisches Anwendungsbeispiel für diesen Index sind Preisanpassungsklauseln bei Gewerberaummieten. Diese Indices sind dimensionslose Größen.

- Aktienindices stellen den aktuellen Status einer bestimmten Gruppe von Aktien dar, wie z. B. der Deutsche Aktienindex DAX oder der Dow-Jones-Index. Der DAX ist eine dimensionslose Größe.

(4.1.-02) ***Bewertung einer Leistung***

- Schulnote
  Bei der Ermittlung einer Schulnote im Jahreszeugnis fließen sowohl objektive Kriterien (die z. B. durch Richtlinien festgelegt werden), als auch subjektive Kriterien (z. B. Schreibstil, Verhalten, Sympathie, Kreativität, …) ein. Die genaue Abgrenzung von subjektiven und objektiven Kriterien kann im Einzelfall schwierig sein. Die Bewertung kann durch Punkte oder Noten erfolgen.

- Bewertung von Mitarbeitern bezüglich Zielerreichung
  Bei betrieblichen Zielvereinbarungen werden häufig sowohl exakt messbare Ziele (Umsatzziele, Fehlerquote, Einhalten definierter Kosten, …), also auch nur grob messbare Ziele (Kooperationsbereitschaft, Motivation, …) definiert. Diese Ziele werden dann am Jahresende im Rahmen eines

Mitarbeitergespräches einzeln mit Zahlen bewertet und daraus ein gewichteter Durchschnitt gebildet. Dem komplexen Verhalten eines Mitarbeiters im Laufe eines Jahres wird also mit diesem Bewertungsverfahren eine Zahl zugeordnet, aus der sich Konsequenzen wie z. B. die Höhe einer Bonuszahlung oder Karrierechancen ergeben. Die Bewertung erfolgt z. B. durch eine Prozentzahl, die angibt, zu wieviel Prozent die geplante Leistung erreicht wurde.

- Bewertung von sportlichen Leistungen, z. B. Bewertung eines Skisprunges
  Beim Skisprung wird eine Gesamtpunktzahl aufgrund der erzielten Weite (exakt messbar), der Haltung (nicht exakt messbar) und ggf. weiterer Faktoren vergeben, indem ein gewichtetes arithmetisches Mittel der einzelnen Bewertungen ermittelt wird.
  Auch hier wird einer komplexen Handlung eine einzige Zahl zugeordnet, die dann über die Platzierung und die Belohnung entscheidet.

(4.1.-03) ***<u>Bewertung einer Ware oder Dienstleistung</u> <u>bezüglich des materiellen Wertes</u>***
Ermittelt ein Verkäufer einen Angebotspreis, so fließen dabei objektive Kriterien (z. B. Kosten) und subjektive Kriterien (z. B. vermutete Verkaufschancen, emotionale Faktoren) ein. Verschiedene Verkäufer können für dieselbe Ware oder Dienstleistung unterschiedliche Angebotspreise ermitteln. Und diese können sich bei neuen Rahmenbedingungen ändern.
Aus Käufersicht fließen beim Nachfragepreis (= Preis, den man zu zahlen bereit ist) sowohl objektive Kriterien (z. B. materieller Nutzen), als auch subjektiven Kriterien (z. B. ein gewisses Glücksgefühl) ein. Unterschiedliche Käufer können bei derselben Ware oder Dienstleistung zu unterschiedlichen Nachfragepreisen kommen, d. h., sie akzeptieren unterschiedliche Preise.

(4.1.-04) ***<u>Bewertung einer Ware oder Dienstleistung</u> <u>bezüglich der Qualität</u>***
- Die Stiftung Warentest führt regelmäßig Tests zu Waren oder Dienstleistungen durch und veröffentlicht die Ergebnisse in der Monatszeitschrift „test“. Beispielsweise wurden im Heft 05/2016 Tests zu Smartphones, Mozzarella und Deos veröffentlicht. Das Ergebnis eines Tests ist ein Qualitätsurteil in Form einer Schulnote mit einer Stelle hinter dem Komma. Dabei werden manche Faktoren exakt gemessen, während andere das arithmetische Mittel der Bewertungen mehrerer Personen sind. Die Bewertungen der einzelnen Faktoren werden auf eine einheitliche Dimension umgerechnet und dann ein Durchschnittswert als Gesamtbewertung ermittelt. Mit dieser Methode wurden beim Mozzarella-Test Noten zwischen 1,7 und 4,0 vergeben.

- Test in Fachzeitschriften
  Auch in vielen Fachzeitschriften werden Tests veröffentlicht, wie z. B. Tests von Autos und Autozubehör, von Kameras oder von Smartphones. Bei diesen Tests wird das Ergebnis oft in Punkten oder Noten dargestellt und durch Worte wie „empfehlenswert“ ergänzt. Die Gesamtpunktzahl bzw. die Gesamtnote ist dabei in der Regel eine Mischung aus objektiven (physikalisch gemessenen) und subjektiven (von Personen beurteilten) Kriterien. Auch hier werden die Einzelbewertungen auf eine einheitliche Dimension umgerechnet und dann zu einer Note oder Punktzahl zusammengefasst.

(4.1.-05) ***<u>Bewertung eines Landes bezüglich der Friedlichkeit</u>***
Der „Global Peace Index“ wird jährlich vom „Institute for Economics and Peace“ ermittelt. Dieser Index ist eine Zahl zwischen 1 und 5 mit drei Stellen hinter dem Komma und ist ein Maß für die Friedlichkeit eines Staates. Im Jahre 2015 lag Island auf dem 1. Platz mit einem Friedlichkeitsindex von

1,148, während Syrien auf dem 162. Platz mit einem Index von 3,645 war. Dabei hatte sich Islands Friedlichkeit von 2014 auf 2015 um 0,002 Punkte verbessert (siehe [VH]).

(4.1.-06) ***Der Intelligenzquotient (IQ) als Bewertung des intellektuellen Leistungsvermögens***
Es gibt je nach Fragestellung und je nachdem, was man unter Intelligenz versteht, unterschiedliche Verfahren und Ergebnisse. Will man Personen bezüglich ihrer Intelligenz vergleichen, so muss dasselbe Verfahren angewendet werden. Der Intelligenzquotient ist eine dimensionslose Zahl.

(4.1.-07) ***Bewertung des Geschäftsklimas***
Der ifo-Geschäftsklimaindex wird monatlich erstellt und in den Medien veröffentlicht (siehe [IF]). Dabei werden ca. 7.000 Unternehmen gebeten, ihre gegenwärtige Geschäftslage zu beurteilen und ihre Erwartungen für die nächsten sechs Monate mitzuteilen. Sie können ihre Lage mit "gut", "befriedigend" oder "schlecht" und ihre Geschäftserwartungen für die nächsten sechs Monate als "günstiger", "gleich bleibend" oder "ungünstiger" kennzeichnen. Aus diesen Einschätzungen der Unternehmen wird dann eine Zahl – der sogenannte ifo-Geschäftsklimaindex – errechnet.

(4.1.-08) ***Bewertung zweier Personen bezüglich Ihrer Übereinstimmung***
Bei der Bewertung zweier Personen durch eine Partnervermittlung spielen sowohl objektiven Kriterien (Einkommen, Schulbildung, …), als auch subjektive oder vage Kriterien (was gefällt (nicht) bei einem Partner, …) eine Rolle. Unterschiedliche Partnervermittlungen können bei denselben Personen zu unterschiedlichen Ergebnissen kommen. Die Übereinstimmung zweier Personen kann z. B. in Punkten oder Prozentzahlen angegeben werden. Siehe z. B. [PV].

(4.1.-09) ***Bewertung eines Landes bezüglich der Gleichstellung der Geschlechter***
Der “Global Gender Gap Index” ist ein vom World Economic Forum erstellter Index, der den Grad der Gleichstellung der Geschlechter quantifiziert (siehe [GG]). Auch hier wird ein Wert aus zahlreichen Einzelbewertungen ermittelt. Der Global Gender Gap Index liegt zwischen null (minimale Gleichstellung) und eins (maximale Gleichstellung).

(4.1.-10) ***Der Refugees-welcome-Index***
Dieser Index, der von Amnesty International erhoben wird, misst den Grad, mit dem Flüchtlinge in einem Land willkommen sind. Es wurden 27 Länder untersucht und das Ergebnis auf einer Skala von 0 bis 100 dargestellt (siehe [AI]).

(4.1.-11) ***Der OECD Better Life Index***
Die OECD (Organisation for Economic Co-operation and Development) misst mit diesem Index das gesellschaftliche Wohlergehen in verschiedenen Staaten (siehe [OE]). Der Index ist eine Zahl zwischen 0 (geringste Zufriedenheit) und 10 (höchste Zufriedenheit) und setzt sich aus 11 subjektiven und objektiven Einzelbewertungen zusammen.

(4.1.-12) ***World Happiness Index***
Für diesen Index wurden 156 Länder von einer Expertengruppe anhand diverser Kriterien untersucht, um das Glücksempfinden in dem jeweiligen Land zu quantifizieren. Der Index ist eine Zahl zwischen 0 und 10 und wird auf drei Stellen hinter dem Komma berechnet. Siehe [WH].

(4.1.-13) ***<u>Bewertung des Gesundheitszustandes</u>***
Es gibt eine Reihe von Apps, die den Gesundheitszustand einer Person mit einem Gesundheitsindex bewerten. Dieser Index kann z. B. eine Zahl zwischen 1 und 1000 sein (siehe z. B. [DA]).

(4.1.-14) ***<u>Bewertung der Kundenzufriedenheit</u>***
Es gibt viele Methoden zur Messung der Kundenzufriedenheit. Um die Entwicklung der Kundenzufriedenheit darzustellen, wird die (möglichst immer gleiche) Methode regelmäßig angewendet. Das Ergebnis wird dann z. B. zu einer Note oder Punktzahl verdichtet.

(4.1.-15) ***<u>Erlebnispunkte für Wanderwege</u>***
Das Deutsche Wanderinstitut vergibt Erlebnispunkte für Wanderwege auf Basis von 34 Einzelkriterien. Die Punktzahl liegt zwischen 0 und 100 Punkte. Siehe [WI].

(4.1.-16) ***<u>Angstindex</u>***
Die Ängste der Deutschen wurden bei der R+V Versicherung untersucht und die verschiedenen Angstarten auf einer Skala von 0% bis 100% dargestellt (siehe [RU]). Der Durchschnitt der 16 Angstarten ist der Angstindex.

4.2 Methode

Aufgrund der vielen Beispiele kann man salopp formulieren: **es gibt nichts, was nicht von jemandem gemessen wird**. Ob man diese Messungen und die daraus abgeleiteten Konsequenzen für sinnvoll hält, ist eine andere Frage. Dieser Drang zum Messen ergibt sich auch aus dem Satz „Was man nicht messen kann, kann man nicht lenken“, der dem Ökonom Peter Drucker zugeschrieben wird.

Die verschiedenen Messungen werden nach einer ähnlichen Methode vorgenommen, die man am einfachsten bei den Tests von Stiftung Warentest ablesen kann:

<u>Schritt 1:</u>
Man definiert möglichst genau,

- was man bewertet
- bezüglich welche Eigenschaft die Bewertung erfolgt
- ggf. warum man diese Bewertung durchführt, also welche Entscheidungen oder Maßnahmen daraus folgen.

<u>Schritt 2:</u>
Man definiert die relevanten Einflussfaktoren für die Bewertung. Diese Einflussfaktoren sollten sich nicht überschneiden; anderenfalls muss sichergestellt werden, dass die Überschneidung nur einmal berücksichtigt wird. Die Einflussfaktoren sind die Bezeichnungen der Zeilen in der Bewertungstabelle.

<u>Schritt 3:</u>

Man bewertet die einzelnen Einflussfaktoren mit spezifischen Verfahren. Diese Werte können unterschiedliche Dimensionen haben. Bewertungen mancher Einflussfaktoren erhält man aufgrund objektiver Messungen, andere sind mehr oder weniger subjektive Einschätzungen oder Durchschnittswerte von mehreren solcher Einschätzungen.

Schritt 4:
Aus den Ergebnissen von Schritt 3 wird dann die Gesamtbewertung ermittelt.
Bei einigen Verfahren, wie z. B. den Warentests, sieht Schritt 4 folgendermaßen aus:
Die Ergebnisse der einzelnen Messungen werden auf die Dimension der Gesamtbewertung umgerechnet (also z. B. auf „Note“ oder „Punkte“), gewichtet und daraus dann eine Summe gebildet.

Überschneidung der Einflussfaktoren:
Bei Schritt 2 wurde die Überschneidung von Einflussfaktoren erwähnt. Rechnet man die Überschneidung nicht heraus, so wird sie in der Gesamtbewertung doppelt berücksichtigt.

Beispiel:
Bei einem Auto werden die beiden zusätzlichen Ausstattungspakete „Premium“ und „Business“ angeboten. „Premium“ bestehe aus den drei Komponenten A, B und C, „Business“ aus den drei Komponenten C, D und E.

Fall 1:
Bewertung bezüglich des materiellen Wertes: Preis P in €
P(Premium) = 3.000 €
P(Business) = 2.500 €
Wenn man alle Ausstattungsmerkmale haben möchte, muss man beide Ausstattungspakete kaufen.
Wenn der Händler fair ist, muss man Merkmal C nicht doppelt bezahlen, also
P(Gesamtzusatzausstattung) = P(Premium) + P(Business) – P(C)

Fall 2:
Bewertung bezüglich der Qualität: Qualität Q in Punkten
Q(Premium) = 12 Punkte
Q(Business) = 10 Punkte
Wenn die Tester nicht aufpassen, ermitteln sie für die gesamte Zusatzausstattung 22 Punkte und zählen dabei die Qualität des Merkmals C doppelt. Richtig wäre wie beim Preis:
Q(Gesamtzusatzausstattung) = Q(Premium) + Q(Business) – Q(C)

4.3 Bewertungen von Kombinationen

In manchen Fällen kann es sinnvoll sein, die Bewertung von einer Kombination von gleichartigen Objekten zu ermitteln, in anderen Fällen bleibt unklar, was „Kombination“ überhaupt zu bedeuten hat. Der Begriff „Kombination“ wird also in ganz allgemeiner Weise benutzt, während in den Kapiteln 3.2 und 3.3 eine ganz bestimmte Zusammenstellung von Elementen einer Menge gemeint war.

In den folgenden Beispielen wird der Begriff Kombination näher erläutert.

(4.3.-01) ***Fall 1***: Die Kombination ist sinnvoll, die Bewertung ist subadditiv
Fasst man mehrere Staaten zusammen (z. B. zu „Europa“ oder „Entwicklungsländer“), so kann man bei manchen Bewertungen einfach die Bewertungen der einzelnen Länder addieren. Kennt man z. B. die Waldfläche oder die Anzahl Einwohner oder den Wasserverbrauch von jedem der n Staaten Europas und bezeichnet man diese Bewertung kurz als B, so ergibt sich:

B(Europa) = B(Staat 1) + … + B(Staat n)

Gibt es ein Gebiet, das von zwei Staaten verwaltet oder beansprucht wird und wird dieses Gebiet in den Zahlen beider Staaten berücksichtigt, so gilt wie oben bei der Sonderausstattung von Autos:

B(Staat 1 kombiniert mit Staat 2) = B(Staat 1) + B(Staat 2) – B(gemeinsames Gebiet)

Diese Formeln haben aber nur dann einen Sinn, wenn das Bewertungsverfahren in allen Staaten gleich ist.

Andere Beispiele sind:

- Bei der 4mal100-Meter-Staffel ergibt sich rechnerisch die Gesamtzeit aus den Einzelzeiten der 4 Läufer/innen minus der Überschneidungszeiten. In der Regel wird aber nur die Gesamtzeit ermittelt, denn nur sie ist entscheidend für die Platzierung.
- Dachpappe wird überlappend verlegt, die abgedeckte Fläche ist also die Summe der Flächen der einzelnen Dachpappenstücke minus die Flächen der Überlappungen.
- Untersucht man die Merkmale A und B in einer Grundgesamtheit, so gilt gemäß Kapitel 2.1.4 sowohl für die relative, also auch für die absolute Häufigkeit H:

  H(Kombination von A und B) = H(A) + H(B) – H(Überlappung von A und B)
- Um entscheiden zu können, ob die Fusion zweier Unternehmen sinnvoll ist, werden die künftigen Kosten K der Unternehmen in 3 Jahren prognostiziert. „Kombination“ bedeutet „Fusion“ und „Überlappung“ sind gleichartige Kosten, die man nach der Fusion aufgrund von Synergieeffekten dann hoffentlich nur noch einmal hat. Dann gilt (natürlich sehr vereinfacht nur unter Berücksichtigung der Kostenreduzierungen):

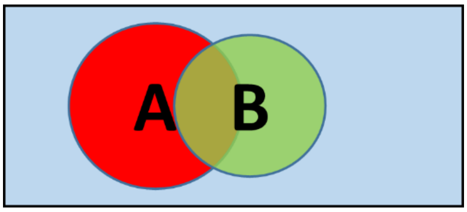
  K(A fusioniert mit B) = K(A) + K(B) – K(Überlappung von A und B)
- Die Schüler A und B werden in der Schule durch Punkte in den folgenden 4 Fächern bewertet:

Fach	**A**	**B**	**A und B**	**A oder B**
Mathematik	13	10	8	15
Sport	10	14	9	15
Englisch	8	12	8	12
Musik	14	10	9	15
Summe	**45**	**46**	**34**	**57**

  Unter „A und B“ stehen die Bewertungen der Fähigkeiten, die beide Schüler haben, also das, was beide können.
  Unter „A oder B“ steht die Bewertung der Fähigkeiten, wenn sie beide ihre Fähigkeiten zusammenlegen, also das, was nur A kann, plus das, was nur B kann, plus das, was beide können.
  Für jedes der 4 Fächer gilt:
  Punkte(A oder B) = Punkte(A) + Punkte(B) – Punkte(A und B)

  In der Summe gilt dann auch für die Gesamtbewertung:
  Punkte(A oder B) = Punkte(A) + Punkte(B) – Punkte(A und B)
  oder in Zahlen:
  57 = 45 + 46 – 34

Grafisch lassen sich solche Bewertungen durch Mengen-Diagramme symbolisieren.



Die Kreisflächen entsprechen den Bewertungen der Objekte A und B.

F(Kombination von A und B) = F(A) + F(B) – F(Überschneidung von A und B)

oder in Mengenschreibweise:

$F(A\cup B) = F(A) + F(B) - F(A\cap B)$.

Bewertungen, für die diese Formeln gelten, heißen **subadditiv**.

Gibt es keine Überschneidungen, so ist einfach

F(Kombination von A und B) = F(A) + F(B)

oder in Mengenschreibweise:

$F(A\cup B) = F(A) + F(B)$

(4.3.-02) ***Fall 2***: Die Kombination ist sinnvoll, die Bewertung ist nicht subadditiv
Fasst man mehrere Staaten zusammen, so kann man bei manchen Bewertungen die einzelnen Bewertungen nicht einfach addieren. Dies ist z. B. beim „World Happiness Index“ oder beim „Refugees Welcome Index“ der Fall, da man anderenfalls über den Maximalwert kommen würde. Stattdessen könnte man gewichtete arithmetische Mittel nehmen, wobei die Gewichtung so gewählt werden sollte, dass simuliert wird, dass es sich um einen Gesamtstaat an Stelle von mehreren Staaten handelt.
Ähnliche Probleme hat man auch bei anderen Indices wie den Mengen- und Preisindices nach Paasche oder Laspayres. Kombiniert man zwei überschneidungsfreie Warenkörbe, so ist die Bewertung des daraus gebildeten Gesamtwarenkorbes nicht die Summe der Einzelwerte.
Weitere Beispiele dafür, dass man die Einzelbewertungen von überschneidungsfreien Objekten nicht einfach addieren darf, da es gegenseitige Abhängigkeiten oder Beeinflussungen gibt, sind:

- Vermischt man Wasser und Alkohol, so hat das Gemisch aufgrund der Volumenkontraktion ein geringeres Volumen als die Summe der einzelnen Volumina.
- Zerteilt man einen Gegenstand G in die Teile A und B, so gilt für die Größe der Oberflächen:

  $F(G) = F(A) + F(B) - 2\cdot\text{Schnittfläche}$

  Analog eine Dimension kleiner: zerschneidet man ein Stück Papier, so gilt die analoge Formel für den Umfang.
- Gibt es Mengenrabatt, so kann man die Einzelpreise nicht einfach addieren:
  Ist z. B. P der Preis,

  A = 5 Brötchen und
  B = 5 andere Brötchen

  und wird Mengenrabatt ab 10 Brötchen gewährt, so ist $A\cap B=\emptyset$, aber es gilt:

  $P(A\cup B) < P(A) + P(B)$

  Auch bei der Preisbündelung hat man den Effekt, dass der Gesamtpreis geringer als die Summe der Einzelpreise ist.

(4.3.-03) ***Fall 3***: Die Kombination ist nicht sinnvoll
Bei wiederum anderen Bewertungen stellt sich die Frage nach der Sinnhaftigkeit von Kombinationen. So hatte z. B. die Stiftung Warentest Wundpflaster in einem Test untersucht und bewertet (siehe [WU]). Aber wie sollte man die Qualität der Kombination von zwei verschiedenen Pflastern bewerten? Und was heißt überhaupt „Kombination von zwei Pflastern“? Dass man sie übereinander klebt?

Oder dass jedes Pflaster nur die Hälfte der Wunde überdeckt? Diese Interpretationen scheinen alle nicht besonders sinnvoll zu sein.
Hat man z. B. Smartphones getestet, so könnte man zwei sich ergänzende Smartphones kaufen und je nach Bedarf mal das eine und mal das andere benutzen. Eine solche Nutzung wäre dann eine Kombination von Smartphones. Das so gebildete virtuelle Supersmartphone hat dann in allen Einzeleigenschaften den jeweils besseren Wert der beiden Smartphones. Man kann sich also mit etwas Mühe den Fall 1 hinbiegen, das Beispiel wirkt aber ziemlich konstruiert.
Aber bei Autoreifen wird es dann aber richtig schwierig, da man nicht je nach Fahrsituation gleich zu den jeweils besten Reifen wechseln kann. Die Kombination der verschiedenen Eigenschaften unterschiedlicher Autoreifen wäre praktisch sinnlos.

(4.3.-04) ***Hinweise***

1. Die Gefahr bei allen Bewertungen ist, dass durch die Angabe von Zahlenwerten eine Genauigkeit und Objektivität suggeriert wird, die tatsächlich oft nicht gegeben ist.

2. Die Bewertungstabelle kann beliebig kompliziert oder einfach sein. Im täglichen Leben bewertet man etwas oft aus dem Bauch heraus oder weil einem etwas besser gefällt als etwas Anderes, ohne dass das im Detail näher untersucht oder hinterfragt wird. Die Bewertungstabelle besteht dann nur aus einer Zeilen, nämlich dem Bauchgefühl. Bei den Warentests dagegen wird eine mehrstufige Bewertung vorgenommen.

3. Bewertet man mehrere Objekte bezüglich einer Eigenschaft, so ist es notwendig, dass das Verfahren für alle Objekte einheitlich ist.

4. Manchmal müssen scheinbar unvereinbare Objekte verglichen werden, um zu Entscheidungen zu kommen. Dazu müssen diese Objekte nach einem einheitlichen Verfahren bewertet werden – sofern man nicht einfach nur eine Bauchentscheidung trifft.
   Wenn ich 10€ habe, kann sich für den Abend die Frage stellen: „Gehe ich ins Kino oder esse ich eine Pizza?“. Das einheitliche Verfahren enthält dann z. B. Kriterien wie „geschmackliche Zufriedenheit“, die beim Kinobesuch gering und bei der Pizza hoch ist, während es sich beim Kriterium „Spaß durch spannende Unterhaltung“ vermutlich umgekehrt verhält. Man muss sich dann festlegen, was einem die einzelnen Kriterien wert sind.

5. Durch die Liste der einzelnen Faktoren in der Bewertungstabelle hat man zunächst das Problem nur auf eine Detailebene tiefer verlagert, denn für die einzelnen Faktoren stellt sich genauso die Frage, wie man sie bewertet – und dann hat man sogar mehrere Faktoren! Idealerweise sind die einzelnen Faktoren aber einfacher zu bewerten als die Ausgangssituation, wie bei den Warentests. Zumindest sorgt die Bewertungstabelle für Transparenz, sodass andere besser erkennen können, was sich der Autor dabei gedacht hat.

6. Unterschiedliche Personen können bei derselben Eigenschaft derselben Objekte zu unterschiedlichen Bewertungen kommen, weil sie
   a. unterschiedliche Einflussfaktoren auswählen
   b. die ausgewählten Faktoren unterschiedlich bewerten
   c. die ausgewählten bewerteten Faktoren unterschiedlich für den endgültigen Wert gewichten

   Darüber hinaus können sich alle drei Aspekte im Laufe der Zeit verändern, da die bewertende Person weitere Erfahrungen macht, die zu einer anderen Bewertung führen.

So kann zum Beispiel die Bewertung durch Stiftung Warentest nach anderen Kriterien als bei einer Fachzeitschrift durchgeführt werden – obwohl dieselben Objekte bewertet werden.

7. Gelegentlich liest man, dass die Bildung eines Notendurchschnitts nicht zulässig ist, da Schulnoten nur einer Ordinalskala, nicht aber einer metrischen Skala genügen (siehe z. B. [BG], Seite 15). Diese Kritik ist aus zweierlei Gründen unzureichend:
    - Schon beim Ermitteln einer einzelnen Schulnote werden verschiedene objektive und subjektive Kriterien bewertet und zu der Note zusammengeführt, also eine Summe oder ein arithmetisches Mittel gebildet. Konsequenterweise müsste man bereits die Existenz jeder einzelnen Schulnote kritisieren.
    - Anhand der vielen Beispiele oben sieht man, dass das Verfahren der Summen- oder Durchschnittsbildung bei sehr vielen Bewertungen genutzt wird.

    Die Ermittlung eines Notendurchschnitts ist also grundsätzlich genauso sinnvoll oder nicht sinnvoll wie alle anderen Bewertungen in den Beispielen oben auch. Richtig ist aber natürlich, dass man Durchschnitte von irgendwelchen Bewertungen mit kritischer Distanz interpretieren sollte.

4.4 Zusammenfassung

Es gibt (fast) nichts, was nicht von einer Person oder einer Organisation bewertet wird. Mit Hilfe einer Bewertung kann zumindest eine Rangordnung festgelegt werden, die Bewertung ist also zumindest komparativ. Für eine Bewertung müssen definiert werden:

- Die Objekte, die betrachtet werden sollen
- Die zu bewertende Eigenschaft der Objekte
- Die Bewertungskategorien
    - mit Worten wie schön/hässlich, gut/schlecht, groß/klein
    - mit Zeichenketten wie AAA oder BB-
    - mit Zahlen und ggf. zusätzlich Maßeinheiten wie Punkte oder €
- Eine genaue Beschreibung des Bewertungsverfahrens
  Dabei werden die einzelnen Einflussfaktoren bestimmt, nach spezifischen Verfahren bewertet, auf eine einheitliche Größe umgerechnet und dann zur Gesamtbewertung verdichtet. Die spezifischen Bewertungen der einzelnen Einflussfaktoren können durch
    - objektive Messungen
    - den Durchschnitt von (mehr oder weniger subjektiven) Bewertungen durch mehrere Personen,
    - ggf. weitere Verfahren

  ermittelt werden. Diese Einzelwerte werden dann zu dem Wert für den Einflussfaktor verdichtet.

Bewertungen von Kombinationen mehrerer Objekte sind manchmal sinnvoll, manchmal nicht.
Wenn die Kombination sinnvoll interpretierbar ist, so kann die Gesamtbewertung in einfachen Fällen subadditiv sein und analog zu einem Mengen-Diagramm dargestellt werden. Für zwei Objekte A und B und eine Bewertung F ist dann:

F(Kombination von A und B) = F(A) + F(B) – F(Überschneidung von A und B)

oder in Mengenschreibweise:

$$F(A \cup B) = F(A) + F(B) - F(A \cap B)$$

Dabei ist wieder Voraussetzung, dass es ein einheitliches Bewertungsverfahren gibt.

5 Definition und Berechnung der Wahrscheinlichkeit

Zusammenfassung
Nun ist es endlich so weit: in diesem Kapitel wird plausibel hergeleitet, dass Wahrscheinlichkeiten einfach nur ein Beispiel für quantitative Bewertungen sind, also etwas eher Unspektakuläres. Es wird eine umfassende Definition der praktischen Wahrscheinlichkeit hergeleitet, die alle Fälle abdeckt – von einmaligen Ereignissen wie dem Grexit bis hin zu standardisierten Versuchsreihen wie dem mehrfachen Würfeln. Darüber hinaus wird ein übergreifendes Verfahren hergeleitet, mit dem man in allen Fällen die Wahrscheinlichkeit konkret bestimmen kann ohne auf Wahrscheinlichkeiten anderer Aussagen zurückgreifen zu müssen. Zusätzlich werden Rechenregeln für Wahrscheinlichkeiten, wie etwa die Subadditivität, plausibel hergeleitet.
Besonders wichtig ist: Begriffe wie Zufall, Experiment, Ereignis, Kolmogoroffsche Axiome oder Ähnliches werden bei dieser Vorgehensweise nicht benötigt.
* * *

Über die möglichen Interpretationen des Begriffs Wahrscheinlichkeit gibt es sehr viele verschiedene Ausführungen in Büchern oder anderen Veröffentlichungen. Es werden die unterschiedlichsten Arten von Wahrscheinlichkeiten beschrieben: die subjektive, die objektive, die tatsächliche, die theoretische, die a-priori, die a-posteriori, die statistische, die geometrische, die Kolmogoroffsche, die axiomatische, die klassische, die Laplacesche, die logische, die statistische, die empirische, die personelle, die induktive, die wissenschaftliche, die mathematische, die epistemische, die kurzfristige, die langfristige, … Wahrscheinlichkeit. Für den Leser trägt das nicht gerade zur Klarheit bei, solange der Begriff Wahrscheinlichkeit nicht definiert ist.

In vielen Veröffentlichungen werden mehrere diese Wahrscheinlichkeitsarten beschrieben (siehe z. B. [SJ], Kapitel 8; [AE], Kapitel 3.2; [AT], Kapitel 3). Dann drängen sich aber folgende Fragen auf:
- Gelten dieselben Rechenregeln für alle Arten von Wahrscheinlichkeiten?
- Wie hängen die verschiedenen Arten von Wahrscheinlichkeiten zusammen?
- Wie stellt man in der Praxis fest, welche Art von Wahrscheinlichkeit vorliegt bzw. anzuwenden ist?
- Was ist zu tun, wenn man eine Mischung aus verschiedenen Wahrscheinlichkeiten hat?
- Gibt es eine übergeordnete Definition, die alle Aspekte umfasst?
- Was hat insbesondere die axiomatische Wahrscheinlichkeit nach Kolmogoroff, die oft als Basis für das Rechnen mit Wahrscheinlichkeiten benutzt wird, mit dem umgangssprachlichen realitätsbezogenen Begriff „Wahrscheinlichkeit“ zu tun?

Und dann wird oft auch noch explizit geschrieben, dass es keine einheitliche Auffassung darüber gibt, was Wahrscheinlichkeit in der Praxis konkret bedeutet, so dass keine konkrete die Realität betreffende Definition dargestellt werden kann (siehe z. B. [HR], Seite 1 und 6). Wenn man aber keine Definition für die konkrete praktische Wahrscheinlichkeit hat, dann bleibt auch unklar, was die Aussage „die Wahrscheinlichkeit, eine 3 zu würfeln, ist gleich $\frac{1}{6}$“, konkret bedeuten soll. Konsequenterweise wären dann alle Abschnitte in Büchern, die sich mit der Anwendung der Wahrscheinlichkeitsrechnung auf konkrete Fragestellungen befassen, hinfällig.

Das sind nicht gerade gute Voraussetzungen für eine Wahrscheinlichkeitsrechnung, die sich mit konkreten praktischen Fragestellungen befassen soll. Und das axiomatische Modell nach Kolmogoroff hilft in der Praxis auch nicht weiter (siehe Kapitel 8.1).

Im Folgenden wird gezeigt, dass zwei Begriffsbildungen ausreichen, um alle Aspekte abzudecken: die praktische auf der Umgangssprache basierende Wahrscheinlichkeit für konkrete Fragestellungen und in der Mathematik die axiomatische Wahrscheinlichkeit nach Kolmogoroff.
Die axiomatische Wahrscheinlichkeit nach Kolmogoroff ist nur ein normiertes Maß im Sinne des mathematischen Teilgebietes Maßtheorie, sodass man das Wort „Wahrscheinlichkeit“ ohne Verlust an Aussagekraft durch ein beliebiges anderes Wort ersetzen könnte. Vorschlag in Kapitel 8.1: wählt man konsequent das Wort „Blubb“ statt „Wahrscheinlichkeit“, so kann es zu keinen Missverständnissen oder Fehlinterpretationen kommen.

Grundlage der Betrachtung ist die Feststellung, dass eine Wahrscheinlichkeit einfach nur ein Beispiel für eine Bewertung von „etwas“ ist. Die Erkenntnisse aus Kapitel 4 können also auf das Beispiel Wahrscheinlichkeit übertragen und dann weiter vertieft werden.
Auf dieser Basis wird dann eine übergreifende Definition der Wahrscheinlichkeit entwickelt, die alle praktisch relevanten Aspekte abdeckt. Zusätzlich werden Verfahren zur Bestimmung einer Wahrscheinlichkeit plausibel hergeleitet, die dann zu einem übergreifenden Verfahren für beliebige Wahrscheinlichkeiten zusammengefasst werden.

5.1 Umgangssprachliche Definition

Ob in der Umgangssprache, in einer Tageszeitung oder in wissenschaftlichen Untersuchungen: überall wird etwas als wahrscheinlich oder unwahrscheinlich dargestellt, manchmal sogar mit einer konkreten Prozentangabe. Typisch sind Sätze der Art wie:

- Ich halte es für sehr unwahrscheinlich, dass es den Yeti gibt.
- Nach meiner Meinung regnet es wahrscheinlich morgen in Hannover.
- Beim Münzwurf beträgt die Wahrscheinlichkeit für Wappen 50 %.

Offensichtlich ist eine Wahrscheinlichkeit ebenso eine Bewertung wie die vielen Beispiele in Kapitel 4.1. Jeder, der sich mit der praktischen Anwendung von Wahrscheinlichkeiten beschäftigt, muss also folgende Fragen beantworten:

- Welche Objekte werden bezüglich welcher Eigenschaft durch eine Wahrscheinlichkeit bewertet?
- Wie ist das Bewertungsverfahren?

Zerlegt man das Wort in seine Bestandteile, so drückt die Wahr-schein-lichkeit aus, wie wahr etwas einer Person zu sein scheint. Es geht also nicht um Wahrheit an sich, sondern um den Anschein, inwieweit etwas wahr ist. „Wahr“ kann mehrere Bedeutungen haben, z. B. bedeutet „wahre Freundschaft“ so viel wie „echte Freundschaft“, aber der Satz „die Freundschaft scheint wahr zu sein“ würde man nicht umformulieren zu „die Freundschaft hat eine hohe Wahrscheinlichkeit“.

„Wahr“ bezieht sich immer auf Aussagen, so dass man die Beispiele oben auch umformulieren kann zu:

- Die Aussage „es gibt den Yeti“ scheint mir sehr unwahr zu sein.

- Die Aussage „morgen regnet es in Hannover“ scheint mir wahr zu sein.
- Die Aussage „beim Münzwurf wird Wappen geworfen“ scheint mir zu 50 % wahr zu sein.

Und was bedeutet das jetzt konkret?

In der Mathematik ist die Situation einfach und objektiv. Für jede Aussage gilt einer der folgenden Fälle:
- Die Aussage ist wahr, weil sie bewiesen wurde, z. B.: der Satz des Pythagoras
- Die Aussage ist falsch, weil sie widerlegt werden kann, z. B.: $2 + 2 = 5$
- Man weiß nicht, ob die Aussage wahr oder falsch ist, wie z. B. die Goldbachsche Vermutung.

Bei Aussagen über die Realität ist das schwieriger, da man im Regelfall nicht alle Informationen für eine sichere Bewertung zur Verfügung hat. Stattdessen kann man nur eine Einschätzung abgeben, wie plausibel einem eine Aussage erscheint oder wie sehr man an die Richtigkeit einer Aussage glaubt. Dabei sind Einschätzungen personen- und zeitabhängig: was dem einen plausibel erscheint, kann für einen anderen unplausibel sein und was mir heute noch unplausibel erscheint, kann ich morgen bereits erklären.

Diese Einschätzung ist die Wahrscheinlichkeit, genauer:

(5.1.-01) ***<u>Definition</u>***
Umgangssprachlich ist die
Wahrscheinlichkeit einer Aussage = Grad des Anscheins, dass die Aussage wahr ist.

Andere Formulierungen sind:
- Grad der Überzeugung, dass die Aussage wahr ist
- Grad des Glaubens, dass die Aussage wahr ist
- Grad der Plausibilität der Aussage oder „Maß für die Plausibilität von Aussagen“ ([KK], Seite 3)

Welche dieser Formulierungen man bevorzugt, ist im Rahmen dieses Buches unwichtig, da sie sehr ähnliche Interpretationen ausdrücken und die mathematischen Verfahren davon unabhängig sind. Die so definierte Wahrscheinlichkeit wird auch kurz als **umgangssprachliche Wahrscheinlichkeit** bezeichnet.
* * *

Diese Definition wird in Kapitel 5.6 noch weiter konkretisiert.

Aus den Beispielen oben wird dann:
- Der Grad meiner Überzeugung, dass die Aussage „es gibt den Yeti“ wahr ist, ist sehr gering.
- Der Grad meines Glaubens, dass die Aussage „morgen regnet es in Hannover“ wahr ist, ist hoch.
- Der Grad des Anscheins, dass die Aussage „beim Münzwurf wird Wappen geworfen“ wahr ist, ist für mich 50 %.

5.2 Erläuterungen zur umgangssprachlichen Definition

(5.2.-01) ***Erläuterung***
Historisch ging es zunächst nur um Berechnungen der möglichen Ergebnisse bei Glücksspielen, also um Kombinatorik und daraus abgeleitete relative Häufigkeiten. Würfelt man z. B. mit zwei Würfeln einmal, so gibt es 36 verschiedene Ergebnisse, von denen 7 die Augensumme 7 haben. Die relative Häufigkeit dieser Augensumme in der Liste der möglichen Ergebnisse ist also $\frac{7}{36}$. Da das, was relativ selten / häufig vorkommt, für mich eine geringe / hohe Wahrscheinlichkeit hat, wurde diese relative Häufigkeit einfach als die Wahrscheinlichkeit interpretiert (klassische Definition der Wahrscheinlichkeit nach Laplace, siehe Kapitel 6.3).

(5.2.-02) ***Erläuterung***
Wichtig für die Definition des Begriffes „Wahrscheinlichkeit" ist, dass damit nur die eigene Einschätzung des Plausibilitätsgrades einer Aussage ausgedrückt wird. Inwieweit diese Aussage die Realität tatsächlich widerspiegelt, bleibt offen.
Wenn ich sage: „Die Wahrscheinlichkeit für den Grexit beträgt 25 %", so ist das meine Einschätzung und nicht eine objektive Tatsache. Andere Personen können andere Einschätzungen haben und somit zu anderen Wahrscheinlichkeiten gelangen.
Wir werden im Folgenden sehen, dass dieses subjektive Element in der Praxis immer einen Einfluss auf die Wahrscheinlichkeit hat, selbst bei Standardbeispielen wie dem Würfeln oder dem Münzwurf. Auch bei diesen Standardbeispielen können verschiedene Personen zu verschiedenen Wahrscheinlichkeiten kommen.

(5.2.-03) ***Erläuterung***
Manchmal wird die Wahrscheinlichkeit auch als „Grad der Gewissheit" definiert (siehe [WK]). Diese Formulierung kann missverstanden werden, denn:

- eine hohe Wahrscheinlichkeit drückt eine hohe Gewissheit (Sicherheit) aus, dass die Aussage richtig ist; siehe z. B.: http://www.t-online.de/unterhaltung/tv/id_80223760/-wer-wird-millionaer-jauch-kandidat-max-glaubt-dem-falschen-joker.html
- eine niedrige Wahrscheinlichkeit drückt eine hohe Gewissheit (Sicherheit) aus, dass die Aussage falsch ist
- hat man überhaupt keine Gewissheit / Sicherheit über den Wahrheitsgehalt einer Aussage, so räumt man der Richtigkeit und der Falschheit der Aussage dieselben Chancen ein.
  „Ich bin total unsicher, ob es morgen regnet", heißt nicht, dass ich total sicher bin, dass es morgen nicht regnet, sondern ich halte beides für gleichermaßen möglich. Umgangssprachlich ist also
  minimale Sicherheit = maximale Unsicherheit = 50 % Sicherheit

(5.2.-04) ***Erläuterung***
Unterschiedliches Wissen kann zu unterschiedlichen Wahrscheinlichkeiten führen. Dabei ist aber zu berücksichtigen:

- Wahrscheinlichkeiten werden auch durch irrationale oder unbewusste Faktoren bestimmt, nicht nur durch Wissen.
- Aus dem Umfang des Wissens folgt nicht die Höhe der Wahrscheinlichkeit: wenn man viel weiß, kann die Wahrscheinlichkeit groß oder klein sein und wenn man wenig weiß, ebenfalls.
- Aus der Höhe der Wahrscheinlichkeit kann nicht auf das Wissen geschlossen werden. Die bewertende Person kann Experte oder völlig ahnungslos sein und trotzdem kann in beiden Fällen dieselbe Wahrscheinlichkeit genannt werden.

- Wissen wird z. B. in der Schule durch eine Schulnote bzw. durch Punkte gemessen: für viel Wissen gibt es viele Punkte, für wenig Wissen gibt es wenig Punkte. Diese Punkte entsprechen aber nicht Wahrscheinlichkeiten, eine Wahrscheinlichkeit ist kein Maß für Wissen.
- Formulierungen wie „Der Zustand des Wissens über eine Aussage wird durch die Wahrscheinlichkeit ausgedrückt“ ([KK], Seite 1) sind also nicht wie eine Messung (viel/wenig Wissen bedeutet hohe/niedrige Wahrscheinlichkeit), sondern nur als Hinweis, dass die Wahrscheinlichkeit vom Wissen abhängen kann, zu verstehen.

5.3 Beispiele für die umgangssprachliche Definition

(5.3.-01) ***Beispiel***
Im Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2015/16 werden Aussagen über die künftige wirtschaftliche Entwicklung gemacht (siehe [SR]). Dabei fließen nicht nur Fakten, sondern auch Erfahrungen und Einschätzungen ein, die man teilen kann oder auch nicht. In diesem Gutachten taucht das Wort „wahrscheinlich“ in verschiedenen Zusammensetzungen 26-mal auf.

(5.3.-02) ***Beispiel***
Bei der Ermittlung der Ausfallwahrscheinlichkeit eines Kredites fließen sowohl objektive, als auch subjektive Kriterien ein. Verschiedene Personen, Banken oder Lieferanten können zu unterschiedlichen Wahrscheinlichkeiten kommen, da sie unterschiedliche Kriterien benutzen.

(5.3.-03) ***Beispiel***
In Krimis wird von erfahrenen Kommissaren gerne gegenüber weniger erfahrenen Kommissaren geäußert „Ich bin mir sehr sicher, dass Herr X nicht der Mörder ist, ich kann aber nicht erklären, warum – das ist nur so ein Bauchgefühl“. Der erfahrene Kommissar ordnet der Aussage „Herr X ist der Mörder“ also eine sehr hohe Wahrscheinlichkeit zu, kann es aber nicht stichhaltig begründen.

(5.3.-04) ***Beispiel***
Beim Glücksspiel (Würfeln, Roulette, Lotto, Kartenspiele, …) oder Wetten schätzen manche Leute die Chancen im Vergleich zur relativen Häufigkeit zu optimistisch ein, da sie von dem Wunsch nach Gewinn beherrscht werden. Die Wahrscheinlichkeitswahrnehmung wird z. B. in der Psychologie, aber auch bei Juristen untersucht. Auch dieser Effekt zeigt, dass bei der Ermittlung einer Wahrscheinlichkeit nicht nur rationale, sondern auch irrationale Faktoren eine Rolle spielen.

(5.3.-05) ***Beispiel***
Der sogenannte Spielerfehlschluss drückt die Annahme aus, dass etwas wahrscheinlicher wird, wenn es deutlich seltener als die relative Häufigkeit aufgetreten ist bzw. dass es unwahrscheinlicher wird, wenn es deutlich häufiger als die relative Häufigkeit aufgetreten ist. Es handelt sich aber nur dann um einen Fehlschluss, wenn die Bedingungen unveränderlich sind und wenn plausibel ist, dass die relative Häufigkeit das richtige Maß für die Wahrscheinlichkeit ist. Der Spielerfehlschluss wird auch manchmal durch Aussagen der Art „Würfel haben kein Gedächtnis“ veranschaulicht.

Wenn man dagegen z. B. auf eine Zielscheibe schießt oder wirft, so wird es immer wahrscheinlicher, dass man ins Schwarze trifft, da man immer besser wird. Zwar haben die Zielscheibe und das Gewehr oder der Dartpfeil kein Gedächtnis, aber der Schütze hat eins.

(5.3.-06) ***Beispiel***
Durch die Software precobs (Pre Crime Observation System) vom Institut für musterbasierte Prognosetechnik sollen mit einer bestimmten Wahrscheinlichkeit Einbrüche vorhergesagt werden können. Näheres siehe [PR].

(5.3.-07) ***Beispiel***
In der forensischen Musikwissenschaft wird untersucht, inwieweit sich zwei Musikstücke ähneln. Daraus kann dann die Wahrscheinlichkeit abgeleitet werden, dass ein Komponist vom anderen abgeschrieben hat. Auf dieser Basis können dann Regressforderungen erhoben werden.

5.4 Quantifizierung der Wahrscheinlichkeit

Man spricht von einer hohen oder einer geringen Wahrscheinlichkeit, dass eine bestimmte Aussage zutrifft, die Wahrscheinlichkeit wird also zumindest grob quantifiziert. Es liegt also nahe, analog zu den Bewertungen in Kapitel 4 hohen Wahrscheinlichkeiten große Zahlen und geringen Wahrscheinlichkeiten kleine Zahlen zuzuordnen. Dies ist der entscheidende Hebel, um mathematische Verfahren anwenden zu können, denn mit Zahlen kann man rechnen, mit „hoch“ und „niedrig“ nicht. Wie aber auch bei den Bewertungen in Kapitel 4 suggerieren exakte Zahlen eine Genauigkeit und Objektivität, die tatsächlich nur bedingt gegeben ist – also Vorsicht!

Man nimmt für die Wahrscheinlichkeiten die Zahlen von 0 bis 1 bzw. von 0 % bis 100 %, denn:
1. Spricht man von einer 0 %-igen oder 100 %-igen Wahrscheinlichkeit, dass eine Aussage richtig ist, so ist man sich absolut sicher, dass diese Aussage falsch bzw. richtig ist.
   Der geringsten Wahrscheinlichkeit wird also die Zahl 0 = 0 % und der höchsten Wahrscheinlichkeit die Zahl 1 = 100 % zugeordnet. Alle anderen Wahrscheinlichkeiten liegen dazwischen.
2. In vielen Fällen ergibt sich die Wahrscheinlichkeit direkt aus einem relativen Anteil, der ja ebenfalls zwischen 0 und 1 liegt. Dieser Aspekt wird in den folgenden Kapiteln näher erläutert.
3. In der Boolschen Algebra und klassischen Aussagenlogik wird „wahr“ mit 1 und „falsch“ mit 0 gleichgesetzt.
4. In bestimmten Fällen ist auch das Produkt zweier Wahrscheinlichkeiten wieder eine Wahrscheinlichkeit. Durch die Festlegung, dass Wahrscheinlichkeiten immer zwischen 0 und 1 liegen, wird sichergestellt, dass auch Produkte wieder zwischen 0 und 1 liegen, also eine Wahrscheinlichkeit darstellen können (siehe auch Kapitel 5.9.2 und 5.9.4).

Nimmt man einen anderen Zahlenbereich, so darf man mit den Wahrscheinlichkeiten nur bestimmte Operationen wie Durchschnittswertermittlung oder Vergleiche durchführen. Beispielsweise findet man bei [SU] für Wahrscheinlichkeiten die ganzen Zahlen von 0 bis 10.

Zur Beschreibung der Höhe einer Wahrscheinlichkeit kann man einen Zusammenhang zwischen umgangssprachlichen Begriffen und Zahlen herstellen. Das kann z. B. eine Zuordnung sein wie:

Begriff	Intervall
sehr unwahrscheinlich	0 % ≤ Wahrscheinlichkeit < 10 %
unwahrscheinlich	10 % ≤ Wahrscheinlichkeit < 30 %
möglich	30 % ≤ Wahrscheinlichkeit < 70 %
wahrscheinlich	70 % ≤ Wahrscheinlichkeit < 90 %
sehr wahrscheinlich	90 % ≤ Wahrscheinlichkeit ≤ 100 %

Andere Zuordnungen zwischen Begriffen und Zahlen findet man z. B. beim Zwischenstaatlichen Ausschuss für Klimaänderung (IPCC) (siehe [IP]) oder bei Tetlock / Gardner [TP], Seite 67.

(5.4.-03) ***Unterschied Wahrscheinlichkeit – fuzzy logic***

In Rahmen der Stochastik können Aussagen stets nur wahr oder falsch sein, nichts dazwischen. Nur der Grad der Überzeugung kann variieren. Bei der fuzzy logic gibt es dagegen auch alle dazwischenliegenden Wahrheitswerte, meine Überzeugung spielt keine Rolle.
Beispiel:
Es klassischer Lichtschalter kann nur an- und ausschalten, es gibt also nur zwei Zustände: das Licht ist an oder das Licht ist aus. Wenn ich im Wohnzimmer bin, dann kann ich zu 90% überzeugt sein, dass ich das Licht im Bad ausgemacht habe, die Wahrscheinlichkeit, dass das Licht aus ist, ist für mich 90%. Wenn im Bad aber ein Dimmer ist und ich bin im Bad, dann kann ich sagen, dass das Licht zu 90% aus ist, wenn der Dimmer fast ganz zugedreht ist. Solche Aussagen werden nicht in der Stochastik, sondern im Rahmen der fuzzy logic untersucht.

(5.4.-04) ***Schreibweise***

Ist A eine Aussage, so wird mit P(A) die Wahrscheinlichkeit der Aussage bezeichnet. P ist die Wahrscheinlichkeitsbewertung oder kurz die Wahrscheinlichkeit und steht für „probability“. P ordnet also einer Aussage nach einem bestimmten Verfahren eine Zahl zu. Die beiden Extremfälle sind:

- P(A) = 0 ist gleichbedeutend mit „die bewertende Person hat keinen Zweifel daran, dass die Aussage A falsch ist“ oder „die bewertende Person ist sich absolut sicher, dass die Aussage A falsch ist“.
- P(A) = 1 ist gleichbedeutend mit „die bewertende Person hat keinen Zweifel daran, dass die Aussage A wahr ist“ oder „die bewertende Person ist sich absolut sicher, dass die Aussage A wahr ist“.

(5.4.-05) ***Einordnung der Wahrscheinlichkeit***

Die folgende Übersicht macht anhand der bereits erwähnten Beispiele in Kapitel 4 deutlich, dass die Wahrscheinlichkeit einfach nur eine weitere quantitative Bewertung neben sehr vielen anderen quantitativen Bewertungen ist.

Objekt	Eigenschaft / Teilaspekt	Bewertungsgröße
Schüler/innen	Leistung	Note
Mitarbeiter/innen	Zielerreichung	%
Volkswirtschaft	Preis-/Mengenentwicklung	dimensionsloser Index
Ware	Qualität	Note oder Punkte
Staat	Friedlichkeit	Global Peace Index; Wert aus [1, 5]
Aussage	***Wahrheit***	***Wahrscheinlichkeit; Wert aus [0, 1]***

Staat	Gleichstellung der Geschlechter	Global Gender Gap Index; Wert aus [0, 1]
Mensch	Glücksgefühl	Glücksindex gemäß World Happiness Report
Mensch	Intelligenz	IQ in %
Dachpappe	Fläche	m²
Mozzarella	Qualität	Note
…	…	…

Die Wahrscheinlichkeit ist also nichts Besonderes oder Mystisches, aber auch nicht klarer oder unklarer als die anderen Bewertungen dieser Liste, sondern einfach eine ganz normale Bewertung wie viele andere Bewertungen auch.

5.5 Ähnliche Begriffe

Statt „Wahrscheinlichkeit“ werden oft andere Begriffe oder Formulierungen benutzt.

(5.5.-01) ***Andere umgangssprachliche Begriffe***
Es gibt viele andere Formulierungen, um die Einschätzung über die Richtigkeit einer Aussage auszudrücken. Statt „Die Wahrscheinlichkeit ist hoch, dass …“ sagt man auch:

- „Es ist wahrscheinlich / es liegt nahe, dass …“
- „Ich bin überzeugt / glaube / vermute / nehme an / erwarte /…, dass …“

(5.5.-02) ***Chance / Risiko***
In Sätzen wie „Die Wahrscheinlichkeit, dass …, beträgt …“ benutzt man statt „Die Wahrscheinlichkeit“ auch

- „Das Risiko“ (= Gefahr), wenn man negative Aspekte betonen möchte
- „Die Chance“, wenn man positive Aspekte betonen möchte.

Auch hier quantifiziert man mit den Begriffen „hoch“/“groß“, „niedrig“/“klein“, „50%-ig“, etc.

Vom Sport kennt man die Aussage: „Er hat eine fast 100%-ige Chance vergeben“, wobei das Wort „fast“ gerne weggelassen wird, um zu dramatisieren.

In betrieblichen Risikomanagementsystemen gemäß dem Gesetz zur Kontrolle und Transparenz im Unternehmensbereich (KonTraG) (siehe [KO]) geht man häufig folgendermaßen vor:

Schritt 1: Ermittlung der Ziele (des Unternehmens, der Abteilung, des Projektes, …)
Schritt 2: Ermittlung der Verfahren zu Erreichung der Ziele
Schritt 3: Ermittlung der Gefahren, die verhindern können, dass die Ziele erreicht werden. Diese Gefahren sind die Risiken.
Schritt 4: Bewertung der Risiken nach Eintrittswahrscheinlichkeit und Schadenshöhe. Häufig wird dazu das Produkt der Eintrittswahrscheinlichkeit mal Schadenshöhe genommen, was voraussetzt, dass man der Eintrittswahrscheinlichkeit eine Zahl zuordnet (was in der betrieblichen Praxis recht spekulativ sein kann).

Schritt 5: Einleitung von Gegenmaßnahmen zur Reduzierung der Risiken unter Berücksichtigung der Wirtschaftlichkeit, gesetzlicher Vorgaben oder anderer Rahmenbedingungen. Zusätzlich werden Maßnahmen zur Abdeckung der dann noch verbleibenden Risiken definiert (z. B. Abdeckung durch Rückstellungen oder Versicherungen).

5.6 Die Definition der praktischen Wahrscheinlichkeit

Die folgenden Definitionen fassen die bisherigen Ergebnisse zusammen.

(5.6.-01) ***Definition***
Unter der **Einstellung** einer Person werden Wissen, Erfahrungen, Annahmen oder Gefühle, also alles, was unser Denken und Handeln beeinflusst, verstanden.

(5.6.-02) ***Definition***

1. Die **Wahrscheinlichkeit** einer Aussage drückt den Grad des Anscheins aus, dass die Aussage wahr ist.
   Andere Formulierungen sind:
   - Grad der Überzeugung, dass die Aussage wahr ist
   - Grad des Glaubens, dass die Aussage wahr ist
   - Grad der Plausibilität der Aussage

   Welche dieser Formulierungen man bevorzugt, ist im Rahmen dieses Buches unwichtig, da sie sehr ähnliche Interpretationen ausdrücken und die mathematischen Verfahren davon unabhängig sind.

2. Die Wahrscheinlichkeit einer Aussage ist eine Bewertung der Aussage bezüglich der Wahrheit. Die Wahrscheinlichkeit wird durch Worte wie „groß / klein“, „hoch / niedrig“ oder durch eine Zahl beschrieben. Die Bewertung ist von der Einstellung der bewertenden Person abhängig.

3. Ist A eine beliebige Aussage und ist P eine quantitative Wahrscheinlichkeitsbewertung, so gilt:
   - $P(A)$ = Wahrscheinlichkeit, dass die Aussage A wahr ist, kurz: Wahrscheinlichkeit von A
   - $0 \leq P(A) \leq 1$
   - $P(A) = 0$ bedeutet: die bewertende Person ist ohne jeden Zweifel davon überzeugt, dass die Aussage A falsch ist
   - $P(A) = 1$ bedeutet: die bewertende Person ist ohne jeden Zweifel davon überzeugt, dass die Aussage A wahr ist

4. Die so definierte Wahrscheinlichkeit wird in Abgrenzung zur axiomatischen Wahrscheinlichkeit nach Kolmogoroff (siehe Kapitel 8.1) als **praktische Wahrscheinlichkeit** – im Folgenden kurz „Wahrscheinlichkeit“ – bezeichnet.

(5.6.-03) ***Hinweis***
Die Definition umfasst alle Aspekte, die man in der Praxis für die Interpretation der Wahrscheinlichkeit benötigt. Damit ist auch die oft gestellte Frage, was Wahrscheinlichkeit konkret bedeutet (siehe z.

B. [HR], Seite 1 und 6), beantwortet. Es ist auch nicht nötig, verschiedene Auffassungen oder Definitionen von Wahrscheinlichkeit zu betrachten, wie das in vielen Veröffentlichungen gemacht wird (siehe z. B. [SJ], Kapitel 8). Insbesondere gibt es nicht den Gegensatz zwischen der subjektiven und der objektiven Wahrscheinlichkeit, sondern beides sind nur Teilaspekte der einheitlichen Definition. Bei den Verfahren zur Ermittlung einer Wahrscheinlichkeit wird auf diesen Aspekt noch konkreter eingegangen.

(5.6.-04) ***Hinweis***
Die Definition basiert auf der umgangssprachlichen Verwendung des Begriffs Wahrscheinlichkeit und kommt deswegen ohne die Begriffe „**Zufall**", „**Ereignis**" und „**Experiment**" aus. Aussagen wie „es regnet morgen" oder „es gibt den Yeti" haben weder etwas mit Zufall, noch mit Experimenten zu tun, aber trotzdem kann man dazu bestimmte Überzeugungen haben und Wahrscheinlichkeiten bestimmen. Die Wahrscheinlichkeitsrechnung und Stochastik sind also nicht Wissenschaften, die den Zufall erforschen oder mit der man den Zufall berechnen kann.
Kurz: „Wahrscheinlichkeit" hat im Allgemeinen nichts mit „Zufall" zu tun.

(5.6.-05) ***Hinweis***
Würde man Wahrscheinlichkeiten in einer Datenbank speichern, so wäre die Datenstruktur im einfachsten Fall:

(*Aussage; Bewertende Person; Datum*; Wahrscheinlichkeit)

Die kursiven Felder bilden den eindeutigen Schlüsselbegriff. Damit wird dokumentiert, dass die Wahrscheinlichkeit nicht nur von der Aussage, sondern auch von der Einstellung der bewertenden Person zu einem bestimmten Zeitpunkt abhängt.

5.7 Die Ermittlung einer elementaren Wahrscheinlichkeit

Eine elementare Wahrscheinlichkeit ist eine praktische Wahrscheinlichkeit, bei deren Bestimmung nicht auf die Wahrscheinlichkeiten anderer Aussagen zurückgegriffen wird. Erst später werden wir uns mit Wahrscheinlichkeiten von Kombinationen mehrerer Aussagen befassen.

Die Wahrscheinlichkeit ist von der Einstellung der bewertenden Person abhängig. Der Einfachheit halber bin ich in diesem Kapitel die bewertende Person, es geht also immer um die durch mich bestimmte Wahrscheinlichkeit der Aussage A.

Im Folgenden werden verschiedene Quellen für eine Wahrscheinlichkeitsbewertung beschrieben.

5.7.1 Das „Bauchgefühl"

Basis für meine Bestimmung der Wahrscheinlichkeit ist mein Bauchgefühl.

Liest man etwas über Wahrscheinlichkeiten in der Tageszeitung, so werden die dargestellten Wahrscheinlichkeiten oft nicht näher begründet. Das ist ganz normal, schließlich entscheidet man selbst auch täglich bewusst oder unbewusst nach Bauchgefühl. Siehe auch Beispiel (5.3.-03).

5.7.2 Einschätzungen anderer Personen

Basis für meine Bestimmung der Wahrscheinlichkeit ist die Einschätzung einer oder mehrerer anderer Personen.

Fall 1: Bekannte, Experten
Eine Person, der ich vertraue, kann ein Bekannter oder ein Experte sein. Die Wahrscheinlichkeitsbewertung dieser Person kann ich als Grundlage für meine Wahrscheinlichkeit nehmen. Die Bewertung dieser Person kann gut und transparent begründet oder auch nur ihr Bauchgefühl sein.
Beispiele:

- Bei den Expertenmeinungen zum Grexit wurden gelegentlich auch konkrete Wahrscheinlichkeiten genannt.
- In der Braunschweiger Zeitung vom 22.05.2015 wurden vor dem letzten Spieltag der Fußballbundesliga Wahrscheinlichkeiten für den Abstieg der einzelnen Mannschaften genannt und kurz begründet. Die Summe aller Wahrscheinlichkeiten (im Artikel „Abstiegsgefahr“ genannt) betrug 300%. Tatsächlich stiegen dann zwei Mannschaften ab, deren Abstiegsgefahr mit 95% bzw. 30% bewertet worden war. Siehe Beispiel (5.10-07).

Fall 2: Schwarmintelligenz (Umfrage)
Bei manchen Fragestellungen gibt es Umfragen oder Abstimmungen, deren Ergebnis ich nutzen kann, z. B. vor der Volksabstimmung zum Brexit am 23.06.2016. Dabei werden relative Häufigkeiten als Wahrscheinlichkeiten benutzt (Näheres dazu in den nächsten Kapiteln):

- Variante 1:
  Fragt man „Glauben Sie, dass am 23.06. mehrheitlich für den Brexit gestimmt wird?“, so kann ich die relative Häufigkeit derer, die mit „ja“ antworten, als Wahrscheinlichkeit nehmen.
- Variante 2:
  Fragt man „Wie werden Sie selbst am 23.06. über den Brexit abstimmen?“, so kann ich die relative Häufigkeit derer, die mit „für den Brexit“ antworten, als Wahrscheinlichkeit nehmen.

Manche Hersteller oder Dienstleister fragen ihre Kunden: „Wie hoch ist die Wahrscheinlichkeit, dass Sie unser Produkt / unsere Dienstleistung weiterempfehlen?“. Der daraus vom Hersteller oder Dienstleister ermittelte Durchschnittswert über alle Kunden wird als die Wahrscheinlichkeit publiziert, die ich dann übernehmen kann.

Fall 3: Schwarmintelligenz (Wettquote)
Der Kehrwert einer Wettquote kann auch als Wahrscheinlichkeit interpretiert werden.
Ist z. B. die Wettquote auf eine bestimmte Aussage 4 : 1, so kann man das so interpretieren, dass die Wahrscheinlichkeit der Aussage gerade $P(A) = 0{,}25$ ist, also der Kehrwert. Setzt man 1 € auf die Richtigkeit der Aussage, so bekommt man 4 € vom Buchmacher zurück, wenn die Aussage richtig ist, man macht also 3 € Gewinn. In der Realität bekommt man aber natürlich weniger, da der Buchmacher Gewinn machen will.

Da aber auch psychologische Faktoren wie die individuelle Risikobereitschaft eine Rolle spielen, kann es durchaus sein, dass Wahrscheinlichkeit und Wettquote nicht miteinander in Einklang stehen, dass man also bei „es wird eine 3 gewürfelt" eine Quote von 4 : 1 akzeptiert, weil man glaubt, eine Glückssträhne zu haben, obwohl man weiß, dass das nur eine von 6 Möglichkeiten ist.

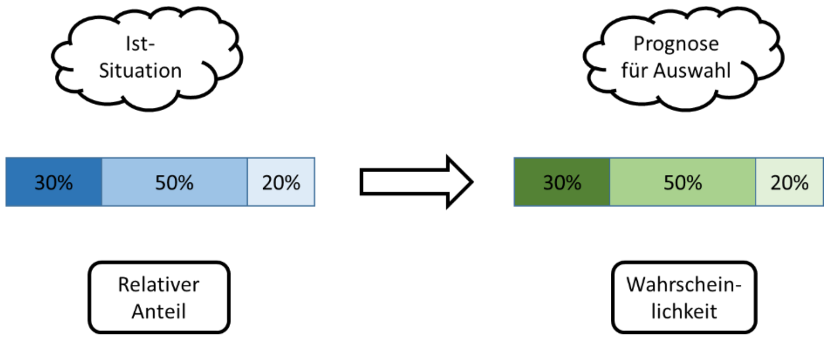
Beim Brexit führte dieses Verfahren zu deutlichen Fehleinschätzungen: am Tag vor der Abstimmung lag die Wahrscheinlichkeit, dass für den Brexit gestimmt wird, bei einem großen Wettanbieter bei 14 %, das Abstimmungsergebnis am 23.06. lag dann aber bei 52 %. Im Internet findet man aus der Zeit vor der Abstimmung viele verschiedene Darstellungen zur Wahrscheinlichkeit, dass für den Brexit gestimmt wird. Siehe Beispiel (5.10-08).

5.7.3 Relativer Anteil (Auswahl)

Basis für meine Bestimmung der Wahrscheinlichkeit ist der relative Anteil aufgrund theoretischer Überlegungen.

Voraussetzung ist, dass man eine Grundgesamtheit hat, bei der der relative Anteil einer bestimmten Eigenschaft bekannt ist. Aus dieser Grundgesamtheit wird eine **Auswahl** vorgenommen.
Es ist plausibel, dass die Wahrscheinlichkeit, dass ein aus dieser Gesamtheit ausgewähltes Objekt diese bestimmte Eigenschaft hat, gleich dem relativen Anteil ist, mit der diese Eigenschaft in der Gesamtheit vorkommt. Je häufiger etwas in der Grundgesamtheit vorkommt, umso eher erwarte ich, dass ich bei einer Auswahl ein Element mir dieser Eigenschaft erhalte. Der Begriff „relativer Anteil" muss aber im Einzelfall genau geklärt werden.

Wenn man annimmt, dass bei der Auswahl kein Element bevorzugt wird, so spricht man auch von einer „fairen" oder „idealen" Auswahl. In der Praxis ist es aber oft schwer nachzuweisen, dass ein idealer Auswahlprozess tatsächlich vorliegt. Oft kann man nur sagen: mir ist nichts bekannt, was dagegen spricht und für Wahrscheinlichkeitsbetrachtungen reicht das auch aus, denn es geht ja um meine subjektive Einschätzung.



(5.7.3.-1) ***Beispiel***
Einmaliges Würfeln mit einem Würfel.
Aussage A = „Mein Spielpartner würfelt eine „6" in einem Würfelspiel".

Der relative Anteil wird ermittelt, indem die verschiedenen Möglichkeiten bewertet werden.

Ich mache ein Würfelspiel mit einem Spielpartner. Da ich die Erfahrung gemacht habe, dass ich bei Glücksspielen oft betrogen wurde, unterstelle ich auch hier, dass ich betrogen werde. Wenn in diesem Glücksspiel also z. B. eine „6“ für einen Spieler besonders gut und eine „1“ besonders schlecht ist, wähle ich zum Beispiel folgende Gewichtung:

Möglichkeit	**Auswahl**	**Gewicht**
A1: Es kann eine „1“ gewürfelt werden	0	0,8
A2: Es kann eine „2“ gewürfelt werden	0	1
A3: Es kann eine „3“ gewürfelt werden	0	1
A4: Es kann eine „4“ gewürfelt werden	0	1
A5: Es kann eine „5“ gewürfelt werden	0	1
A6: Es kann eine „6“ gewürfelt werden	1	1,2
A7: der Wurf kann ungültig sein (Würfel liegt auf einer Kante oder fällt vom Tisch)	0	0,25

Die Gewichte 0,8 und 1,2 sind ein Kompromiss zwischen „der Betrug soll sich lohnen“ und „der Betrug soll nicht auffallen“. Natürlich kann man auch andere Gewichte wählen, sie sind nicht eindeutig definiert. Bei „Auswahl“ werden die betrachteten Aussagen mit „1“ gekennzeichnet, alle anderen erhalten eine „0“.
Insgesamt ergibt sich dann als relativer Anteil das gewichtete arithmetische Mittel von „Auswahl“:

$$P(\text{„Partner würfelt eine 6“}) = \frac{0{,}8\cdot 0+4\cdot 1\cdot 0+1{,}2\cdot 1+0{,}25\cdot 0}{0{,}8+4\cdot 1+1{,}2+0{,}25} = \frac{1{,}2}{6{,}25} = 19{,}2\ \%.$$

Wenn ich würfle, so sind die Gewichte von „1“ und „6“ zu vertauschen (denn ich glaube ja, betrogen zu werden), sodass sich ergibt:

$$P(\text{„ich würfle eine 6“}) = \frac{1{,}2\cdot 0+4\cdot 1\cdot 0+1\cdot 0{,}8+0{,}25\cdot 0}{1{,}2+4\cdot 1+0{,}8+0{,}25} = \frac{0{,}8}{6{,}25} = 12{,}8\ \%.$$

(5.7.3.-2) ***Beispiel***
Wie im vorangegangenen Beispiel mit den Änderungen:
1. Es werden nur gültige Würfe gewertet, Zeile A7 entfällt.
2. Da ich die Erfahrung gemacht habe, dass es bei Glücksspielen fair zugeht, unterstelle ich auch hier, dass es fair zugeht. „Fair“ bedeutet in diesem Zusammenhang, dass die Gewichtung aller Augenzahlen in den Zeilen A1–A6 gleich ist. Aus dem gewichteten arithmetischen Mittel wird einfach das arithmetische Mittel.
Dann ist

$$P(\text{„Partner würfelt eine 6“}) = P(\text{„ich würfle eine 6“}) = \frac{5\cdot 1\cdot 0+1\cdot 1}{6} = \frac{1}{6}$$

Diesen Wert erhält man auch, wenn man berücksichtigt, dass die relative Häufigkeit jeder Augenzahl auf den Würfelseiten = $\frac{1}{6}$ ist.

Unter diesen Annahmen kann man also einfach die relative Häufigkeit der jeweiligen Augenzahl auf der Würfeloberfläche als Wahrscheinlichkeit nehmen (klassische Definition der Wahrscheinlichkeit nach Laplace, siehe Kapitel 6.3).
Oder gleichwertig: man schreibt alle möglichen Ergebnisse auf und ermittelt dann die relative Häufigkeit des betrachteten Ereignisses.

(5.7.3.-3) ***Beispiel***
Eine Torte wird in 12 Stücke mit gleichem Volumen geteilt. In der Torte ist eine Kirsche versteckt, ich habe aber keinerlei Hinweise darauf, wo sie versteckt ist. Wie beim Würfelbeispiel könnte ich auch hier eine Tabelle mit 12 Zeilen erstellen, in jeder Zeile steht „Kirsche ist in Tortenstück Nummer …“

und die Gewichtung ist jeweils 1. Die Wahrscheinlichkeit, dass diese Kirsche in einem bestimmten Tortenstück verborgen ist, ist dann = $\frac{1}{12}$ = 8,3 % und das ist genau der Volumenanteil. Wenn aber ein Tortenstück etwas größer wäre, nämlich z. B. 9,1 % des Gesamtvolumens hätte, so wäre die Wahrscheinlichkeit, dass die Kirsche in diesem Tortenstück verborgen ist, gleich 9,1 %.

(5.7.3.-4) ***Beispiel***
Auf einem großen Platz sind fast alle Parkplätze besetzt, die Parkplätze sind gleichmäßig verteilt. Durch die Mitte des Parkplatzes verläuft eine Straße, so dass beide Teile des Parkplatzes dieselbe Fläche haben. Ich habe einen Autoschlüssel gefunden und suche das passende Auto. Am Autoschlüssel sind keinerlei Hinweise, um welches Auto es sich handelt und es ist ein rein mechanischer Schlüssel, also ohne Funk. Um die Wahrscheinlichkeit zu bestimmen, dass das gesuchte Auto auf dem linken Teil des Platzes parkt, kann man folgendermaßen vorgehen:

Variante 1: man zählt alle Autos und kommt zu dem Ergebnis: auf dem linken Teil parken 156 Autos, auf dem rechten Teil 157 Autos. Also ist die Wahrscheinlichkeit, dass das gesuchte Auto links steht, gleich

$\frac{156}{313}$ = 49,84 %.

Variante 2: man stellt fest, dass die beiden Parkplatzteile gleich groß sind und dass aus Symmetriegründen auf beiden Parkplatzhälften gleich viele Parkplätze vorhanden sind. Da fast alle Parkplätze belegt sind, ergibt sich näherungsweise, dass das gesuchte Auto mit 50 % Wahrscheinlichkeit auf der linken Hälfte steht.

In Variante 1 hat man als exakten Anteil die relative Häufigkeit, während man in Variante 2 näherungsweise den Flächenanteil zur Bestimmung der Wahrscheinlichkeit nutzt. Variante 1 ist genauer, aber aufwändiger als Variante 2.

(5.7.3.-5) ***Beispiel***
Hat man eine Urne mit 10 Kugeln, von denen 7 rot und 3 schwarz sind, so sind die relativen Häufigkeiten:

f(„Kugel ist rot“) = 0,7
f(„Kugel ist schwarz“) = 0,3

Zieht man eine Kugel, wobei keine Hinweise darauf vorliegen, dass dabei eine Kugel bevorzugt wird, so ist plausibel, dass man eher eine rote als eine schwarze Kugel erhält. Es ist also wieder sinnvoll, die relativen Häufigkeiten als Wahrscheinlichkeiten zu nehmen:

P(„gezogene Kugel ist rot“) = 0,7
P(„gezogene Kugel ist schwarz“) = 0,3

Auch bei der Ziehung von Kugeln kann es zu Unregelmäßigkeiten kommen, die man als Außenstehender nicht wahrnimmt. Zum Beispiel könnten die Kugeln unterschiedlich vorgewärmt sein, siehe z. B. [SZ].

(5.7.3.-6) ***Beispiel***
Eine Münze hat zwei Seiten, jede Seite hat also die relative Häufigkeit = $\frac{1}{2}$. Die Auswahl der Seite erfolgt durch einen Münzwurf. Weiß man nichts Zusätzliches über den Münzwurf und ignoriert insbesondere die Tatsache, dass die beiden Seiten der Münze eine unterschiedliche Oberfläche haben, so ist:

P(„es wurde Wappen geworfen“) = P(„es wurde Zahl geworfen“) = $\frac{1}{2}$.

Dass es in der Realität nicht so ist, sondern tatsächlich 3 Möglichkeiten bestehen, zeigt der Münzwurf von Rotterdam, bei dem die Münze im weichen Boden auf der Kante stecken blieb.

(5.7.3.-7) ***Beispiel***
Ein Bus fährt alle 15 Minuten ab einer bestimmten Haltestelle, man hat aber vergessen, um welche Uhrzeiten. Man geht zu einem beliebigen Zeitpunkt von zu Hause los und hat keine Möglichkeit, den Bus vor der Ankunft zu sehen (und schnell zu laufen), da die Haltestelle direkt hinter einer Ecke liegt. Wie groß ist die Wahrscheinlichkeit, dass man an der Haltestelle höchstens 10 Minuten auf den Bus warten muss? Unter der Annahme, dass man keine weiteren Informationen hat, ergibt sich:

Variante 1: Zeitanteil (bei minutengenauer Messung)

P(„ich muss höchstens 1 Minute warten") = $\frac{1}{15}$

…

P(„ich muss höchstens k Minuten warten") = $\frac{k}{15}$

…

P(„ich muss höchstens 15 Minuten warten") = $\frac{15}{15} = 1$

Die Wahrscheinlichkeit ergibt sich also aus dem relativen Zeitanteil.

Variante 2: relative Häufigkeit (bei sekundengenauer Messung)
Ich schreibe alle möglichen Wartezeiten auf und bestimme die relative Häufigkeit der Zeiten von höchstens 10 Minuten.
Es gibt 600 Zeiten mit höchstens 10 min: 0:01; 0:02; …; 9:59; 10:00
Es gibt 300 Zeiten mit mehr als 10 min: 10:01; 10:02; …; 14:59; 15:00
Also ist

P(„ich muss höchstens 10 Minuten warten") = $\frac{600}{900} = \frac{2}{3}$

Die Wahrscheinlichkeit ergibt sich also aus einer relativen Häufigkeit.

(5.7.3.-8) ***Beispiel***
Dass es nicht nur beim Münzwurf (Beispiel (5.7.3.-6)), sondern auch bei der Ziehung der Lottozahlen vorkommen kann, dass die üblichen idealisierenden Annahmen falsch sind, zeigt die Ziehung der Lottozahlen am 03.04.2013, bei der 2 Kugeln nicht in der Trommel landeten. Der Fehler wurde zunächst nicht bemerkt, aber später musste die Auslosung dann wiederholt werden.

(5.7.3.-9) ***Beispiel***
Eine Reißzwecke hat auf harten ebenen Boden zwei mögliche Lagen: entweder zeigt die Spitze nach oben oder schräg nach unten. Die Auswahl der beiden Lagen erfolgt wie üblich durch einen Wurf. Wenn man sich keine weiteren Gedanken macht, sind die Wahrscheinlichkeiten für die beiden Positionen entsprechend den relativen Häufigkeiten jeweils 50 %.
Aufgrund der ausgeprägten Asymmetrie kann man aber auch andere Wahrscheinlichkeiten nehmen.

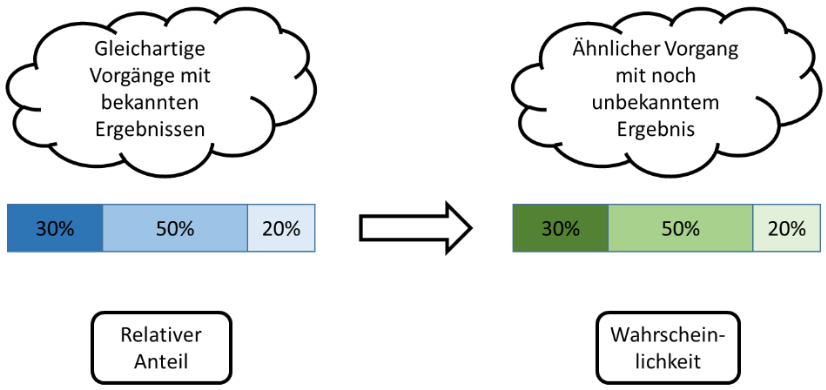
(5.7.3.-10) ***Definition***
Die relativen Anteile wurden, bevor man eine praktische Untersuchung durchgeführt hat, nur aufgrund von theoretischen Überlegungen ermittelt. Man nennt die daraus abgeleiteten Wahrscheinlichkeiten **a-priori-Wahrscheinlichkeiten**.

In allen Fällen wurde ein relativer Anteil als Wahrscheinlichkeit interpretiert.

5.7.4 Relativer Anteil (Erfahrung)

Basis für meine Bestimmung der Wahrscheinlichkeit ist meine Erfahrung aufgrund vieler ähnlicher Fälle.

Die relativen Häufigkeiten der Ergebnisse vieler gleichartiger Vorgänge können als Grundlage für die Wahrscheinlichkeit für das noch unbekannte Ergebnis eines weiteren solchen Vorganges genommen werden. Unter der Annahme, dass bestimmte Bedingungen und Abläufe gleichbleiben, ist die Wahrscheinlichkeit für ein mögliches neues Ergebnis gleich der relativen Häufigkeit dieses Ergebnisses bei den bisherigen bekannten Vorgängen, man leitet die Wahrscheinlichkeit also aus der **Erfahrung** ab. Das ist eine praktische Anwendung des Gesetzes der großen Zahlen. Analoges gilt natürlich auch wieder für andere Anteile an Stelle der relativen Häufigkeiten.



(5.7.4.-1) ***Beispiel***
Würfelt man 600-mal und kommt dabei die Augenzahl „4" genau 107-mal vor, so ist die aus dieser Erfahrung abgeleitete Wahrscheinlichkeit, dass beim nächsten Wurf eine „4" gewürfelt wird, gleich der relativen Häufigkeit bei den bisherigen Würfen $= \frac{107}{600}$. Das Ergebnis unterscheidet sich also von der a-priori-Wahrscheinlichkeit. Dabei wurde angenommen, dass bestimmte bisherige Versuchsbedingungen auch für den nächsten Wurf gelten. Welche Bedingungen gleichgeblieben sind und welche variabel sind, ist in der Praxis nicht immer einfach zu bestimmen. Sicher ist nur, dass es variable Bedingungen gibt, denn sonst wäre das Ergebnis bei jedem Wurf gleich.

(5.7.4.-2) ***Beispiel***
Ein Antibiotikum hat im Durchschnitt der letzten 5 Jahre in 80 % der Fälle bei Krankheit X geholfen. Also wird es auch jetzt bei einem Kranken mit der Krankheit X mit 80 %-iger Wahrscheinlichkeit helfen. Tatsächlich kann es aber sein, dass aufgrund von Resistenzen die Wahrscheinlichkeit inzwischen geringer geworden ist, also vor 5 Jahren 90 % betrug, aber jetzt nur noch 70 % beträgt. Die 80 % sind

also nur ein Durchschnittswert über die letzten 5 Jahre. Aussagekräftiger wäre eine Regressionsfunktion für die beiden Merkmale „Zeitabschnitt“ (z. B. Quartal) und „relative Häufigkeit der Wirksamkeit des Medikamentes“.

(5.7.4.-3) ***Beispiel***
Die relative Häufigkeit, dass ein Mädchen geboren wurde, beträgt im Durchschnitt der letzten 5 Jahre in einem bestimmten Gebiet 48,6 %. Also ist die Wahrscheinlichkeit, dass bei der nächsten Geburt ein Mädchen geboren wird, ebenfalls 48,6 %. Tatsächlich kann es aber sein, dass sich in den letzten Jahren der Anteil der Mädchen bei den Neugeborenen aufgrund von Umweltfaktoren oder weil Frauen heute später schwanger werden als früher oder aus anderen Gründen verändert hat, so dass der Durchschnittswert 48,6 % jetzt nicht mehr gilt. Auch hier wäre eine Regressionsfunktion aussagekräftiger.

(5.7.4.-4) ***Beispiel***
Mannschaft A hat in den letzten 10 Jahren immer gegen Mannschaft B gewonnen. Daraus kann man schließen, dass Mannschaft A auch heute gegen Mannschaft B gewinnen wird (in der Sportberichterstattung vielzitiertes „Gesetz der Serie“). Dieser Schluss ist aber erfahrungsgemäß mit großen Unsicherheiten behaftet, weil es zu viele variable Bedingungen gibt.

(5.7.4.-5) ***Beispiel***
Ein Bauer betrachtet 95 % seines Feldes und stellt fest, dass das Korn sehr gut gewachsen ist, er erwartet also hohe Erträge. Lediglich 5 % des Feldes kann er nicht einsehen, da es in einer kleinen Senke liegt, aber da 95 % des Feldes einen hohen Ertrag erwarten lassen, geht er davon aus, dass das auch im restlichen 5 % so sein wird. Er hat dabei übersehen, dass das Korn zur Senke hin immer schlechter gewachsen ist und in der Senke selbst besonders schlecht ist. Der gute Ertrag für die einsehbaren 95 % ist nur ein Durchschnittswert, von dem die Realität immer mehr abweicht, je näher man der Senke kommt.

(5.7.4.-6) ***Beispiel***
Wahrscheinlichkeitsaussagen zum Brexit auf Basis von relativen Anteilen wurden bereits in Kapitel 5.7.2 besprochen.

(5.7.4.-7) ***Definition***
Da die Wahrscheinlichkeiten ermittelt wurden, nachdem man eine praktische Untersuchung durchgeführt hat, nennt man die daraus abgeleiteten Wahrscheinlichkeiten **a-posteriori-Wahrscheinlichkeiten**.

(5.7.4.-8) ***Hinweise***
Die Beispiele zeigen deutlich, dass die a-posteriori-Wahrscheinlichkeit problematisch sein kann, da man vom Durchschnitt der bisherigen Erfahrung auf die Zukunft schließt und eine eventuelle Entwicklung außer Acht lässt. Konstante Bedingungen kann man eventuell im Labor erzeugen, ist man dagegen nur passiver Beobachter einer Entwicklung, so muss man eher mit veränderlichen Bedingungen rechnen, so dass die Durchschnittserfahrung der Vergangenheit nicht als Wahrscheinlichkeit für die nächste Beobachtung taugt. Trotzdem werden solche Beispiele in vielen Veröffentlichungen ohne diese kritischen Anmerkungen dargestellt – also Vorsicht!

Aussagekräftiger wäre eine Regressionsfunktion. Die unabhängige Variable x der Regressionsfunktion repräsentiert die absolute Größe (Zeiten, Häufigkeiten, Flächen, …) und die abhängige Variable y die relativen Anteile des Ereignisses. Die gesuchte Wahrscheinlichkeit ist dann der y-Wert für den entsprechenden nächsten noch nicht beobachteten x-Wert (nächster Zeitpunkt, nächster Versuch, nächstes Flächenstück). Die Regressionsfunktion ist nicht eindeutig und hängt von der Fragestellung und der Erfahrung ab. Besonders einfache Fälle sind:

- Die empirische Regressionsgerade. Sie wird nach dem Prinzip der kleinsten Quadrate bestimmt.
- Eine waagerechte Gerade durch den letzten Punkt = Punkt mit dem größten x-Wert (Anwendung vom Gesetz der großen Zahlen). Es wird also einfach der relative Anteil des betrachteten Ereignisses an den bisherigen Beobachtungen, aber keine Entwicklungen innerhalb der Beobachtungen berücksichtigt.
  Die a-posteriori-Wahrscheinlichkeit ist also dann sinnvoll, wenn eine waagerechte Gerade (durch den letzten Punkt) als Regressionsfunktion angemessen ist.

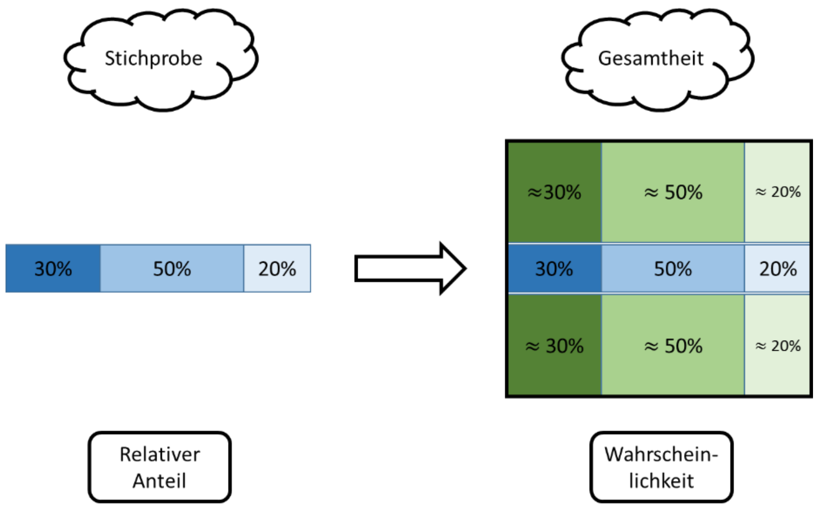
Vor plötzlichen massiven Änderungen ist man aber nie geschützt, wie die Bankenkrise 2008 oder die Truthahn-Illusion zeigen.

5.7.5 Relativer Anteil (Hochrechnung)

Basis für meine Bestimmung der Wahrscheinlichkeit ist eine Hochrechnung.

Voraussetzung ist, dass man eine Stichprobe hat, bei der der relative Anteil einer bestimmten Eigenschaft bekannt ist. Daraus wird dann mit einer **Hochrechnung** auf den relativen Anteil der Eigenschaft in der Gesamtheit geschlossen.

Im einfachsten Fall überträgt man die relative Häufigkeit der Eigenschaft aus der Stichprobe einfach im gleichen Verhältnis auf die Gesamtheit: wenn 27 % in der Stichprobe von 1000 Personen eine bestimmte Partei wählen, dann wählen auch ca. 27 % aus der Gesamtheit von 40 Millionen diese Partei. Eine solche einfache Schlussweise ist natürlich mit hohen Risiken behaftet, da die Stichprobe die Verhältnisse in der Gesamtheit völlig falsch widerspiegeln kann.



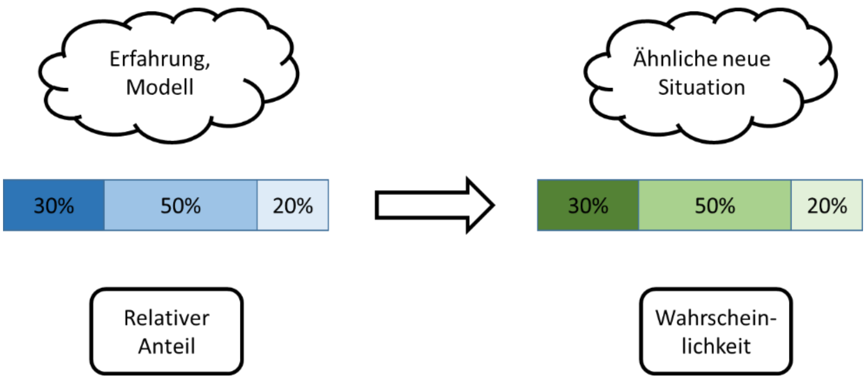
In der schließenden Statistik werden anspruchsvollere Methoden angewendet, indem man bestimmte Erfahrungen nutzt, wie z. B.:

- man wählt eine Stichprobe, die „repräsentativ" sein sollte (was aber nicht sicher feststellbar ist)
- man nimmt eine bestimmte Art der Datenverteilung in der Gesamtheit an (z. B eine Normalverteilung) und nutzt dann die Stichprobe, um die Parameter der Verteilung zu schätzen

Auf dieser Basis kann man dann z. B. angeben, wie groß die Wahrscheinlichkeit ist, dass die Gesamtheit die Partei mit 25 % – 29 % der Stimmen wählt. Weiterführende Hinweise findet man in vielen Stochastikbüchern.

5.7.6 Allgemeines zu relativen Anteilen

(5.7.6.-1) ***Hinweis***
Es ist plausibel, dass das, was aus Erfahrung oder in einem realistischen Modell überwiegend – also mit einem hohen relativen Anteil – vorkommt, dann auch in einer neuen vergleichbaren Situation eher passiert. Es ist also plausibel, den relativen Anteil als Maß für die Wahrscheinlichkeit zu nehmen, sofern man über keine weiteren Informationen verfügt.
Wie man in den Beispielen (5.7.3.-1) und (5.7.3.-2) sieht, können unterschiedliche Einstellungen der bewertenden Personen zu unterschiedlichen Gewichtungen und damit zu unterschiedlichen Wahrscheinlichkeiten führen. In diesem Sinne sind auch relative Anteile subjektiv.



Der relative Anteil kann sich dabei auf eine Häufigkeit, Fläche, Gewicht, Zeit, etc. beziehen. Wahrscheinlichkeiten, die sich als relative Anteile von Strecken, Flächen oder Volumina darstellen lassen, nennt man auch **geometrische Wahrscheinlichkeiten**.

(5.7.6.-2) ***Hinweis***
In den vorangegangenen Kapiteln haben wir verschiedene Methoden kennengelernt, wie man aus relativen Anteilen Wahrscheinlichkeiten ableiten kann. Dabei ist die Aussagekraft der Ergebnisse recht unterschiedlich:

1. Auswahl

Die Ungewissheit besteht darin, inwieweit der Auswahlprozess den Annahmen entspricht, ob der Prozess also z. B. tatsächlich „fair“ ist, also kein Objekt bei der Auswahl bevorzugt wird. Wenn eine bestimmte Eigenschaft in den bekannten Fällen einen Anteil von 27 % hatte, dann ist die Wahrscheinlichkeit, dass ein daraus zufällig ausgewählter Fall diese Eigenschaft hat, ebenfalls 27 %.

2. Erfahrung
   Im einfachsten Fall nimmt man den Durchschnitt der bisherigen Erfahrungen, eine zeitliche oder räumliche Entwicklung wird nicht berücksichtigt. Wenn eine bestimmte Eigenschaft in den bekannten 999 Fällen einen Anteil von 27 % hatte, dann ist die Wahrscheinlichkeit, dass im 1000. Fall diese Eigenschaft auftritt, ebenfalls 27 %.
   Die Aussagekraft wird erhöht, wenn man stattdessen die zeitliche oder räumliche Entwicklung berücksichtigt und eine geeignete Regressionsfunktion benutzt.

3. Hochrechnung
   „Hochrechnung“ ist eine Verallgemeinerung von „Erfahrung“: Wenn eine bestimmte Eigenschaft in den bekannten 999 Fällen einen Anteil von 27 % hatte, dann kommt diese Eigenschaft in der Gesamtheit von 40 Millionen ebenfalls zu 27 % vor. Oder anders formuliert: die Wahrscheinlichkeit, dass ein Element der Gesamtheit diese Eigenschaft hat, ist ebenfalls 27 %. Wie im vorangegangenen Kapitel bereits dargestellt wurde, ist diese Methode spekulativ. Mit zusätzlichen Erfahrungen und Methoden der schließenden Statistik lassen sich aber seriösere Aussagen machen.

(5.7.6.-3) ***Hinweis***
Anhand der Beispiele in den vorangegangenen Kapiteln sieht man, dass es zwei Varianten für a-priori und a-posteriori gibt:

Variante 1: eine Fragestellung, zwei Antworten

Frage:	Wie groß ist die Wahrscheinlichkeit, beim nächsten Wurf eine 4 zu würfeln?
Antwort a-priori:	Wahrscheinlichkeit = relative Häufigkeit auf den 6 Würfelseiten = $\frac{1}{6}$
Antwort a-posteriori:	Wahrscheinlichkeit = relative Häufigkeit bei den letzten 600 Würfen = $\frac{107}{600}$.

Variante 2: zwei Fragestellungen, eine Antwort

Antwort:	Die Wahrscheinlichkeit für ein Mädchen beträgt 48,6 %.
Frage a-priori:	Wie groß ist die Wahrscheinlichkeit, dass ein aus der Menge der geborenen Kinder ausgewähltes Kind ein Mädchen ist?
Frage a-posteriori:	Wie groß ist die Wahrscheinlichkeit, dass bei der nächsten Geburt ein Mädchen geboren wird?

(5.7.6.-4) ***Hinweis***
Manchmal wird Wahrscheinlichkeit auch als der Wert definiert, gegen den die relativen Häufigkeiten streben, wenn man den Versuch unendlich oft unter gleichen Bedingungen wiederholt. Das ist eine ungenaue Interpretation des Gesetzes der großen Zahlen. In vielen Büchern wird darauf hingewiesen, dass eine solche Interpretation aus mehreren Gründen problematisch ist:

- Man kann einen Versuch nicht unendlich oft wiederholen, irgendwann ist Schluss.
- Auch nach sehr vielen Versuchen gibt es immer noch Schwankungen in der relativen Häufigkeit, es gibt also keinen eindeutigen Grenzwert.

- Es ist fragwürdig, ob es gelingt, bei einer sehr großen Anzahl von Versuchen die relevanten Rahmenbedingungen konstant zu halten. Z. B. nutzt sich ein Würfel langsam ab.
- Es ist schwer zu bestimmen, was „konstante Bedingungen“ konkret sind. Am Beispiel der Geschlechter von Neugeborenen sieht man, dass nicht vollständig klar ist, welche Bedingungen konstant sind und welche nicht, bzw. welche für das Geschlecht relevant sind und welche nicht.
- Der Ansatz suggeriert, dass von Anfang an eine tatsächliche oder objektive Wahrscheinlichkeit existiert, die man leider nicht kennt und deshalb viele Versuche macht, um sich ihr zu nähern. Aber entweder hat man eine Wahrscheinlichkeit bereits aufgrund einer a-priori-Betrachtung – dann dienen die Versuche höchstens zu Plausibilisierung – oder man hat keine a-priori-Wahrscheinlichkeit, dann wird die Wahrscheinlichkeit erst durch die Versuchsreihe erzeugt, nämlich als relative Häufigkeit nach dem letzten Versuch.
- Fälle, die nur einmal vorkommen, wie ein eventueller Grexit, sind durch diese Definition nicht abgedeckt.

(5.7.6.-5) ***<u>Hinweis</u>***
Der Ablauf in allen Fällen im Kapitel 5.7.6 ist also:
1. Genaue Formulierung der Aussage, deren Wahrscheinlichkeit zu bestimmen ist
2. Plausible Ableitung, welche Art von relativem Anteil (relative Häufigkeit, relativer Flächenanteil, …) die Wahrscheinlichkeit repräsentiert
3. Konkrete Bestimmung des relativen Anteils
4. Interpretation des relativen Anteils als Wahrscheinlichkeit.

Mathematische Verfahren finden nur im 3. Schritt statt, alles andere ist Interpretation. Die üblichen Beispiele zur Bestimmung einer Wahrscheinlichkeit sind also tatsächlich einfach nur Beispiele für die Bestimmung eines relativen Anteils, der dann als Wahrscheinlichkeit interpretiert wird: Wahrscheinlichkeit von Ergebnissen bei Glücksspielen wie Münzwurf, Würfeln, Lotto oder Fragen zur Wahrscheinlichkeit einer bestimmten Krankheit oder des Geschlechts von Neugeborenen oder … – rechnerisch geht es dabei immer nur um relative Anteile.
Es bleibt die Frage, ob es in der Schule oder in den Statistikbüchern überhaupt Fälle gibt, bei denen die elementaren Wahrscheinlichkeiten nicht einfach nur interpretierte relative Anteile sind. Wenn das nicht so ist, so könnte man alle Darstellungen und Aufgaben so umformulieren, dass der Begriff Wahrscheinlichkeit durch den entsprechenden relativen Anteil ersetzt wird, der Begriff Wahrscheinlichkeit kommt dann gar nicht mehr vor. In diesen Fällen ergibt sich dasselbe mathematische Ergebnis, nur die Interpretation ist etwas unterschiedlich.

(5.7.6.-6) ***<u>Hinweis</u>***
In manchen Veröffentlichungen wird beschrieben, dass es verschiedene Interpretationen von Wahrscheinlichkeit gibt: einmal der frequentistische Ansatz, der besagt, dass eine relative Häufigkeit ein Maß für die Wahrscheinlichkeit ist und der subjektivistische Ansatz, der besagt, dass Wahrscheinlichkeit der Grad der Plausibilität sei. Dabei wird verkannt, dass das, was relativ häufig vorkommt, plausibler ist, als das, was selten vorkommt. Der frequentistische Ansatz ist also nur ein Sonderfall des subjektivistischen Ansatzes.
Zitat: „Wahrscheinlichkeiten werden also nicht nur als Häufigkeiten interpretiert, sie repräsentieren darüber hinaus die Plausibilität von Aussagen“ ([KK], Seite 1).
Besser wäre:
1. Umgekehrt: Relative Häufigkeiten werden als Wahrscheinlichkeiten interpretiert
2. Wahrscheinlichkeiten repräsentieren immer den Grad der Plausibilität, der oft durch relative Häufigkeiten ermittelt wird.

5.7.7 Eigene Wettquote

Die Wettquote eines Wettbüros wurde bereits unter Kapitel 5.7.2, Fall 3, besprochen. Analog kann man für sich selbst eine individuelle Wettquote bestimmen. Diese Wettquote dann auch durch ein Bauchgefühl oder relative Anteile oder andere Begründungen zustande kommen. Der Kehrwert der eigenen Wettquote kann dann als Wahrscheinlichkeit bezeichnet werden (siehe z. B. auch [SJ], Kapitel 8.5). Im Allgemeinen ist die eigene Wettquote aber nur ein Teilaspekt bei der Bestimmung einer Wahrscheinlichkeit.

Dies wird im nächsten Kapitel ausführlich erläutert.

5.8 Die konsolidierte elementare Wahrscheinlichkeit

In Kapitel 4 wurde dargestellt, nach welcher Methode man eine quantitative Bewertung von „etwas" durchführen kann. Nach demselben Muster kann man auch Wahrscheinlichkeiten ermitteln, wobei auch hier sowohl exakt messbare oder objektive, als auch eher vage oder subjektive Faktoren einfließen können. Eine Wahrscheinlichkeit kann also ähnlich ungenau und subjektiv sein, wie der Glücksindex im World Happiness Report, die Bewertung von Statistikkursen durch Studierende, der Erfüllungsgrad einer Zielvereinbarung in Unternehmen oder das Qualitätsurteil über Mozzarella.

- Die Ermittlung einer Wahrscheinlichkeit ist in der Praxis nicht so exakt und objektiv wie das Messen einer Fläche oder die Angabe eines Kontostandes

- Es gibt in der Praxis keine objektive oder tatsächliche Wahrscheinlichkeit, diese gibt es höchstens bei idealisierten Gedankenexperimenten

- Es gibt keinen allgemeinen Algorithmus oder „robot" (siehe [JE]), der zu einer beliebigen Aussage die Wahrscheinlichkeit ermittelt.

Da es kein allgemeines exakt vorgegebenes Verfahren zur Bestimmung einer Wahrscheinlichkeit gibt, sollte bei jeder Wahrscheinlichkeit die Herleitung transparent gemacht werden.

Leider gibt es im Gegensatz zu den Bewertungen von Smartphones, Winterreifen oder Mozzarella kein Buch und keine Fachzeitschrift, in der eine allgemeingültige Bewertungstabelle zur Bestimmung der Wahrscheinlichkeit einer beliebigen Aussage dargestellt wird. Wenn man aber bedenkt, wie unterschiedlich Aussagen sein können, dann ist verständlich, dass es die einheitliche Bewertungstabelle nicht geben kann. Es gibt ja schließlich auch nicht die übergreifende Bewertungstabelle, mit der man beliebige Geräte wie Smartphones, Kameras und Musikanlagen bewerten kann, sondern nur individuelle Bewertungstabellen. Übergreifend ist nur die allgemeine Methodik.

Analog zu Kapitel 4 nimmt man eine Bewertungstabelle, in die man alle Wahrscheinlichkeiten aus Kapitel 5.7 zu dieser Aussage aufnimmt und daraus ein gewichtetes arithmetisches Mittel bildet. Wie auch bei beliebigen Bewertungen in Kapitel 4 können unterschiedliche Personen bei derselben Aussage zu unterschiedlichen Wahrscheinlichkeiten kommen, da sie

a. unterschiedliche Einflussfaktoren auswählen
b. bei den ausgewählten Faktoren unterschiedliche Wahrscheinlichkeiten ermitteln
c. die ausgewählten Wahrscheinlichkeiten unterschiedlich für den endgültigen Wert gewichten

Darüber hinaus können sich alle drei Aspekte im Laufe der Zeit verändern, da die bewertende Person weitere Erfahrungen macht, die zu einer anderen Bewertung führen.

Auch hieraus folgt wieder: bei der Angabe einer Wahrscheinlichkeit muss der Weg zur Ermittlung der Wahrscheinlichkeit transparent sein – oder man glaubt der bewertenden Person einfach.

(5.8.-01) ***<u>Verfahren zur Bestimmung einer elementaren Wahrscheinlichkeit</u>***

<u>Schritt 1:</u>
Man definiert möglichst genau, welche Aussage A man bezüglich der Wahrheit bewerten will.

<u>Schritt 2:</u>
Bestimmung der Einflussfaktoren. Die möglichen Einflussfaktoren sind gemäß Kapitel 5.7:

- Bauchgefühl, keine oder nur schwache Begründung
- Einschätzung anderer Personen
    - Einzelpersonen, wie Experten oder Bekannte
    - Schwarmintelligenz: Umfragen
    - Schwarmintelligenz: Kehrwert einer Wettquote
- Relative Anteile auf Basis von
    - Auswahl (a priori)
    - Erfahrung (a posteriori)
    - Hochrechnung
- Kehrwert der eigenen Wettquote
- Ggf. weitere Faktoren

Zu jedem Einflussfaktor wird die Wahrscheinlichkeit ermittelt.

<u>Schritt 3:</u>
Für jeden Einflussfaktor wird ein Gewicht in Form einer positiven Zahl festgelegt. Formal:
Hat man (i = 1, …, n)

- einen Einflussfaktor,
- die zugehörige Wahrscheinlichkeitsbewertung P_i mit $0 \leq P_i(A) \leq 1$ und
- das zugehörige Gewicht $c_i > 0$,

so ist $c_i \cdot P_i(A)$ das gewichtete Bewertungsergebnis der Aussage A bezüglich des i-ten Einflussfaktors.

<u>Schritt 4:</u>
Die gesuchte Wahrscheinlichkeit ist das gewichtete arithmetische Mittel der Bewertungen der Einflussfaktoren und liegt damit im Intervall [0; 1], also:

$$P(A) = \frac{\sum_{i=1}^{n} c_i \cdot P_i(A)}{\sum_{i=1}^{n} c_i} = \sum_{i=1}^{n} d_i \cdot P_i(A)$$

mit den relativen Gewichten

$$d_i = \frac{c_i}{\sum_{j=1}^{n} c_j}$$

Die Abhängigkeit von Rahmenbedingungen sieht man besonders deutlich am Beispiel der Wetten auf den Brexit. Die Wahrscheinlichkeit des Brexit unterlag ständigen kurzfristigen Veränderungen. Da die genauen Rahmenbedingungen mit ihren Veränderungen nicht oder nur mühselig ermittelbar sind, muss in diesem Fall die Angabe der Wahrscheinlichkeit zumindest mit der Angabe des Zeitpunkts ergänzt werden. Die Aussage „die Wahrscheinlichkeit des Brexit beträgt x %“ ist ohne Angabe des Zeitpunktes und der Bewertungsmethode sinnlos.

(5.8.-02) ***Beispiel***
Beim Würfeln hat man sowieso immer eine a-priori-Wahrscheinlichkeit und oft auch eine a-posteriori-Wahrscheinlichkeit – aber für welche soll man sich entscheiden? Da man nicht weiß, ob der Würfelprozess wirklich absolut ideal ist, kann man zur Ermittlung der Wahrscheinlichkeit, beim nächsten Wurf eine bestimmte Augenzahl zu würfeln, z. B. das arithmetische Mittel von a-priori- und a-posteriori-Wahrscheinlichkeit nehmen.

(5.8.-03) ***Beispiel***
Beim Brexit vor der Volksabstimmung am 23.06.2016 konnte man ein gewichtetes arithmetisches Mittel aus Ergebnissen von Expertenmeinungen, Umfragen und Kehrwerten von Wettquoten bilden. Das Gewicht bei den Wahrscheinlichkeiten von Experten ergibt sich aus ihrer Glaubwürdigkeit und der Aktualität ihrer Aussage.

(5.8.-04) ***Beispiel***
Bislang war für eine bestimmte Aussage A die relative Häufigkeit f(A) = 0,4 und damit die a-posteriori-Wahrscheinlichkeit $P_1(A) = 0{,}4$. Da sich die Bedingungen (neue Gesetze, neue Technik, …) geändert haben, wird mittelfristig $P_2(A) = 0{,}8$ erwartet. Da man sich mitten im Umbruch befindet, ist jetzt

$$P(A) = \frac{1}{2}\cdot P_1(A) + \frac{1}{2}\cdot P_2(A) = 0{,}6$$

(5.8.-05) ***Hinweis***
Manchmal ist zu lesen, dass man Wahrscheinlichkeiten von Aussagen nicht immer quantifizieren kann. Das oben angegeben Verfahren zeigt dagegen, dass das immer möglich ist, aber die Ergebnisse natürlich – wie bei jeder Bewertung – mit Vorsicht zu interpretieren sind, da sie eine Genauigkeit und Objektivität ausstrahlen, die tatsächlich nicht gegeben ist. Insofern bewegt man sich in einer Grauzone: manchmal kann es seriöser sein, nur Worte wie „hoch“ / „niedrig“ statt Zahlen zu benutzen, aber unmöglich sind Zahlen deshalb nicht.
Wenn man nichts über eine Aussage weiß, außer dass es sich um eine Aussage im Sinne von Kapitel 2.1.3 handelt, sie also nicht versteht (denn zum Verständnis gehört schon Wissen), so ist $P(A) = P(\overline{A}) = 0{,}5$ plausibel.

(5.8.-06) ***Hinweis***
Kennt man bei der Bestimmung einer Wahrscheinlichkeit Gründe, die für die Richtigkeit einer Aussage sprechen, aber keine Gründe, die dagegen sprechen, so kann man die Aussage als eher wahr einstufen. Wenn man die Gründe als schwerwiegend ansieht, so ist die Wahrscheinlichkeit sehr hoch, wenn die Gründe eher schwach sind, so wird man als Wahrscheinlichkeit, dass die Aussage wahr ist, nur knapp über 50% bestimmen. Diese Vorgehensweise wird bei der Subadditivität (Kapitel 5.9.1),

der bedingten Wahrscheinlichkeit (Kapitel 5.9.2) und dem Indifferenzprinzip (Kapitel 8.3) angewendet.

(5.8.-07) ***Hinweis***
Wichtig ist, dass im täglichen Leben die Ermittlung einer Wahrscheinlichkeit kein rein rationaler Prozess ist, der nur auf präsentem, bewusst nutzbarem Wissen und von jedem Menschen akzeptierten rein logischen Schlüssen beruht. Menschen sind keine deterministisch funktionierenden Roboter. Die Ermittlung einer Wahrscheinlichkeit ist keine Mathematikaufgabe, für die es zwar vielleicht verschiedene Wege, aber nur ein objektives, richtiges Ergebnis gibt. Sondern es fließen subjektive und objektive, bewusste und unbewusste, klare und schwammige Bewertungen in die Ermittlung einer Wahrscheinlichkeit ein. Und deswegen können in der Realität verschiedene Personen zu unterschiedlichen Ergebnissen kommen – auch wenn sie ausgewiesene, rationale denkende Experten sind, denen man mehr als nur gesunden Menschenverstand zubilligen würde.

Es entspricht nicht unbedingt der Erfahrung, dass zwei Experten nach entsprechender Diskussion über eine Fragestellung wie der Wahrscheinlichkeit eines Grexit zu demselben Wissen und zu derselben Bewertung des Wissens und damit letztlich zu derselben Wahrscheinlichkeit kommen.

Jetzt werden sich alle Studierenden freuen, denn die nächste Statistikklausur ist gerettet: wenn nach irgendwelchen Wahrscheinlichkeiten gefragt wird, dann schreibt man einfach eine Zahl zwischen 0 und 1 hin und erklärt, dass das das Bauchgefühl ist und das ist laut Herrn Stegen als Wahrscheinlichkeit erlaubt – einfacher und schneller kann man keine Klausur bestehen.
Aber zu früh gefreut: in Klausuren werden üblicherweise konkrete Rahmenbedingungen für die Aufgaben vorgegeben, die zu einer eindeutigen Wahrscheinlichkeit führen. Die Frage, ob diese Rahmenbedingungen in der Realität so vorkommen und ob bzw. wie man sie im Einzelfall nachweisen kann, wird dabei natürlich wohlweislich ausgeklammert, denn dann würde es brenzlig. Siehe auch Kapitel 8.2 über den „idealen“ Würfel.

5.9 Komplexe Wahrscheinlichkeiten

Eine komplexe Wahrscheinlichkeit ist eine Wahrscheinlichkeit, bei deren Bestimmung auf die Wahrscheinlichkeiten anderer Aussagen zurückgegriffen wird.

Basis für die Bestimmung der Wahrscheinlichkeit der Aussage A sind Wahrscheinlichkeiten von weiteren Aussagen.
Das kann sein:
- Zerlegung der Aussage A in mehrere Teilaussagen (Subadditivität)
- Bewertung der Aussage A, indem man Annahmen über die Wahrheit weiterer Aussagen trifft (bedingte Wahrscheinlichkeit).

5.9.1 Die Subadditivität der Wahrscheinlichkeit

In diesem Kapitel wird die Subadditivität von Wahrscheinlichkeiten plausibel hergeleitet. Basis ist Kapitel 4.3, insbesondere (4.3.-01) Fall 1.

Zunächst ist klar, dass Kombinationen von zwei Aussagen A und B sinnvoll sind, wenn man Kombination mit „oder“ und Überlappung mit „und“ gleichsetzt. Damit ist die erste Hürde genommen.

Als zweites ist zu klären, ob für die Wahrscheinlichkeit P und die Aussagen A und B die Subadditivität

P(A oder B) = P(A) + P(B) – P(A und B)

sinnvoll ist. Das wäre sehr hilfreich, denn in diesem Falle hätte man besonders einfach Rechenregeln. Ferner kann dann auch z. B. Mengen-Diagramme zur Veranschaulichung einsetzen.

Tatsächlich gibt es gute Argumente für die Sinnhaftigkeit der Subadditivität und deshalb hat sich dieser Ansatz auch allgemein durchgesetzt.

(5.9.1-01) ***Argumente***

1. In der klassischen Aussagenlogik gibt es nur wahr und falsch, also die Wahrscheinlichkeiten 1 und 0. Anhand der beiden Wahrheitstafeln für „oder“ und „und“ lässt sich leicht die Gültigkeit der Subadditivität nachweisen:

„oder“	**P(B) = 0**	**P(B) = 1**
P(A) = 0	P(A oder B) = 0	P(A oder B) = 1
P(A) = 1	P(A oder B) = 1	P(A oder B) = 1

„und“	**P(B) = 0**	**P(B) = 1**
P(A) = 0	P(A und B) = 0	P(A und B) = 0
P(A) = 1	P(A und B) = 0	P(A und B) = 1

   (1) A wahr und B wahr: P(A oder B) = P(A) + P(B) – P(A und B) ⇔ 1 = 1 + 1 – 1
   (2) A wahr und B falsch: P(A oder B) = P(A) + P(B) – P(A und B) ⇔ 1 = 1 + 0 – 0
   (3) A falsch und B wahr: P(A oder B) = P(A) + P(B) – P(A und B) ⇔ 1 = 0 + 1 – 0
   (4) A falsch und B falsch: P(A oder B) = P(A) + P(B) – P(A und B) ⇔ 0 = 0 + 0 – 0
   Für beliebige Aussagen gilt also die Subadditivität, wenn man als Wahrscheinlichkeiten nur die Werte 0 und 1 zulässt.

2. In der Methode zur Bestimmung einer elementaren Wahrscheinlichkeit wurde bereits dargelegt, dass man relative Anteile zur Bestimmung von Wahrscheinlichkeiten heranziehen kann.
   Am Beispiel der Regenwahrscheinlichkeit sieht man, dass es verschiedene mögliche Interpretationen der Aussage „Die Regenwahrscheinlichkeit beträgt x % für den Zeitraum y im Gebiet z“ gibt. Es könnte ein relativer Flächenanteil, relativer Zeitanteil, relative Häufigkeit von Expertenmeinungen oder ähnliches gemeint sein. Beim Wetterbericht im Fernsehen ist übrigens damit gemeint, dass es in x % vergleichbarer Wetterlagen geregnet hat.
   Für relative Anteile gilt stets die Subadditivität, also in diesem Fall auch für Wahrscheinlichkeiten.

3. Die Subadditivität ist konform zur Bestimmung einer konsolidierten Wahrscheinlichkeit
   Gilt im Verfahren gemäß Kapitel 5.8 für jede einzelne elementare Wahrscheinlichkeit die Subadditivität, so gilt sie auch für die konsolidierte elementare Wahrscheinlichkeit:

$$\begin{aligned}
P(\text{A oder B}) &= \sum_{i=1}^{n} d_i \cdot P_i(\text{A oder B}) = \sum_{i=1}^{n} d_i \cdot (P_i(A) + P_i(B) - P_i(\text{A und B})) \\
&= \sum_{i=1}^{n} d_i \cdot P_i(A) + \sum_{i=1}^{n} d_i \cdot P_i(B) - \sum_{i=1}^{n} d_i \cdot P_i(\text{A und B}) \\
&= P(A) + P(B) - P(\text{A und B})
\end{aligned}$$

Man kann zwar die Subadditivität der Wahrscheinlichkeit nicht für alle Fälle zwingend beweisen – insbesondere nicht für die Bauchgefühl-Wahrscheinlichkeit –, aber durch die bisherigen Betrachtungen wurde sie für viele Fälle plausibel gemacht. Für die Fälle, für die man keine stichhaltigen Argumente findet, kann man die Subadditivität per Definition erzwingen, um einheitliche Regeln für Wahrscheinlichkeiten beliebiger Aussagen zu haben.

Dabei können aber folgende Probleme auftreten:

- Man berechnet die vier Wahrscheinlichkeiten gemäß den Methoden für elementare Wahrscheinlichkeiten (die gelten auch für „A und B“ und „A oder B“) und stellt fest: alle vier Wahrscheinlichkeiten liegen zwar in [0; 1], aber die Subadditivität ist verletzt.
- Man berechnet drei Wahrscheinlichkeiten gemäß den Methoden für elementare Wahrscheinlichkeiten, bestimmt die vierte Wahrscheinlichkeit aus der Subadditivität und stellt fest: die Subadditivität ist zwar erfüllt, aber die vierte Wahrscheinlichkeit liegt nicht in [0; 1].

In diesen Fällen muss dann solange an den einzelnen Wahrscheinlichkeiten „gedreht“ werden, bis das Problem gelöst ist. Also ein sehr pragmatischer Ansatz wie beim Heimwerker: was nicht passt, wird passend gemacht.

Es gibt also gewichtige Gründe, die für die Subadditivität der Wahrscheinlichkeit sprechen und keine Gründe, die dagegen sprechen. Damit gilt die Subadditivität solange, bis Probleme auftauchen – was aber nach den bisherigen Erfahrungen nicht zu erwarten ist (siehe auch Hinweis (5.8.-06)).

Ab jetzt wird bei allen Wahrscheinlichkeiten die Subadditivität vorausgesetzt.

Wenn man mehrere Aussagen hat, ihre Wahrscheinlichkeiten ermittelt und diese dann über Gleichungen oder Ungleichungen in einen Zusammenhang bringen will, so können Widersprüche auftreten, wenn die einzelnen Bewertungsverfahren nicht widerspruchsfrei zueinander passen. Analog zu den Warentests, dem Global Peace Index oder anderen Bewertungen in Kapitel 4 muss man ein einheitliches Bewertungsverfahren für alle Aussagen, die man in einen Zusammenhang bringen will, definieren.

Dafür ist folgende Definition sinnvoll:

(5.9.1-02) ***Definition***
S sei eine endliche Menge von Aussagen.
K sei die Menge aller endlichen Verknüpfungen von Aussagen aus S mit den Verknüpfungen „und“, „oder“ oder „nicht“.
P sei die Wahrscheinlichkeit der bewertenden Person für die Aussagen aus K.
Dann heißt (S, K, P) ein **endlicher Wahrscheinlichkeitsraum für Aussagen**.

(5.9.1-03) ***Hinweise***

- Das Bewertungsverfahren P ist für alle Aussagen aus K gleich.
- Der Zusatz „für Aussagen“ wurde gewählt, da in Kapitel 6.3 der Sonderfall auf Basis von Ereignissen dargestellt wird.
- Für Aussagen $A \in S$ kann $P(A) = 1$ sein. In jedem Falle ist $P(A \text{ oder } \overline{A}) = 1$ für alle $A \in K$. Bei endlichen Wahrscheinlichkeitsräumen für Ereignisse gibt es genau ein Ereignis A mit $P(A) = 1$ (siehe (6.3.-3)).
- Für Aussagen $A \in S$ kann $P(A) = 0$ sein. In jedem Falle ist $P(A \text{ und } \overline{A}) = 0$ für alle $A \in K$. Bei endlichen Wahrscheinlichkeitsräumen für Ereignisse gibt es genau ein Ereignis A mit $P(A) = 0$ (siehe (6.3.-3)).

(5.9.1-04) ***Satz***

Gegeben sei ein endlicher Wahrscheinlichkeitsraum für Aussagen (S, K, P).
Dann gibt es eine endliche Teilmenge $K_e \subset K$, so dass es zu jeder Aussage $A \in K$ eine logisch äquivalente Aussage $A_e \in K_e$ gibt: $A \Leftrightarrow A_e$.

Beweis:
Der folgende Beweis basiert auf [JE], Kapitel 1.6.
Es sei $S = \{A_1, \dots, A_n\}$. Es sei durch $B = f(A_1, \dots, A_n)$ eine logische Funktion definiert, die dadurch entsteht, dass die Aussagen $A_1, \dots, A_n$ mit „und“, „oder“ und „nicht“ kombiniert werden.
Dann gibt es nach Formel (3.2.-04) 2^n mögliche Kombinationen der Wahrheitswerte „wahr“ und „falsch“ für diese n Aussagen. Zu jeder dieser 2^n Kombinationen gibt es unabhängig voneinander 2 Möglichkeiten für den Wahrheitswert der logischen Funktion, insgesamt gibt es also $2^{(2^n)}$ mögliche logische Funktionen. Jede Kombination von Aussagen kann auf eine dieser Kombinationen logisch äquivalent umgeformt werden.

Beispiel:
Ist n = 2, so gibt es für die beiden Aussagen $A_{1,2}$ vier Kombinationen von Wahrheitswerten und insgesamt höchstens $2^{(2^2)} = 2^4 = 16$ logische Funktionen:

Wahrheitswerte von A_1 und A_2	WW	WF	FW	FF
Zugehöriger Wahrheitswert von $B_1 = f_1(A_1, A_2)$	W	W	W	W
Zugehöriger Wahrheitswert von $B_2 = f_2(A_1, A_2)$	W	W	W	F
…				
Zugehöriger Wahrheitswert von $B_{16} = f_{16}(A_1, A_2)$	F	F	F	F

Ob es tatsächlich alle 16 Funktionen gibt, ist unwesentlich, wichtig ist nur, dass es keine weiteren Funktionen gibt.

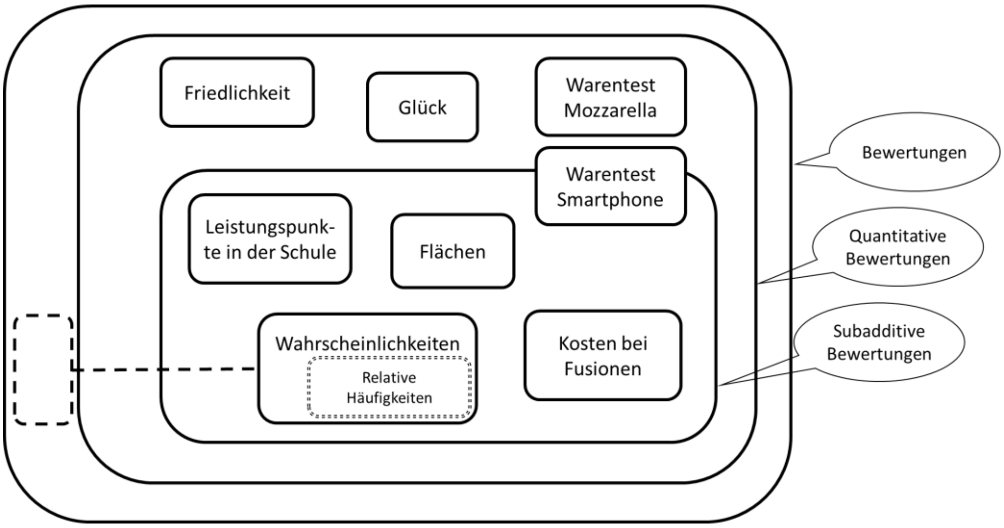
(5.9.1-05) ***Satz***

Für beliebige Aussage A und B gilt: wenn A und B logisch äquivalent sind, also $A \Leftrightarrow B$, so sind auch die Wahrscheinlichkeiten gleich: $P(A) = P(B)$.

Beweis:

Ist A ⇔ B, so ist A wahr, wenn B wahr ist, und umgekehrt. Wenn der Grad des Anscheins, dass A wahr ist, gleich x ist, also P(A) = x, dann ist der Grad des Anscheins, dass B wahr ist, aufgrund der Äquivalenz ebenfalls gleich x, also P(B) = x, also insgesamt: P(A) = P(B).

(5.9.1-06) ***Grafik***
Die folgende Grafik ordnet die Wahrscheinlichkeit in die verschiedenen Arten von Bewertungen ein. Sie macht auf einen Blick deutlich, dass die Wahrscheinlichkeit nur eine von vielen Bewertungen ist, also etwas eher Unspektakuläres und Normales:



Die durchgezogene Umrandung von „Wahrscheinlichkeiten“ gilt, wenn nur Zahlen als Wahrscheinlichkeiten zugelassen werden. Die gestrichelte Erweiterung gilt, wenn auch Worte wie „hoch“ oder „niedrig“ zugelassen werden. Das Feld „Relative Häufigkeiten“ soll symbolisieren, dass Wahrscheinlichkeiten in vielen Fällen auch aus relativen Häufigkeiten abgeleitet werden.

5.9.2 Die bedingte Wahrscheinlichkeit

Bei jeder Wahrscheinlichkeit muss die Einstellung der bewertenden Person dargestellt werden, also insbesondere, was als Wissen oder Annahmen vorausgesetzt wird. Will man deutlich machen, dass darüber hinaus eine zusätzliche Bedingung gilt, so benutzt man den Begriff der bedingten Wahrscheinlichkeit. Dabei ist rechnerisch unerheblich, ob es sich bei der zusätzlichen Bedingung um Wissen, Vermutungen, Gedankenexperimente („was wäre, wenn …“) oder nur Bauchgefühl handelt.

Im Folgenden sei wieder ein endlicher Wahrscheinlichkeitsraum für Aussagen (S, K, P) mit A, B ∈ K gegeben.

(5.9.2-01) ***Definition***
Die neue Wahrscheinlichkeitsbewertung P_B erhält man, indem bei P zusätzlich die Bedingung

P(B) = 1

in die Bewertung einfließt. P_B heißt dann „**Wahrscheinlichkeit unter der Zusatzbedingung B**“ oder kurz: **bedingte Wahrscheinlichkeit**.

Statt $P_B(A)$ schreibt man auch P(A|B) und so wird das auch im Folgenden meistens gemacht.

$P_B(A)$ ist die Wahrscheinlichkeit von A unter der Zusatzbedingung P(B) = 1. Der Schritt P $\rightarrow$ P_B charakterisiert die Veränderung von Wahrscheinlichkeiten aufgrund einer veränderten Einstellung – nämlich der zusätzlichen Erkenntnis oder Annahme, dass die Aussage B richtig ist. Diese Bedingung muss in die Bewertung jedes einzelnen Einflussfaktors einbezogen werden.
Da P_B eine neue Wahrscheinlichkeitsbewertung ist, wird dadurch auch ein neuer endlicher Wahrscheinlichkeitsraum für Aussagen (S, K, P_B) definiert. Wegen $P_B(B) = 1$ gibt es in diesem Fall immer mindestens eine Aussage, deren Wahrscheinlichkeit gleich 1 ist.

Eine weitere Methode, um P(A|B) zu bestimmen, liefert die folgende Formel, die einen Zusammenhang zwischen den beiden Bewertungsverfahren P und P_B herstellt:

(5.9.2-02) ***Satz***
Gegeben seien die endlichen Wahrscheinlichkeitsräume für Aussagen (S, K, P) und (S, K, P_B) mit A, B $\in$ K und P(B) $\neq$ 0. Dann ist

$$\mathbf{P_B(A) = \frac{P(A\text{ und }B)}{P(B)}}$$

Interpretation:
Diese Formel ist folgendermaßen zu interpretieren:
Für die gesamte Formel gilt zunächst dasselbe Bewertungsverfahren P. Auf der linken Seite der Gleichung steht die Wahrscheinlichkeit für die Richtigkeit der Aussage A, wenn man zusätzlich weiß oder annimmt, dass die Aussage B richtig ist. Diese Zusatzannahme gilt nicht für die rechte Seite (sonst könnte man den Nenner weglassen).

Eine andere Formulierung ist:
Auf der linken Seite der Gleichung steht die Wahrscheinlichkeit von A, nachdem man festgestellt oder angenommen hat, dass B richtig ist, während die Wahrscheinlichkeiten auf der rechten Seite ermittelt werden, bevor man festgestellt oder angenommen hat, dass B richtig ist.

Plausible Begründung:
Die Formel gilt gemäß (2.1.4-16) für relative Anteile und somit auch für Wahrscheinlichkeiten, wenn sie nur aus relativen Anteilen abgeleitet werden. In der grafischen Darstellung bei (2.1.4-16) sind dann die Strecken als Wahrscheinlichkeiten statt als relative Anteile zu interpretieren.
Wie bei den Argumenten für die Subadditivität gilt: die Formel in Satz (5.9.2-02) wird für beliebige Wahrscheinlichkeiten vorausgesetzt, da es gewichtige Gründe gibt, die für die Formel sprechen und keine Gründe, die dagegen sprechen. Damit gilt die Formel solange, bis Probleme auftauchen – was aber nach den bisherigen Erfahrungen nicht zu erwarten ist (siehe auch Hinweis (5.8.-06)).

(5.9.2-03) ***Hinweis***
In manchen Veröffentlichungen wird für den Fall, dass (auf der rechten Seite der Gleichung) P(B) = 0 ist, definiert (siehe z. B. [AE], Kapitel 3.1, Seite 143):

$P_B(A) = 0, \ A \in K$

Dann wäre aber insbesondere auch $P_B(B) = 0$, und das ist ein Widerspruch.

(5.9.2-04) ***Beispiel***

Aussage A = „X-land tritt bis <Datum> aus dem Euro aus (Xexit)“
Um die Wahrscheinlichkeit zu ermitteln, bewertet man alle Argumente, die für die Richtigkeit der Aussage sprechen (Argumente, die dagegen sprechen, werden verneint). Im Beispiel wird angenommen, dass es nur zwei Szenarien gibt, unter denen A eintreten kann. Wenn beide Szenarien nicht eintreten, dann gibt es auch keinen Xexit. A kann also nur zusammen mit B oder zusammen mit C (oder beidem) eintreten. Die Wahrscheinlichkeiten seien:

Argument	**Wahrscheinlichkeit**	**Wirksamkeit**
B: Die EU fordert X-land zum Austritt auf	0,3	0,5
C: Die X-ländische Bevölkerung stimmt für einen Austritt	0,15	0,8

Die Wahrscheinlichkeiten in der Tabelle beziehen sich auf die Argumente. Die Wirksamkeit sagt aus, mit welcher Wahrscheinlichkeit der Xexit erfolgt, wenn das Argument wahr ist, ist also einfach eine bedingte Wahrscheinlichkeit.
Z. B. bedeutet in der ersten Zeile: Die Wahrscheinlichkeit, dass B wahr wird, beträgt 0,3. Wenn die EU X-land zum Austritt aufgefordert hat, dann beträgt die Wahrscheinlichkeit, dass X-land dann auch tatsächlich austritt: $P(A|B) = 0{,}5$.
Würde die Tabelle nur aus Zeile B bestehen, so wäre also:

$$P(A) = P(A \text{ und } B) = P(B)\cdot P(A|B) = 0{,}3\cdot 0{,}5 = 0{,}15$$

Zusammen mit C ergibt sich dann:

$$P(A) = 0{,}3\cdot 0{,}5 + 0{,}15\cdot 0{,}8 - a\cdot b = 0{,}27 - a\cdot b$$

Dabei ist

a = Wahrscheinlichkeit, dass B und C gleichzeitig wahr sind, also

$$a = P(B \text{ und } C) \leq \min(0{,}3;\, 0{,}15) = 0{,}15 \qquad \Rightarrow \quad 0 \leq a \leq 0{,}15$$

b = Wirksamkeit von B und C für A, also

$$b = P(A|B \text{ und } C) \geq \max(0{,}5;\, 0{,}8) = 0{,}8 \qquad \Rightarrow \quad 0{,}8 \leq b \leq 1$$

also

$$0 \leq a\cdot b \leq 0{,}15$$

Allgemein ergibt sich, wenn es nur zwei Argumente B und C gibt:

$$\begin{aligned} P(A) &= P(A \text{ und } (B \text{ oder } C)) = P((A \text{ und } B) \text{ oder } (A \text{ und } C)) \\ &= P(A \text{ und } B) + P(A \text{ und } C) - P(A \text{ und } (B \text{ und } C)) \\ &= P(B)\cdot P(A|B) + P(C)\cdot P(A|C) - P(B \text{ und } C)\cdot P(A|B \text{ und } C) \end{aligned}$$

Alle verwendeten Zahlen sind mehr oder weniger subjektiv, aber das Verfahren macht zumindest das „Bauchgefühl“ zum Xexit etwas transparenter.

(5.9.2-05) ***Hinweis***
Es gibt also 2 Verfahren, um P(A|B) und 3 Verfahren, um P(A und B) zu ermitteln:

$$P(A|B) \quad = \begin{cases} \text{direkte Bewertung von A mit der Zusatzinformation "B ist richtig"} \\ \dfrac{P(A \text{ und } B)}{P(B)} \end{cases}$$

$$P(A \text{ und } B) = \begin{cases} \text{direkte Bewertung von "A und B", keine Einzelbewertungen} \\ P(A) + P(B) - P(A \text{ oder } B) \\ P(A|B)\cdot P(B) \end{cases}$$

Die direkte Bewertung auf Basis einer Bewertungstabelle mit Einflussfaktoren und Gewichten ist immer möglich, während die anderen Methoden nur dann sinnvoll sind, wenn man die jeweiligen Wahrscheinlichkeiten schon hat oder zumindest leicht ermitteln kann.

(5.9.2-06) ***Hinweis***
Wie man an der Grafik in (2.1.4-16) sieht, kann P(A) durch die Zusatzbedingung, dass B richtig ist, größer werden, kleiner werden oder gleich bleiben:

1. Je stärker sich A und B überschneiden, umso größer wird P(A|B). Ist sogar B ein Teil von A, d. h., wenn B wahr ist, dann ist auch erst recht A wahr, so ist
   A und B = B
   und somit
   $P(A|B) = \frac{P(A \text{ und } B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

2. Je weniger sich A und B überschneiden, umso kleiner wird P(A|B). Überschneiden sich A und B überhaupt nicht, so folgt aus dem Wissen, dass B richtig ist, dass A nicht richtig sein kann, also
   $P(A|B) = 0$.

(5.9.2-07) ***Definition***
Überschneiden sich A und B so, dass
P(A) = P(A|B)
ist, so ändert sich die Wahrscheinlichkeit von A durch die zusätzliche Bedingung, dass B richtig ist, nicht. Man sagt in diesem Fall auch: „**A ist stochastisch unabhängig von B“.**
Ist dagegen
P(A) ≠ P(A|B),
so ist **A von B stochastisch abhängig**.

(5.9.2-08) ***Folgerungen***
1. **Ist A von B stochastisch unabhängig, so ist auch B von A stochastisch unabhängig**.
   Man sagt also einfach: A und B sind voneinander stochastisch unabhängig.
   Beweis:
   A sei von B stochastisch unabhängig, also P(A|B) = P(A). Dann folgt
   $P(B|A) = \frac{P(B \text{ und } A)}{P(A)} = \frac{P(A \text{ und } B)}{P(B)} \cdot \frac{P(B)}{P(A)} = P(A|B) \cdot \frac{P(B)}{P(A)} = P(A) \cdot \frac{P(B)}{P(A)} = P(B)$,
   B ist also von A stochastisch unabhängig.
   Hinweis: die Formel $P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)}$ ist der Satz von Bayes.

2. **A und B sind genau dann voneinander stochastisch unabhängig, wenn**
   P(A und B) = P(A)·P(B) ist.
   Beweis:
   $P(A|B) = P(A) \Leftrightarrow \frac{P(A \text{ und } B)}{P(B)} = P(A) \Leftrightarrow P(A \text{ und } B) = P(A) \cdot P(B)$

3. **Ist A von B stochastisch unabhängig, so ist A auch von $\overline{B}$ stochastisch unabhängig**
   Beweis:
   Nach dem Distributivgesetz ist
   $A = A \text{ und } (B \text{ oder } \overline{B}) = (A \text{ und } B) \text{ oder } (A \text{ und } \overline{B})$

und damit

$$\begin{aligned} P(A) &= P(A \text{ und } B) + P(A \text{ und } \overline{B}) - P((A \text{ und } B) \text{ und } (A \text{ und } \overline{B})) \\ &= P(A \text{ und } B) + P(A \text{ und } \overline{B}) \end{aligned}$$

Daraus folgt

$$P(A|\overline{B}) = \frac{P(A \text{ und } \overline{B})}{P(\overline{B})} = \frac{P(A) - P(A \text{ und } B)}{P(\overline{B})} = \frac{P(A) - P(A)\cdot P(B)}{1 - P(B)} = \frac{P(A)\cdot(1-P(B))}{1-P(B)} = P(A)$$

4. **Schließen sich A und B gegenseitig aus, so gilt: P(A|B) = 0**
   Beweis:
   Schließen sich A und B gegenseitig aus, so können A und B nicht gleichzeitig wahr sein, es ist also P(A und B) = 0 und damit folgt:
   $$P(A|B) = \frac{P(A \text{ und } B)}{P(B)} = 0$$

5. **Folgt aus der Richtigkeit von B stets die Richtigkeit von A (B ist dann eine Teilaussage von A), so gilt: P(A|B) = 1**
   Beweis:
   Aus den Voraussetzungen folgt
   A und B = B
   und somit
   $$P(A|B) = \frac{P(A \text{ und } B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(5.9.2-09) ***Beispiele***

1. Es wurde eine Wahl durchgeführt, bei der 4 Frauen und 3 Männer zur Auswahl standen. Da keine weiteren Informationen bekannt sind, geht man davon aus, dass die Wahrscheinlichkeit, gewählt zu werden, für jede Kandidatin/jeden Kandidaten gleich der relativen Häufigkeit, also = $\frac{1}{7}$ ist. Während der Auszählung sickert durch, dass eine Frau gewinnen wird, da sie bereits uneinholbar vorne liegt. Die Wahrscheinlichkeit, gewählt zu werden, steigt für jede Frau auf $\frac{1}{4}$ und fällt für jeden Mann auf 0. Das tatsächliche Ergebnis der Wahl wird durch die Zusatzinformationen nicht beeinflusst, sondern nur die Einschätzung der Wahrscheinlichkeiten für den Wahlausgang.

2. Wird bei einer Wahl zunächst unter verschiedenen Kandidaten gewählt und findet dann im zweiten Durchgang eine Stichwahl zwischen den beiden Kandidaten mit den meisten Stimmen statt, so beeinflusst das Ergebnis des 1. Schrittes das Ergebnis des 2. Schrittes.

3. Bei einer Quizsendung im Fernsehen gibt es 4 Antworten auf eine Frage, wobei nur genau eine Antwort richtig ist. Rät der Kandidat einfach nur, so ist die Gewinnwahrscheinlichkeit gleich der relativen Häufigkeit, also = $\frac{1}{4}$. Hat er aber zumindest Teilwissen oder werden 2 Antworten ausgeschlossen (50 %-Joker), so kann er die Gewinnwahrscheinlichkeit steigern. Die tatsächliche korrekte Antwort wird durch die Zusatzinformationen nicht beeinflusst, da sie schon von Beginn an feststand.

4. Man hat einen roten und einen grünen Würfel und geht davon aus, dass
   - bei jedem Wurf die Wahrscheinlichkeit für jede Augenzahl gleich = $\frac{1}{6}$ ist
   - und sich die Würfel beim Würfeln gegenseitig nicht beeinflussen.

   Man würfelt mit jedem Würfel einmal. Zur Abkürzung sei R = „der rote Würfel hat die Augenzahl“ und G = „der grüne Würfel hat die Augenzahl“. Dann ist nach Regel 2:

$P(R=a \text{ und } G=b) = P(R=a)\cdot P(G=b) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$

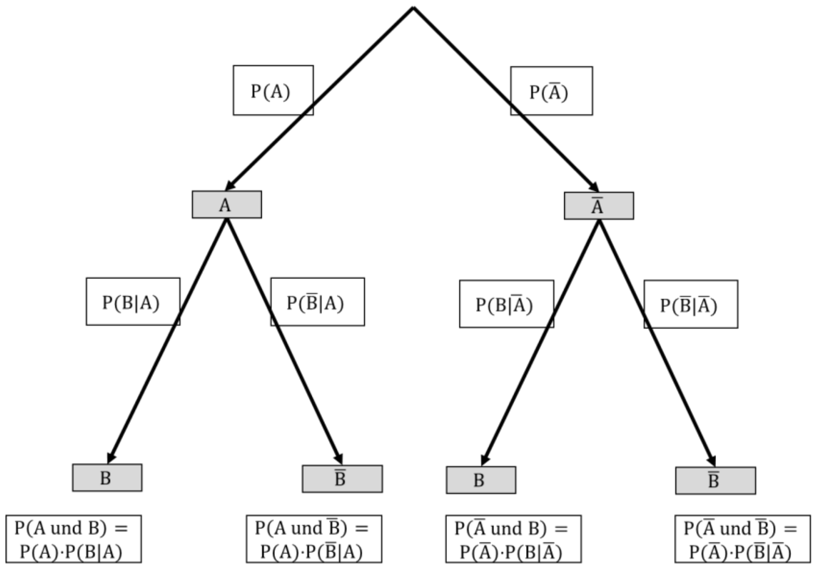
(5.9.2-10) ***Hinweis***
Wenn A und B logisch voneinander unabhängig sind, dann spielt die Einschätzung über den Wahrheitsgehalt von B in der Bewertung von A keine Rolle und umgekehrt. Logische und stochastische Unabhängigkeit haben nur bedingt etwas miteinander zu tun:

- Sind A und B logisch voneinander abhängig, sodass man etwas Zusätzliches über A sagen kann, wenn man etwas über B weiß, dann kann – muss aber nicht – das Einfluss auf die Wahrscheinlichkeiten haben, d. h., A und B können in diesem Fall stochastisch voneinander abhängig oder unabhängig sein.
- Sind A und B logisch voneinander unabhängig, sodass man nichts Zusätzliches über A sagen kann, wenn man etwas über B weiß, so sind sie auch stochastisch unabhängig.
- Stochastische Unabhängigkeit ist symmetrisch, logische Unabhängigkeit nicht:
  Ist A = „Stufe 1 des Projektes ist erfolgreich“ und B = „Stufe 2 des Projektes ist erfolgreich“, dann ist B von A logisch abhängig, aber nicht umgekehrt.

5.9.3 Baumdiagramme

Gegeben sei ein endlicher Wahrscheinlichkeitsraum für Aussagen (S, K, P) mit A, B ∈ K. Hat man mehrere Aussagen, so kann man Aussagen oder ihre Verneinungen mit „und“ verknüpfen und die zugehörigen Wahrscheinlichkeiten durch einen **Baum** darstellen.

Im einfachsten Fall zweier Aussagen A und B ergibt sich folgender Baum:



Die Summe der Wahrscheinlichkeiten unter einer Verzweigung ist immer = 1

Ebene 1: $P(A) + P(\overline{A}) = 1$

Ebene 2: $P(B|A) + P(\overline{B}|A) = 1$ und $P(B|\overline{A}) + P(\overline{B}|\overline{A}) = 1$

Beweis für Ebene 2 - links:

$$P(B|A) + P(\overline{B}|A) = \frac{P(B \text{ und } A)}{P(A)} + \frac{P(\overline{B} \text{ und } A)}{P(A)} = \frac{P(B \text{ und } A) + P(\overline{B} \text{ und } A)}{P(A)} = \frac{P((B \text{ und } A) \text{ oder } (\overline{B} \text{ und } A))}{P(A)}$$

$$= \frac{P((B \text{ oder } \overline{B}) \text{ und } A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

Erste Pfadregel (Produktregel):
Die Gleichungen am unteren Ende jedes Pfades drücken aus, dass die Wahrscheinlichkeit, dass ein bestimmter Pfad durchlaufen wird, gerade das Produkt der Einzelwahrscheinlichkeiten ist. Dies folgt unmittelbar aus der Gleichung für die bedingte Wahrscheinlichkeit.

Zweite Pfadregel (Summenregel):
Werden Aussagen am unteren Ende der Pfade mit „oder“ verknüpft, so sind die zugehörigen Wahrscheinlichkeiten zu addieren. Das folgt aus der Subadditivität und daraus, dass sich die Aussagen gegenseitig ausschließen.
Beispiel:
$P(B) = P((A \text{ oder } \overline{A}) \text{ und } B) = P((A \text{ und } B) \text{ oder } (\overline{A} \text{ und } B))$
$= P(A \text{ und } B) + P(\overline{A} \text{ und } B) – P((A \text{ und } B) \text{ und } (\overline{A} \text{ und } B))$
$= P(A \text{ und } B) + P(\overline{A} \text{ und } B)$

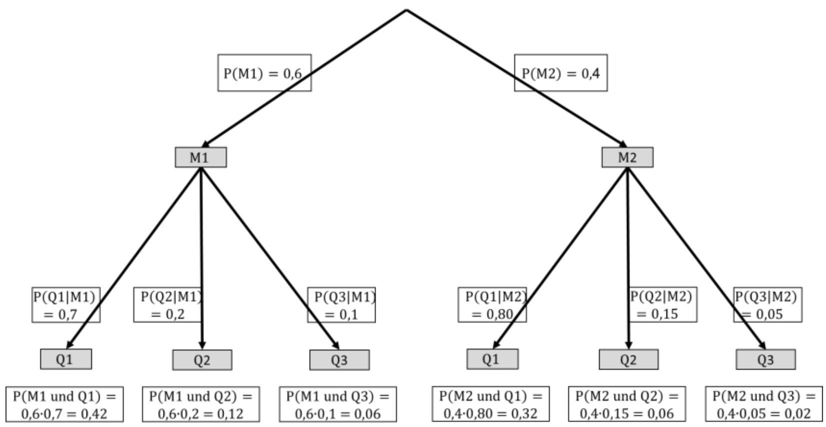
Sind die Aussagen stochastisch unabhängig voneinander, so ist
$P(A|B) = P(A)$ und $P(A \text{ und } B) = P(A) \cdot P(B)$
In diesem Fall ist es vom Aufwand her egal, ob erst A und dann B oder erst B und dann A im Baumdiagramm betrachtet wird.

Sind die Aussagen aber stochastisch abhängig voneinander, so kann der Aufwand für das Baumdiagramm je nach Reihenfolge unterschiedlich sein.

(5.9.3-01) ***<u>Beispiel</u>***
60 % eines Produktes wird mit der Maschine M1 und 40 % mit der Maschine M2 produziert. Die Produkte werden nach drei Qualitätsstufen Q1, Q2, Q3 klassifiziert. Die Anteile der Produkte sind bei den Maschinen:
M1: Q1 – 70 %, Q2 – 20 %, Q3 – 10 %
M2: Q1 – 80 %, Q2 – 15 %, Q3 – 5 %

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Produkt eine bestimmte Qualitätsstufe hat und von einer bestimmten Maschine stammt?

<u>Variante 1:</u>
Es sei kurz
Mx = „Das Produkt wurde auf Maschine x produziert“; x = 1, 2
Qy = „Das Produkt hat Qualitätsstufe y“; y = 1, 2, 3
Wählt man die Aussagen Mx als 1. Ebene und die Aussagen Qy als 2. Ebene, so erhält man ohne weitere Umrechnung:



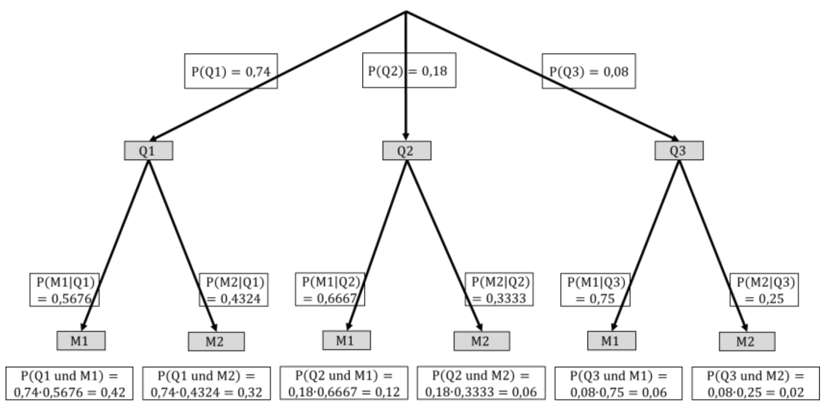
Die gesuchten Wahrscheinlichkeiten sind ganz unten im Diagramm abzulesen.

Variante 2:
Will man die Qualitätsstufen auf Ebene 1 und die Maschinen auf Ebene 2 darstellen, so muss man die Zuordnung Maschine <-> Qualitätsstufe umkehren. Der Aufwand ist wesentlich höher, die Ergebnisse sind aber dieselben. Das Verfahren ist nur sinnvoll, wenn auch die Zwischenergebnisse wichtig sind. Mit den Ergebnissen aus Variante 1 ergibt sich:

$$
\begin{aligned}
P(Q1) &= P((M1 \text{ oder } M2) \text{ und } Q1) \\
&= P(M1 \text{ und } Q1) + P(M2 \text{ und } Q1) = 0{,}42 + 0{,}32 = 0{,}74 \\
P(Q2) &= P(M1 \text{ und } Q2) + P(M2 \text{ und } Q2) = 0{,}12 + 0{,}06 = 0{,}18 \\
P(Q3) &= P(M1 \text{ und } Q3) + P(M2 \text{ und } Q3) = 0{,}06 + 0{,}02 = 0{,}08
\end{aligned}
$$

$$
\begin{aligned}
P(M1|Q1) &= \frac{P(M1 \text{ und } Q1)}{P(Q1)} = \frac{0{,}42}{0{,}74} = 0{,}5676 \\
P(M1|Q2) &= \frac{P(M1 \text{ und } Q2)}{P(Q2)} = \frac{0{,}12}{0{,}18} = 0{,}6667 \\
P(M1|Q3) &= \frac{P(M1 \text{ und } Q3)}{P(Q3)} = \frac{0{,}06}{0{,}08} = 0{,}75
\end{aligned}
$$

$$
\begin{aligned}
P(M2|Q1) &= \frac{P(M2 \text{ und } Q1)}{P(Q1)} = \frac{0{,}32}{0{,}74} = 0{,}4324 \\
P(M2|Q2) &= \frac{P(M2 \text{ und } Q2)}{P(Q2)} = \frac{0{,}06}{0{,}18} = 0{,}3333 \\
P(M2|Q3) &= \frac{P(M2 \text{ und } Q3)}{P(Q3)} = \frac{0{,}02}{0{,}08} = 0{,}25
\end{aligned}
$$



Wären die Qualitätsstufen unabhängig von der Maschine, so wäre es egal, ob man erst Mx und dann Qy oder erst Qy und dann Mx im Baumdiagramm darstellt. Der Aufwand ist derselbe.

5.9.4 Zusammenhänge

Um andere Vorgehensweisen kurz zu beleuchten, wird in diesem Kapitel nur

$P(A) \geq 0$ und

$P_B(A) \cdot P(B) = P(A \text{ und } B)$ (siehe Satz (5.9.2.-02))

vorausgesetzt, aber nicht $P(A) \leq 1$ und nicht die Subadditivität. Dann folgt daraus:

1. Für jede wahre Aussage S gilt:

   $P(S) = 1$

   Wäre die maximale Wahrscheinlichkeit, also die Wahrscheinlichkeit für eine wahre Aussage, allgemein = x, so würde aus der Formel in Satz (5.9.2-02) für B = A folgen:

   Linke Seite: $P_A(A) = x$ (Tautologie: unter der Bedingung, dass A wahr ist, ist A wahr)

   Rechte Seite: $\frac{P(A \text{ und } A)}{P(A)} = 1$

   Es folgt also x = 1. Oder umgekehrt formuliert: wäre die maximale Wahrscheinlichkeit $x \neq 1$ (siehe z. B. [SU]), so wäre die Formel in Satz (5.9.2-02) falsch.

   Die Behauptung ist natürlich nicht überraschend, da sie für relative Anteile gilt und die Formel in Satz (5.9.2.-02) aus der entsprechenden Formel für bedingte relative Anteile abgeleitet wurde.

2. Da keine Aussage plausibler als eine wahre Aussage sein kann, folgt für jede beliebige Aussage A:

   $0 \leq P(A) \leq 1$

3. Eine Aussage A ist in dem Maße plausibel, wie die Gegenaussage $P(\overline{A})$ unplausibel ist. Ist also z. B. $P(A) = 0{,}2$, so ist $P(\overline{A}) = 0{,}8$, also insgesamt (wie bei relativen Anteilen):

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

   Da P_B auch eine Wahrscheinlichkeit ist, folgt:

   $P_B(A) + P_B(\overline{A}) = 1$

   oder gleichwertig:

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

4. Auf Basis von Punkt 3. folgt die Subadditivität:

$$
\begin{aligned}
P(A \text{ oder } B) &= P(\overline{\overline{A} \text{ und } \overline{B}}) = 1 - P(\overline{A} \text{ und } \overline{B}) = 1 - P(\overline{A}|\overline{B}) \cdot P(\overline{B}) \\
&= 1 - (1 - P(A|\overline{B})) \cdot P(\overline{B}) \\
&= (1 - P(\overline{B})) + P(A|\overline{B}) \cdot P(\overline{B}) \\
&= P(B) + P(A \text{ und } \overline{B}) \\
&= P(B) + P(\overline{B}|A) \cdot P(A) \\
&= P(B) + (1 - P(B|A) \cdot P(A)) \\
&= P(B) + P(A) - P(B|A) \cdot P(A) \\
&= P(A) + P(B) - P(A \text{ und } B)
\end{aligned}
$$

Diese etwas andere Vorgehensweise ist ein weiteres Indiz dafür, dass die Formeln für die Subadditivität und die bedingte Wahrscheinlichkeit zueinander konsistent sind.

5.10 Weitere Beispiele für die praktische Wahrscheinlichkeit

Im Gegensatz zu den Beispielen in Kapitel 5.3 beinhalten die folgenden Beispiele eine konkrete Zahl als Wahrscheinlichkeit.

(5.10-01) ***Beispiel***
Bestimmung eines Trennungsrisikos für Paare, siehe z. B.:
„… Er seziert Beziehungen regelrecht und benennt das Trennungsrisiko mit 94-prozentiger Wahrscheinlichkeit.“ aus: https://www.welt.de/gesundheit/psychologie/article144220686/So-einfach-laesst-sich-das-Trennungsrisiko-berechnen.html
Siehe auch:
http://www.t-online.de/lifestyle/partnerschaft/id_74781442/liebe-das-trennungsrisiko-laesst-sich-voraussagen.html

(5.10-02) ***Beispiel***
Bei einem Tennisspiel treten die Spieler X und Y gegeneinander an. Ich kenne mich sehr gut im Tennis aus und halte die beiden Spieler für momentan gleich stark, also:

P(„Spieler X gewinnt“) = P(„Spieler Y gewinnt“) = 0,5

(5.10-03) ***Beispiel***
Ich weiß, dass ein Tennisspiel stattfindet, bei dem die zwei Spieler X und Y gegeneinander antreten. Ich habe keine Ahnung von Tennis und weiß lediglich, dass am Ende einer von beiden gewinnt, es gibt also kein Unentschieden. Ich habe folgende Möglichkeiten:

- Da ich keine Informationen habe, die einen Spieler bevorzugen, setze ich („Indifferenzprinzip“):
  P(„Spieler X gewinnt“) = P(„Spieler Y gewinnt“) = 0,5
- Da mir der Name des ersten Spielers besser gefällt als der Name des zweiten oder da ich glaube, dass ein Erstgenannter eher gewinnt oder weil ich so ein Gefühl habe oder weil ich zocken will oder … setze ich:
  P(„Spieler X gewinnt“) = 0,7
  P(„Spieler Y gewinnt“) = 0,3
- Da die genannten Möglichkeiten reine Spekulation sind, ordne ich keine Wahrscheinlichkeiten zu und sage nur: „Ich habe keine Einschätzung, wer gewinnt!“

(5.10-04) ***Beispiel***
Ein Mikromort ist eine Maßeinheit für Risiko und bezeichnet die Wahrscheinlichkeit von 0,000001 (eins zu einer Million), zu sterben.

(5.10-05) ***Beispiel***
Viele Risiken werden mit konkreten Prozentwerten angegeben, wie z. B. das Risiko einer atomaren Katastrophe

Fußnote

50

Prozent

beträgt das Risiko einer atomaren Katastrophe vom Ausmaß Tschernobyls (oder schlimmer) in den kommenden 27 Jahren. Das geht aus einer Analyse von Forschern in Zürich und Aarhus hervor. Sie haben dafür so gründlich wie niemand zuvor die Geschichte bisheriger Störfälle und Desaster ausgewertet.

FOTO: KELD NAVNTOFT / DPA (O. M.)

SPIEGEL, 25.04.2015

(5.10-06) ***Beispiel***
Die Wahrscheinlichkeit, dass ein Musikstück ein Hit wird, wird aus vielen Daten in den sozialen Medien ermittelt.

Summe kosten.

Natürlich hat auch Apple schon seine hauseigenen Analyseexperten: Die Firma hat, weitgehend unbemerkt von der Öffentlichkeit, am Anfang des Jahres das britische Start-up Semetric gekauft, das hinter dem Dienst Music Metric steht. Die Firma spielt eine wesentliche Rolle in der gerade auf der Entwicklerkonferenz von Apple angekündigten Strategie, mit dem eigenen Streamingdienst „Apple Music" Spotify und Co. die Kunden streitig machen zu wollen. Auch die Neulinge im Apple-Konzern glauben, anhand von Daten mit zunehmender Genauigkeit den Erfolg von Musikern vorhersagen zu können. Geht es um Weltstars wie Katy Perry, werten die Großrechner der Londoner mehr als 19 000 einzelne Hinweise aus, um den Erfolg eines einzelnen Songs vorherzusagen. Für Musiker, die noch nicht ganz so bekannt sind, ist die Anzahl geringer. Laut Geschäftsführer Gregory Mead ist es der Firma schon gelungen, die Nummer eins in den amerikanischen Charts mit 90-prozentiger Treffsicherheit vorherzusagen. Drei Monate im Voraus.

FASZ, 14.06.2015

(5.10-07) ***Beispiel***
Darstellung der Abstiegsgefahr vor dem letzten Spieltag der Bundesliga.

Platz 13
35 Punkte, -15 Tore
Hertha kann bei einer Niederlage mit zwei Toren Differenz allenfalls auf den Relegationsplatz abrutschen. Zudem müsste dann Stuttgart siegen, Hannover und Freiburg die Punkte teilen.
Abstiegsgefahr: 10 %

Platz 14
34 Punkte, -10 Tore
Den Freiburgern würde bereits ein Unentschieden in Hannover zur Rettung reichen. Sollten weder der HSV noch der VfB gewinnen, darf sich der SC sogar eine Niederlage leisten.
Abstiegsgefahr: 30 %

Platz 15
34 Punkte, -17 Tore
Definitiv gerettet ist Hannover nur bei einem Sieg gegen Freiburg. Im Falle eines Unentschiedens oder einer Niederlage ist der Klub von den Resultaten der Konkurrenz abhängig.
Abstiegsgefahr: 35 %

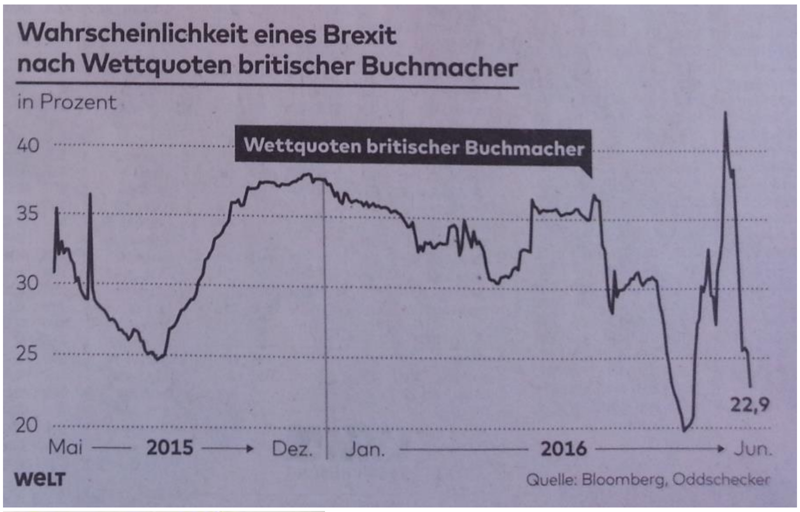
Platz 16
33 Punkte, -19 Tore
Mit einem Sieg beim Tabellenletzten Paderborn hält Stuttgart auf jeden Fall die Klasse. Möglicherweise reicht den Schwaben sogar schon ein Unentschieden, um nicht abzusteigen.
Abstiegsgefahr: 50 %

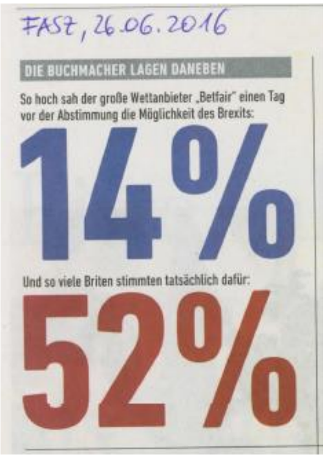
Platz 17
32 Punkte, -27 Tore
Aus eigener Kraft kann der Bundesliga-Dino den Klassenerhalt nicht mehr schaffen. Der ganz schwache Auftritt beim 1:2 gegen Stuttgart dient zudem nicht als Mutmacher.
Abstiegsgefahr: 80 %

Platz 18
31 Punkte, -33 Tore
Paderborn kann aufgrund der schlechten Tordifferenz maximal nur noch den Relegationsplatz erreichen. Ein Sieg gegen Stuttgart ist Pflicht, zudem darf der HSV nicht gewinnen.
Abstiegsgefahr: 95 %

BRAUNSCHWEIGER ZEITUNG 22.05.2015

(5.10-08) ***Beispiel***
Wahrscheinlichkeit von Brexit (vor dem Referendum am 23.06.2016) und Grexit





kein Geld mehr haben. Ohne neue Notkredite ist es unwahrscheinlich, dass die Banken nächste Woche wieder öffnen können."

BZ, 01.07.2015

Ist ein Grexit unausweichlich?
Das Bankhaus Berenberg hat die Wahrscheinlichkeit für einen Ausstieg Griechenlands aus der Eurozone, einen Grexit, von 40 auf 55 Prozent angehoben – weil, so Schmieding, die griechische Regierung für ein „Nein" im griechischen Referendum werbe. Die

(5.10-09) ***Beispiel***
Risiken bei der Geldanlage

Das Risiko immer im Griff

ANZEIGE

Scalable Capital setzt bei der Vermögensverwaltung auf moderne Technologie

Professor Dr. Stefan Mittnik (wissenschaftlicher Beirat und Gründer der digitalen Vermögensverwaltung Scalable Capital)

Zinsen? Nahe Null. Lebensversicherungen? Werfen kaum noch Rendite ab. Gesetzliche Rente? Reicht bald nicht mehr aus. Die meisten Deutschen wissen, dass sie sich dringend um ihren Vermögensaufbau kümmern sollten. Aber das Misstrauen gegenüber der Finanzbranche ist groß. Zu Recht. Fondshäuser knöpfen dem Anleger meist viel zu hohe Gebühren ab. Und Bankberater drängen dem Kunden oft Produkte auf, die der gar nicht braucht. In dieser Situation hilft der digitale Vermögensverwalter Scalable Capital. Er bietet eine Geldanlage mit wissenschaftlich fundiertem Risikomanagement zu niedrigen Kosten an. Wie sie funktioniert, erklärt Scalable-Mitgründer Stefan Mittnik, der auch Professor für Finanzökonometrie an der Ludwig-Maximilians-Universität in München ist.

Herr Professor Mittnik, was macht Scalable Capital?

Wir sind eine Online-Vermögensverwaltung. Für jeden Kunden stellen wir ein global gestreutes Portfolio aus ETFs zusammen. ETFs sind kostengünstige Fonds, die einen Wertpapier-Index nachbilden, zum Beispiel den DAX. Wie wir das Portfolio bestücken, hängt von der Risikobereitschaft des Anlegers ab.

Ein ETF-Portfolio kann sich ein Anleger mit etwas Börsenwissen auch selbst bauen. Warum soll er sein Geld von Scalable Capital verwalten lassen?

Aus zwei Gründen. Erstens gibt es rund 1.500 ETFs. Da ist es für Nicht-Profis kaum möglich, die besten auszuwählen. Und zweitens überwachen und steuern wir die Kundenportfolios fortlaufend durch unser dynamisches Risikomanagement. Das kann ein Privatanleger nicht selbst in die Hand nehmen.

Das Risikomanagement ist das Herzstück der Vermögensverwaltung von Scalable Capital. Woher wissen Sie, wie viel Risiko ein Anleger verträgt?

Bei uns gibt es 23 Risikokategorien. Wir ermitteln die geeignete für den Kunden und stellen ihm dazu bei der Depoteröffnung entsprechende Fragen, unter anderem zu seinen finanziellen Verhältnissen und seinen Anlagezielen.

Weiß der Anleger auch, wie viel Risiko er damit wirklich eingeht?

Die meisten Anbieter bezeichnen ihre Risikoklassen mit vagen Begriffen wie konservativ, moderat oder chancenorientiert. Das ist viel zu schwammig. Kein Anleger weiß, welche Verluste damit einhergehen können. Wir ordnen jeder Kategorie eine nachvollziehbare, konkrete Verlustgefahr zu, mithilfe der Risikokennziffer Value-at-Risk (VaR). Im Klartext heißt das: Wenn Sie die Risikokategorie mit einem VaR von zehn Prozent wählen, dann verliert Ihr Depot im kommenden Jahr mit einer Wahrscheinlichkeit von 95 Prozent nicht mehr als zehn Prozent. So bekommen Sie einen viel besseren Eindruck, welches Risiko Ihre Anlage tatsächlich birgt.

Das Portfoliorisiko kann sich täglich ändern – je nach Börsenlage. Wie sorgen Sie dafür, dass die Verlustgefahr nicht plötzlich steigt?

Durch den Einsatz moderner Computertechnologie. Wir ermitteln Tag für Tag für jedes Kundenportfolio den Value-at-Risk, indem wir Zehntausende Simulationsrechnungen durchführen – nach neuesten wissenschaftlichen Erkenntnissen. Weicht der VaR von der Vorgabe des Anlegers ab, schichten wir um. In risikoärmere Anlageklassen, wenn das Risiko über die festgelegte Schwelle zu steigen droht; in risikoreichere, wenn es darunter sinkt. So passen wir das Kundenportfolio laufend an die aktuelle Marktsituation an, und das Risiko bleibt unter Kontrolle. Anders ist das bei Vermögensmanagern, die die ETF-Portfolios ihrer Kunden lediglich in festen Abständen „rebalancen", das heißt: die Gewichte der jeweiligen Anlageklassen auf den Ausgangswert zurückführen. In diesem Fall kann die Verlustgefahr für den Anleger enorm schwanken, denn nur die Portfolio-Gewichte bleiben hier mehr oder weniger konstant, nicht aber die Risiken.

Was kostet die Vermögensverwaltung von Scalable Capital?

Unsere All-in-Jahresgebühr liegt bei 0,75 Prozent auf das durchschnittlich verwaltete Vermögen und beinhaltet auch alle Handelskosten. Hinzu kommen im Schnitt noch 0,25 Prozent pro Jahr für die ETF-Verwaltung. Die fließen aber nicht an uns, sondern an die Anbieter der Papiere, und würden auch dann fällig, wenn der Anleger die ETFs selbst gekauft hätte. Insgesamt ist Scalable Capital weit günstiger als die Geldanlage vom Bankberater. Wir schlagen auch keine versteckten Kosten oder Ausgabeaufschläge drauf, wie es in der Finanzbranche leider immer noch üblich ist.

FASZ, 18.12.2016

www.scalable.capital

Der Wert einer Vermögensanlage kann sowohl steigen als auch fallen. Anleger müssen deshalb bereit und in der Lage sein, Verluste des eingesetzten Kapitals hinzunehmen. Anlageergebnisse aus der Vergangenheit lassen keine Rückschlüsse auf die zukünftige Wertentwicklung zu. Weitere Informationen hierzu finden Sie auf unserer Website.

scalable .CAPITAL

5.11 Rechenregeln für Wahrscheinlichkeiten

In diesem Kapitel werden allgemeingültige Rechenregeln aus der Definition der praktischen Wahrscheinlichkeit und der Subadditivität hergeleitet.

Rechenregeln

Gegeben sei ein Wahrscheinlichkeitsraum (S, K, P) mit A, B, C, $A_i \in K$.
Dann gelten folgende Rechenregeln:

(1) **P(A und $\overline{A}$) = 0**
P(A oder $\overline{A}$) = 1

(2) **P(A oder B) = P(A) + P(B)** ⇔ **P(A und B) = 0**
⇔ A und B schließen sich gegenseitig aus
⇔ A und B können nicht beide richtig sein

(3) **P($\overline{A}$) = 1 – P(A)**
Beweis:
Setzt man B = $\overline{A}$ in Regel (2) ein und nutzt Regel (1), so ergibt sich:
P(A) + P($\overline{A}$) = P(A oder $\overline{A}$) = 1 ⇔ P($\overline{A}$) = 1 – P(A)
Oder als reine Plausibilitätsbetrachtung: A ist in dem Maße plausibel, wie $\overline{A}$ unplausibel ist, also
P($\overline{A}$) = 1 – P(A)

(4) **P(A oder B oder C) = P(A) + P(B) + P(C) – P(A und B) – P(A und C) – P(B und C) + P(A und B und C)**
Beweis:
Aus der Subadditivität folgt:
P(A oder B oder C) = P(A oder (B oder C)
= P(A) + P(B oder C) – P(A und (B oder C))
Wegen
P(B oder C) = P(B) + P(C) – P(B und C)
und
P(A und (B oder C)) = P((A und B) oder (A und C))
= P(A und B) + P(A und C) – P((A und B) und (A und C))
= P(A und B) + P(A und C) – P(A und B und C)
ergibt sich dann insgesamt:
P(A oder B oder C) = P(A) + P(B) + P(C) – P(A und B) – P(A und C) – P(B und C)
+ P(A und B und C)
Zusatz: schließen sich A, B und C paarweise gegenseitig aus, so gilt einfach:
P(A oder B oder C) = P(A) + P(B) + P(C)

(5) Hat man k Aussagen A_i (i = 1, …, k), die sich paarweise gegenseitig ausschließen, so folgt als Verallgemeinerung von Regel (4):
$P(A_1$ oder A_2 oder … oder $A_k)$ = $P(A_1) + P(A_2) + … + P(A_k)$
Beweis:
Die Formel folgt durch mehrfache Anwendung von Regel (2)

(6) Hat man k Aussagen A_i (i = 1, …, k), die sich paarweise gegenseitig ausschließen und sind alle Einzelwahrscheinlichkeiten gleich, also $P(A_1) = P(A_2) = … = P(A_k)$, so gilt für jedes i:
$$\mathbf{P(A_i) = \frac{P(A_1 \text{ oder } A_2 \text{ oder } \dots \text{ oder } A_k)}{k} \quad (i = 1, \dots, k)}$$

Beweis:
aus Regel (5) folgt:

$$\begin{aligned} P(A_1 \text{ oder } A_2 \text{ oder } \dots \text{ oder } A_k) &= P(A_1) + P(A_2) + \dots + P(A_k) \\ &= P(A_i) + P(A_i) + \dots + P(A_i) \\ &= k \cdot P(A_i), \end{aligned}$$

also

$$P(A_i) = \frac{P(A_1 \text{ oder } A_2 \text{ oder } \dots \text{ oder } A_k)}{k}$$

(7) Folgt aus der Richtigkeit der Aussage A, dass dann auch die Aussage B richtig sein muss, kurz: A ⇒ B, so gilt:

$P(A) \leq P(B)$

Beweis:

$$\begin{aligned} P(B) &= P(B \text{ und } (A \text{ oder } \overline{A})) = P((B \text{ und } A) \text{ oder } (B \text{ und } \overline{A})) \\ &= P(B \text{ und } A) + P(B \text{ und } \overline{A}) - P((B \text{ und } A) \text{ und } (B \text{ und } \overline{A})) \end{aligned}$$

Da B automatisch aus A folgt, ist: „B und A“ = „A“
Da sich A und $\overline{A}$ gegenseitig ausschließen, ist $P((B \text{ und } A) \text{ und } (B \text{ und } \overline{A})) = 0$
Damit folgt:

$$P(B) = P(A) + P(B \text{ und } \overline{A}) - 0 \geq P(A)$$

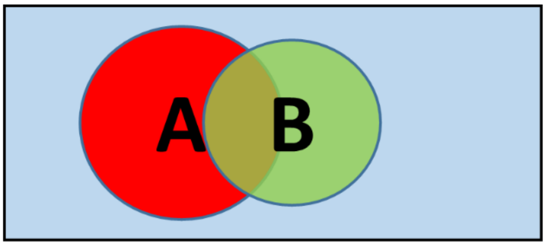
Beispiel:
A = „über meinem Rasen regnet es“
B = „mein Rasen ist nass“
Wenn es regnet, ist mein Rasen nass. Wenn es nicht regnet, kann mein Rasen trotzdem nass sein, nämlich, weil ich den Rasensprenger in Betrieb habe. P(B) ist also mindestens so groß wie P(A). Man kann natürlich auch die beiden Aussagen in den Beweis einsetzen und erhält dasselbe Ergebnis.

(8) Sind A und B logisch äquivalent, kurz A ⇔ B, so gilt:

$P(A) = P(B)$

Beweis:
Aus (7) folgt unter den genannten Voraussetzungen: $P(A) \leq P(B)$ und $P(B) \leq P(A)$, also insgesamt die Behauptung (siehe auch Satz (5.9.1-05)).

Man kann sich die Regeln wieder gut mit Mengen-Diagrammen oder mit Strecken veranschaulichen, wenn man „Aussage“ durch „Fläche“ bzw. „Strecke“ und P durch „Flächeninhalt“ bzw. „Länge“ ersetzt:


A
B


A
"Rest"
B
0
1

6 Vorgänge

Zusammenfassung
In diesem Kapitel wird eine wichtige Anwendung untersucht, nämlich die Wahrscheinlichkeiten von Ergebnissen bei Vorgängen. Vorgänge können Experimente oder passive Beobachtungen sein. Wie üblich werden Ereignisse als Zusammenfassung mehrerer Ergebnisse definiert. Zum Schluss werden Vorgänge betrachtet, die sich aus mehreren Vorgängen zusammensetzen, wie z. B. dem mehrfachen Würfeln.

* * *

6.1 Ergebnisse

(6.1-01) ***Definition***
Betrachtet man bei einem Vorgang mit einem definierten Ende einen bestimmten Aspekt, so nennt man jeden möglichen Ausgang dieses Vorgangs bezüglich dieses Aspektes ein **Ergebnis** des Vorgangs.
Die **Menge aller möglichen Ergebnisse** nennt man auch **Ergebnismenge** oder **Ereignisraum** für diesen Aspekt des Vorgangs, symbolisiert durch S oder **Ω**. S ist eine endliche Menge.
Die Ergebnisse eines Vorgangs müssen so definiert sein, dass bei dem Vorgang nur genau ein Ergebnis tatsächlich eintreten kann.

(6.1-02) ***Erläuterungen***
1. **Ω** ist der griechische Buchstabe „großes Omega“ und ist der letzte Buchstabe im griechischen Alphabet.

2. Die Ergebnismenge S ist nur im Zusammenhang mit einem klar beschriebenen Vorgang und dem dabei betrachteten Teilaspekt interpretierbar.

3. Es werden nur endliche Ergebnismengen betrachtet, da es in der Realität aufgrund der beschränkten Messgenauigkeit und des beschränkten Messbereichs stets nur endlich viele Ergebnisse geben kann. Lässt man auch unendliche Mengen zu, so müssen die Definitionen und Sätze des mathematischen Gebietes „Maßtheorie“ berücksichtigt werden, aufgrund derer manche Teilmengen von überabzählbaren Mengen nicht zulässig sind, da sie nicht messbar sind.

4. Ein „mögliches Ergebnis“ ist ein Ergebnis, das man für möglich hält. Wenn man mehr Wissen hat, so kann es sein, dass man die Menge S verkleinern kann, da manche Ergebnisse tatsächlich nicht auftreten können oder dass man die Menge S vergrößern muss, da man manche Ergebnisse, die auftreten können, vorher nicht berücksichtigt hat. Der Ereignisraum ist also nicht unbedingt eindeutig, sondern hängt vom Wissen und der Erfahrung ab und auch davon, was interessiert. Gerade wenn man sich in neue Untersuchungsgebiete begibt, kann es sein, dass man S zunächst falsch bestimmt.

5. Die Ergebnisse müssen **überschneidungsfrei** und **vollständig** sein, damit bei einem Vorgang nur genau ein Ergebnis eintreten kann:

- Überschneidungsfreiheit: es können nicht mehrere Ergebnisse, sondern nur ein Ergebnis aus S eintreten
- Vollständigkeit: jedes mögliche Ergebnis des Vorgangs liegt in S

6. Oft wird der Begriff „Zufall“ bei der Betrachtung von Vorgängen und ihren Ergebnissen benutzt. Man benutzt den Begriff Zufall, wenn man keine kausale Erklärung für ein Ergebnis hat. Dabei werden zwei Aspekte diskutiert:
   - Der Zufall existiert in der Realität
   - Der Zufall existiert nicht in der Realität, sondern ist nur Ausdruck mangelnden Wissens.

   Im Rahmen der Stochastik ist diese Unterscheidung belanglos: in jedem Fall weiß man nicht genug, um eine sichere Aussage über das Ergebnis machen zu können. Ob das daran liegt, dass das prinzipiell nicht geht oder daran liegt, dass das prinzipiell ginge, man aber im konkreten Fall nicht alle benötigten Informationen hat, ist für die Bewertung egal. Auch hier ist die Einschätzung personenabhängig: was für den einen zufällig erscheint, kann für den anderen zwangsläufig sein.

7. In der bisherigen Darstellung spielte der **Zufall** keine Rolle. Alle Definitionen und Regeln wurden ohne einen Zusammenhang zum Zufall abgeleitet. Der Zufall spielt also keine grundlegende Rolle in der Stochastik und tritt höchstens bei ganz bestimmten Beispielen auf. Eine Betrachtung vor dem Referendum am 23.06.2016, ob die Bevölkerung von Großbritannien für oder gegen den Brexit stimmen wird, hat mit Zufall wenig zu tun und trotzdem kann man davon eine Wahrscheinlichkeit angeben. Niemand wird sagen „Die Briten haben zufällig für den Brexit gestimmt“ oder „Großbritannien scheidet zufällig aus dem Euro aus“, während „ich habe zufällig eine 6 gewürfelt“ problemlos ist. Stochastik als Wissenschaft des Zufalls zu bezeichnen, ist eine Reduzierung auf ganz bestimmte Einzelfälle und geht am Kern der Stochastik vorbei.

8. Statt „Vorgang“ werden auch oft die Begriffe „**(Zufalls)experiment**“ oder „**(Zufalls)versuch**“ benutzt. Diese Begriffe beschreiben aber nur ganz bestimmte Vorgänge, da sie aktives Eingreifen beinhalten. Oft geht es aber nur um das Beobachten von Vorgängen und Erheben von Daten, ohne den Prozess zu beeinflussen. Das reine Beobachten von Prozessen, wie dem Verhalten von Menschen, wirtschaftlichen Prozessen oder Vorgängen in der Natur, ist aber kein Experiment. Die Beobachtung des Brexit wird man wohl kaum als Experiment bezeichnen.

9. Beobachtet man ähnliche Vorgänge, so hat man sowohl konstante, als auch veränderliche Rahmenbedingungen, die dazu führen, dass es unterschiedliche Ergebnisse geben kann. Wären alle Bedingungen konstant, so würde auch stets dasselbe Ergebnis herauskommen (Ausnahmen gibt es möglicherweise in der Quantentheorie). Die häufig zu findende Aussage „Eine Beobachtung wurde mehrmals **unter gleichen Bedingungen** gemacht“ oder: „Ein Experiment wurde mehrmals unter gleichen Bedingungen durchgeführt“ ist also zu interpretieren als „Eine Beobachtung wurde mehrmals gemacht / ein Experiment wurde mehrmals durchgeführt, wobei bestimmte Bedingungen, nämlich … <genaue Beschreibung> … konstant blieben, während alle anderen Bedingungen (welche auch immer das sein mögen) nicht betrachtet wurden“. Solche Annahmen werden z. B. beim „Gesetz der großen Zahlen“ gemacht.

(6.1.-03) ***Beispiel***

Beim Wetter interessiert oft nur der Teilaspekt „Regen“ und die Beobachtungsergebnisse

r = „es regnet“
n = „es regnet nicht“

Die Menge der möglichen Ergebnisse ist dann S = {r, n}.
Als Gärtner interessiert der Aspekt Regen detaillierter, zum Beispiel

r_k = „es hat in meinem Garten gestern k Liter/m² geregnet“ mit k = 0, 1, 2, …
Die Menge der möglichen Ergebnisse ist dann

$S = \{r_0, r_1, r_2, \dots\}$ oder kürzer: $S = \{0, 1, 2, \dots\}$

Andere Personen interessieren sich beim Wetter vielleicht eher für die Teilaspekte Windstärke oder Sonnenstunden.

(6.1.-04) ***Beispiel***
Beim Spiel „Mensch ärgere Dich nicht“ interessiere der Teilaspekt „Würfelergebnis“.

1. Zu Beginn interessiert zunächst nur:
   s = „es wird eine 6 gewürfelt“
   k = „es wird keine 6 gewürfelt“
   = „es wird eine 1 oder 2 oder 3 oder 4 oder 5 gewürfelt“

   Die Menge der möglichen Ergebnisse ist dann
   $S = \{s, k\}$

2. Im weiteren Verlauf ist dann zunächst jedes Würfelergebnis wichtig
   e = „es wird eine 1 gewürfelt“
   z = „es wird eine 2 gewürfelt“
   …
   s = „es wird eine 6 gewürfelt“

   Die Menge der möglichen Ergebnisse ist dann:
   $S = \{e, z, d, v, f, s\}$.

3. Während des Spiels kann es dann für einen Spieler wichtig sein, ob er eine 2 (dann kann er einen Mitspieler rauswerfen) oder eine 5 (dann kann er „ins Haus“) würfelt. Es ist dann
   z = „es wird eine 2 gewürfelt“
   f = „es wird eine 5 gewürfelt“
   w = „es wird weder 2 noch 5 gewürfelt“
   = „es wird eine 1 oder 3 oder 4 oder 6 gewürfelt“

   Die Menge der möglichen Ergebnisse ist dann:
   $S = \{z, f, w\}$

4. Wenn man den Spielverlauf ganz genau dokumentieren will, dann hat jedes Ergebnis z. B. die Struktur:
   (Uhrzeit; Spieler; Augenzahl; gesetzte Figur),
   oder man dokumentiert die Position der Figuren nach jeder Veränderung.

Die Beispiele zeigen: zu einem Vorgang kann es je nach Fragestellung oder Interessenlage unterschiedliche Ergebnismengen geben. Im Allgemeinen nimmt man eine Ergebnismenge, die

- alle möglicherweise interessierenden Ergebnismengen abdeckt
- die alles, was voraussichtlich nicht interessiert, nicht berücksichtigt

Das erreicht man am einfachsten, indem man das, was man direkt beobachtet, als Ergebnisse definiert und dabei das, was nicht interessiert, weglässt. Andere Definitionen von Ergebnissen sind dann äquivalent zu Mengen von diesen direkt beobachteten Ergebnissen. Dieser Aspekt wird im Kapitel 7.1 Zufallsvariable näher untersucht.

Im Beispiel (6.1.-04) ergibt sich dann:

Die einfachste umfassende Ergebnismenge lässt sich als Menge der möglichen Augenzahlen schreiben (ungültige Würfe werden wieder ignoriert):

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

Die Ergebnisse sind dann in den verschiedenen Fällen äquivalent zu Teilmengen von S:

1. Das Ergebnis s ist äquivalent zu $E_1 = \{6\}$
   Das Ergebnis k ist äquivalent zu $E_2 = \{1, 2, 3, 4, 5\}$
2. Das Ergebnis e ist äquivalent zu $E_1 = \{1\}$
   …
   Das Ergebnis s ist äquivalent zu $E_6 = \{6\}$
3. Das Ergebnis z ist äquivalent zu $E_1 = \{2\}$
   Das Ergebnis f ist äquivalent zu $E_2 = \{5\}$
   Das Ergebnis w ist äquivalent zu $E_3 = \{1, 3, 4, 6\}$

Es ist also sinnvoll, eine Menge von Ergebnissen mit einem eigenen Begriff zu belegen. Da sich etwas ereignet hat, nimmt man den Begriff „Ereignis“.

6.2 Ereignisse

Man könnte zunächst beliebige Mengen von Ergebnissen als Ereignisse bezeichnen und in vielen Fällen wird das auch so gemacht, aber manchmal ist es zweckmäßig nur bestimmte Mengen von Ergebnissen zu betrachten. In Hinsicht auf Kapitel 6.3 müssen auch Durchschnitt, Vereinigung und Komplement von Ereignissen wieder Ereignisse sein. Man definiert daher:

(6.2.-01) ***Definition***

Hat man zu einem Vorgang bezüglich eines bestimmten Aspektes eine endliche Ergebnismenge S definiert und ist $\mathfrak{P}(S)$ die zugehörige **Potenzmenge**, so heißt jedes nichtleere Mengensystem $\mathfrak{A} \subset \mathfrak{P}(S)$ eine **σ-Algebra** auf S, wenn auch Durchschnitt, Vereinigung und Komplement wieder in $\mathfrak{A}$ liegen, also:

Sind $A, B \in \mathfrak{A}$ (also $A, B \subset S$), dann sind auch $A \cup B$, $A \cap B$, $\overline{A} \in \mathfrak{A}$.

Jedes $A \in \mathfrak{A}$ heißt ein **Ereignis**.

- Ist bei einem Vorgang ein Ergebnis $a \in A$ eingetreten, so sagt man: das Ereignis A ist **eingetreten** oder das Ereignis A hat **stattgefunden**.
- Besteht A aus nur einem Element, so heißt A ein **Elementarereignis**.
- Das Komplement von A heißt **Gegenereignis** zu A, symbolisiert durch $\overline{A}$.
- Die leere Menge $\emptyset$ heißt **unmögliches Ereignis**.
- Die Ergebnismenge S heißt **sicheres Ereignis**.

(6.2.-02) ***Erläuterungen***

1. Die konkrete Bedeutung der Begriffe „Ereignis“, „Gegenereignis“, „unmögliches Ereignis“ und „sicheres Ereignis“ sind stets an einen ganz bestimmten Vorgang mit einer Menge S gekoppelt.

2. Je nach Zusammenhang kann das Ereignis in der Vergangenheit, Gegenwart oder Zukunft liegen. In der Definition ist egal, „ist eingetreten“, „tritt ein“ oder „wird eintreten“ steht.

3. Beziehung zu Aussagen
   a. Ereignisse können als Aussagen geschrieben werden durch:
      A = das Ereignis <Beschreibung> ist eingetreten / tritt ein / wird eintreten
      oder
      A = das Ereignis <Beschreibung> hat stattgefunden / findet statt / wird stattfinden
      Damit kann alles, was über Aussagen und ihre Wahrscheinlichkeiten dargestellt wurde, auf Ereignisse übertragen werden.
   b. Aussagen, die keine Ereignisse beschreiben, sind z. B. zeitlose Feststellungen wie „2 + 3 = 5“ oder Naturgesetze wie „$E = m \cdot c^2$“ oder „im Zentrum der Milchstraße befindet sich ein schwarzes Loch“.
   c. Da im Rahmen der Stochastik Aussagen nur wahr oder falsch sein können, können Ereignisse stets nur eintreten oder nicht eintreten, nichts dazwischen.
   d. Da Verknüpfungen von Aussagen mit „und“, „oder“ oder „nicht“ wieder Aussagen ergeben, ist es sinnvoll, dass auch Vereinigung, Durchschnitt und Komplement von Ereignissen wieder Ereignisse sind (siehe auch (6.2.-07)).

4. Besteht die Menge S aus n Elementen, so besteht die Potenzmenge $\mathfrak{P}(S)$ aus 2^n Elementen. Denn will man eine Teilmenge $A \subset S$ bilden, so kann man für jedes der n Elemente von S entscheiden, ob es zu A gehören soll oder nicht. Und da diese n Entscheidungen unabhängig voneinander sind, gibt es insgesamt $2 \cdot 2 \cdot \ldots \cdot 2 = 2^n$ Entscheidungen, wie die Menge A gebildet werden soll und damit gibt es 2^n verschiedene Teilmengen.

5. Besteht ein Ereignis aus nur einem Element, so heißt es Elementarereignis. Aber umgekehrt muss nicht jede einelementige Teilmenge von S ein Elementarereignis sein, wie das folgende Beispiel zeigt:
   $S = \{1, 2, 3, 4, 5, 6\}$
   $\mathfrak{A} = \{\emptyset, A, \overline{A}, S\}$ mit $A = \{1, 2, 3\}$ und $\overline{A} = \{4, 5, 6\}$
   $\mathfrak{A}$ ist eine σ-Algebra, enthält aber keine Elementarereignisse. Z. B. ist $\{1\}$ kein Ereignis, also auch kein Elementarereignis.

6. Wählt man im Teil 1 des Würfelbeispiels (6.1.-04) …
   - $S = \{s, k\}$, so sind $\{s\}$ und $\{k\}$ die Elementarereignisse
   - $S = \{1, 2, 3, 4, 5, 6\}$, so sind $\{1\}, \ldots, \{6\}$ die Elementarereignisse.

7. Ein unmögliches Ereignis ist z. B. beim Würfeln das Ereignis „es wurde eine 7 gewürfelt“ oder bei der Wetterbeobachtung das Ereignis „hier und jetzt regnet es und regnet es nicht“ (siehe Beispiel (6.1.-03)). Betritt man dagegen Neuland, so kann es Überraschungen geben und es treten Ereignisse ein, von denen man überzeugt war, dass sie unmöglich sind. Siehe z. B.
   http://www.desy.de/infos__services/presse/pressemeldungen/2013/pm_191213/index_ger.html
   (01.02.2017).

8. Ob ein Ereignis möglich oder unmöglich ist, hängt auch vom Vorgang ab: „es wurde Wappen geworfen“ ist beim Werfen einer Münze möglich, während es beim Würfeln unmöglich ist.

9. Ein sicheres Ereignis ist z. B. beim Würfeln das Ereignis „es wurde irgendetwas gewürfelt“ (wenn man ungültige Würfe ignoriert) oder bei der Wetterbeobachtung das Ereignis „es regnet oder es regnet nicht“.

10. Bei den Begriffen Ergebnis und Elementarereignis herrscht in der Literatur Uneinheitlichkeit.

a. Im vorliegenden Buch wird die Begrifflichkeit aus der Mengenlehre angewendet, bei der Elemente und Mengen sauber unterschieden werden:

Ergebnis = Element von S
Ereignis = Teilmenge von S, Element von $\mathfrak{A}$
Elementarereignis = Ereignis, das nur aus einem Element besteht

Diese Festlegung hat den Vorteil, dass es immer um Mengen geht, wenn von Ereignissen die Rede ist. Insbesondere sind Elementarereignisse auch Ereignisse.

b. Im Axiomensystem nach Kolmogorff ist (siehe Kapitel 8.1 Kolmogoroff):

Elementares Ereignis = Element von S
Zufälliges Ereignis = Teilmenge von S

Kein eigener Begriff für zufällige Ereignisse, die nur aus einem Element bestehen.

c. In manchen Büchern wird definiert

Ergebnis = Elementarereignis = Element von S (also keine Teilmenge von S)
Ereignis = Teilmenge von S

In diesem Fall sind Elementarereignisse keine Ereignisse und daher sollten solche Begriffsbildungen vermieden werden.

Die Verwendung des Begriffs „Ereignis“ in der Stochastik entspricht nur begrenzt dem umgangssprachlichen Gebrauch.

- Umgangssprachlich ist ein Ereignis etwas Außergewöhnliches. Man spricht z. B. von einem Naturereignis, gesellschaftlichen Ereignis oder freudigen Ereignis. Etwas Belangloses wie „ein Staubkorn ist auf den Boden gefallen“ wird man wohl kaum als Ereignis bezeichnen.
- Umgangssprachlich kann es verschiedene Beschreibungen eines Ereignisses geben, obwohl nur <u>ein</u> Ereignis stattfand. Nach einem Konzert wird jeder eine etwas andere Beschreibung abgeben, obwohl sich alle einig sind, von demselben Ereignis zu sprechen. In der Stochastik dagegen bezeichnet man die Beschreibung eines realen Ereignisses mit einer Menge bereits als „Ereignis“, sodass es zu einem realen Ereignis viele verschiedene stochastische Ereignisse geben kann. Außerdem ist es in der Stochastik belanglos, ob das Ereignis außergewöhnlich oder völlig belanglos war.

(6.2.-03) ***<u>Beispiel</u>***

Einmaliges Würfeln mit dem Ergebnis Augenzahl = 4 und S = {1, 2, 3, 4, 5, 6}.

- Exakte Beschreibung: „Es wurde eine 4 gewürfelt“, kurz: E = {4}
- Grobe Beschreibung: „Es wurde eine gerade Zahl gewürfelt“, kurz: E = {2, 4, 6}
- Sehr grobe Beschreibung: „Es wurde keine 5 gewürfelt“, kurz: E = {1, 2, 3, 4, 6}
- Nichtssagende Beschreibung: „Es wurde irgendwas gewürfelt“, kurz: E = {1, 2, 3, 4, 5, 6}

Jede dieser Beschreibungen desselben realen Ereignisses ist in der Stochastik ein „Ereignis“. Da es auf Basis von S genau $2^5 = 32$ solche Beschreibungen / Mengen gibt, die die Zahl 4 enthalten, sind demnach gleichzeitig 32 Ereignisse eingetreten – und das beim einmaligen Würfeln! Für einen Nicht-Stochastiker klingt das ziemlich schräg, aber man gewöhnt sich daran.

(6.2.-04) ***<u>Checkliste</u>***

Um festzustellen, ob ein Ereignis eingetreten ist, muss vorher klar definiert werden:

- Welcher Vorgang wird beobachtet? -> Beschreibung des Vorgangs
- Welche Ergebnisse sind möglich? -> Definition von S
- Welches Ereignis wird betrachtet? -> Beschreibung Ereignis $\subset$ S
- Wer stellt fest, dass das Ereignis eingetreten ist? -> Verantwortlicher

- Wann wird festgestellt, dass das Ereignis eingetreten ist? -> Zeitpunkt
- Wie wird festgestellt, dass das Ereignis eingetreten ist? -> Methode

(6.2.-05) ***Beispiel***
Checkliste, um beim Fußball ein Tor festzustellen

- Vorgang: „Schuss aufs Tor beim Fußballspiel“
- Ergebnismenge: S = {„Tor“, „Ecke“, „Abschlag“, „Elfmeter“, „Freistoß“, „Abpfiff“}
- Ereignis: E = {„Tor“}
- Verantwortlicher: „der Schiedsrichter“
- Zeitpunkt: „vor der Freigabe des Balles zum Weiterspielen“
- Methode: „der Schiedsrichter folgt seiner eigenen Wahrnehmung oder der Wahrnehmung des Linienrichters“

Dabei ist unerheblich, was tatsächlich passiert ist, also ob z. B. der Ball tatsächlich im Tor war.

(6.2.-06) ***Beispiel***
Bei der Wette, dass außerirdisches Leben existiert, muss klar angegeben werden, wer bis wann offiziell feststellen muss, dass außerirdisches Leben existiert.
Siehe http://www.wettanbieter.cc/wettanbieter-news/verruckte-wetten-auserirdische-und-yetis-beim-britischen-wettanbieter-william-hill.html (01.02.2017)

(6.2.-07) ***Mengenoperationen***
Wie in der Mengenlehre üblich können die bei den Aussagen bereits benutzten logischen Operationen „und“, „oder“ und „nicht“ durch entsprechende Mengenoperationen dargestellt werden.
Sind A und B Ereignisse, so ist:

$\overline{A}$	→ $\overline{A} = S \setminus A$	Komplement zu A, also $\overline{A}$ = Menge der Ergebnisse, die nicht zu A gehören (Gegenereignis)
A und B	→ $A \cap B$	Schnittmenge = Menge der Ergebnisse, die sowohl zu A als auch zu B gehören
A oder B	→ $A \cup B$	Vereinigungsmenge = Menge der Ergebnisse, die zu A oder zu B gehören
$A \Rightarrow B$	→ $A \subset B$	A ist Teilmenge von B: wenn ein Element zu A gehört, dann gehört es auch zu B

(6.2.-08) ***Regeln***
Es gelten die Regeln und Begriffe der Mengenlehre, wie z. B.:

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

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

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

$A \cup \overline{A} = S$

$|A|$ = Anzahl der Elemente der Menge A

(6.2.-09) ***Hinweis***

Um aus einer sprachlichen Formulierung zu einer Ergebnismenge zu kommen, übersetzt man die einzelnen Elemente der Formulierung in Mengen, verknüpft sie und schneidet dann mit S. Unterschiedliche sprachliche Formulierungen können zu derselben Menge führen.
Beispiele:
Es wird mit einem Würfel einmal gewürfelt mit S = {1, 2, 3, 4, 5, 6}. Dann ist:

- „es wurde eine 7 gewürfelt“ → $\{7\} \cap S = \emptyset$
- „es wurde gleichzeitig eine 3 und eine 4 gewürfelt“ → $(\{3\} \cap \{4\}) \cap S = \emptyset$
- „es wurde eine Zahl > 3 gewürfelt“ → $]3; \infty[\cap S = \{4, 5, 6\}$

Es gibt nur eine leere Menge, aber viele sprachliche Beschreibungen für ein unmögliches Ereignis. Die leere Menge ∅ repräsentiert „kein Ergebnis aus der Menge S“. Die sprachlich verschiedenen Darstellungen in den ersten beiden Fällen werden zunächst durch unterschiedliche Mengen repräsentiert, die sich dann aber zu der einen leeren Menge vereinfachen lassen.

(6.2.-10) ***Beispiel***
Das folgende Beispiel zeigt, warum es notwendig ist, dass bei einem Vorgang nur genau ein Ergebnis eintreten darf.
Beim Spiel „Mensch ärgere dich nicht“ seien für das Setzen der einen Figur die Augenzahlen „2“ oder „4“ von Bedeutung, während für die andere Figur desselben Spielers „2“ oder „5“ wichtig sind. Es interessieren in diesem Fall also nur die Ergebnisse

a = „es wurde 2 oder 4 gewürfelt“
b = „es wurde 2 oder 5 gewürfelt“
c = Rest = „es wurde eine 1, 3, oder 6 gewürfelt“

also S = {a, b, c}.
Die Ergebnisse a und b können bei einem Vorgang gleichzeig eintreten, nämlich wenn eine 2 gewürfelt wurde. Andererseits ist formal wegen $a \neq b$:

$$\{a\} \cap \{b\} = \emptyset$$

also das unmögliche Ereignis und das ist ein Widerspruch.
Daher wird der Fall, dass zwei Ergebnisse bei einem Vorgang gleichzeitig auftreten können, in der Definition von „Ergebnis“ bereits ausgeschlossen.

6.3 Wahrscheinlichkeit eines Ereignisses

Die abkürzende Schreibweise für P(„das Ereignis E ist eingetreten“) ist: P(E)
Die Wahrscheinlichkeit P bildet also in diesem Fall eine Menge von Ergebnissen auf eine reelle Zahl ab.

(6.3.-1) ***Definition***
Ist ein Vorgang mit der endlichen Ergebnismenge S, der σ-Algebra $\mathfrak{A}$ und der Wahrscheinlichkeit P gegeben, so nennt man das Tripel (S, $\mathfrak{A}$, P) einen **endlichen Wahrscheinlichkeitsraum für Ereignisse**, ab jetzt kurz: **Wahrscheinlichkeitsraum.**

(6.3.-2) ***Hinweise***

- Die Definition ergibt sich unmittelbar aus der Definition des Wahrscheinlichkeitsraumes für Aussagen, wenn man Aussagen durch Ereignisse ersetzt.
- Die Definition der σ-Algebra ist plausibel, wenn man berücksichtigt, dass die Subadditivität erfüllt sein muss.
- Die komplette anwendungsbezogene Beschreibung wäre
  (Vorgang, Aspekt, S, 𝔄, P),
  aber für den rechnerischen Anteil der Betrachtung genügt (S, 𝔄, P) und so wird der Wahrscheinlichkeitsraum auch üblicherweise definiert. Die praktische Interpretation ist aber natürlich nur möglich, wenn man auch den Vorgang und den dabei betrachteten Aspekt einbezieht.
- Da S die Menge aller möglichen Ergebnisse ist, ist auch jedes Ereignis $A \neq \emptyset$ eine Menge von möglichen Ergebnissen, also $P(A) \neq 0$. Für jedes Ereignis $A \neq S$ gilt analog $P(A) \neq 1$ (siehe (6.3.-3), Rechenregel (1)). In der Definition nach Kolmogoroff (siehe Kapitel 8.1) muss das nicht so sein.
- Ein formaler Unterschied zwischen den beiden Definitionen (5.9.1-02) und (6.3.-1) ist:
  - Aussagen: ist $A \in S$, so gilt: $A \in K$ und $P(A)$ existiert
  - Ereignisse: ist $a \in S$, so kann (muss aber nicht) gelten: $\{a\} \in K$ und $P(\{a\})$ existiert

Aus den bisherigen Rechenregeln für die Wahrscheinlichkeiten von Aussagen folgen die

(6.3.-3) ***<u>Rechenregeln</u>*** für die Wahrscheinlichkeiten von Ereignissen
Es sei ein bestimmter Vorgang mit dem Wahrscheinlichkeitsraum (S, 𝔄, P) und der Ergebnismenge $S = \{a_1, \ldots, a_n\}$ gegeben. A, B, C, A_i seien Elemente von 𝔄, also Ereignisse. Dann folgt aus den Rechenregeln in Kapitel 5.11:

(1) **$A = \emptyset$** ⇒ A enthält keine möglichen Ergebnisse ⇒ **$P(A) = P(\emptyset) = 0$**
$A = S$ ⇒ A enthält alle möglichen Ergebnisse ⇒ **$P(A) = P(S) = 1$**
$A \neq \emptyset$ ⇒ A enthält mögliche Ergebnisse ⇒ **$P(A) \neq 0$**
$A \neq S$ ⇒ A enthält nicht alle möglichen Ergebnisse ⇒ **$P(A) \neq 1$**

(2) **$P(A \cup B) = P(A) + P(B) - P(A \cap B)$**
$P(A \cup B) = P(A) + P(B)$ ⇔ **$P(A \cap B) = 0$**
⇔ $A \cap B = \emptyset$ (siehe Hinweis (6.3.-2))
⇔ A und B können nicht gleichzeitig eintreten

(3) **$P(\bar{A}) = 1 - P(A)$**

(4) **$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$**

(5) Hat man k Ereignisse A_i ($i = 1, \ldots, k$), die sich paarweise gegenseitig ausschließen – also $A_i \cap A_j = \emptyset$ für $i \neq j$ –, so folgt:
$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$

(6) Hat man k Ereignisse A_i ($i = 1, \ldots, k$), die sich paarweise gegenseitig ausschließen und sind alle Einzelwahrscheinlichkeiten gleich, also $P(A_1) = P(A_2) = \ldots = P(A_k)$, so ist für jedes i:
$P(A_i) = \frac{P(A_1 \cup A_2 \cup \ldots \cup A_k)}{k}$ $\quad (i = 1, \ldots, k)$
<u>Sonderfall</u>: Ist 𝔄 = 𝔓(S) und sind die Wahrscheinlichkeiten aller n Elementarereignisse gleich, so ist für jedes i:

$P(\{a_i\}) = \frac{1}{n}$ (i = 1, …, n)
Diese Formel ergibt sich, wenn man in der allgemeinen Formel k = n und $A_i = \{a_i\}$ setzt und $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(S) = 1$ berücksichtigt.

(7) **$A \subset B \Rightarrow P(A) \leq P(B)$**

(8) P(A|B) ist die Wahrscheinlichkeit für das Eintreten des Ereignisses A, wenn man zusätzlich weiß oder voraussetzt, dass das Ereignis B eingetreten ist.
$P(A|B) = \frac{P(A \cap B)}{P(B)}$; $P(B) \neq 0$

Während Aussagen sehr unterschiedlich sein können, findet bei Ereignissen eine Reduktion auf Ergebnisse von Vorgängen statt, also eine zusätzliche Abstraktion und damit Vereinfachung. Verschiedene Aussagen können zu demselben Ergebnis führen. Eine „1“ beim Würfeln ist etwas anderes als eine „1“ bei einer Untersuchung über „Anzahl Kinder“, aber das Ergebnis ist in beiden Fällen formal dasselbe. Bei der Berechnung von Wahrscheinlichkeiten muss dieser Kontext dann natürlich wieder berücksichtigt werden. Da diese Vereinfachungen häufig ausreichen und da man mit Mengen einfacher umgehen kann als mit Aussagen, werden oft Ereignisse statt Aussagen betrachtet.

(6.3.-4) ***<u>Hinweis</u>***
Ist $\mathfrak{A} = \mathfrak{P}(S)$, so gilt für jedes Ereignis mit geeigneten Ergebnissen a_i:

$$P(A) = P(\{a_1, a_2, a_3, \ldots\}) = P(\{a_1\} \cup \{a_2\} \cup \{a_3\} \cup \ldots) = P(\{a_1\}) + P(\{a_2\}) + P(\{a_3\}) + \ldots$$

In diesem Fall ergibt sich als Methode zur Bestimmung der Wahrscheinlichkeit eines Ereignisses:

- <u>Erst</u> den Vorgang und den betrachteten Aspekt definieren
- <u>dann</u> die Ergebnismenge S mit der σ-Algebra $\mathfrak{A} = \mathfrak{P}(S)$ bestimmen,
- <u>dann</u> die Wahrscheinlichkeiten der Elementarereignisse mit dem gleichen Bewertungsverfahren bestimmen, also den Wahrscheinlichkeitsraum definieren,
- <u>dann</u> die Wahrscheinlichkeit für das betrachtete Ereignis ausrechnen

Die einzige Herausforderung besteht in diesem Fall also darin, die Wahrscheinlichkeiten der Elementarereignisse zu bestimmen. Auch hieran sieht man, dass die Beschränkung auf Vorgänge und ihre Ergebnismenge erhebliche Vereinfachungen nach sich zieht.

Ein historisch wichtiger Sonderfall ist der Fall, dass einfach die relative Häufigkeit als Wahrscheinlichkeit genommen wird:

(6.3.-5) ***<u>Definition</u>***
Es wird ein Vorgang mit dem folgenden Wahrscheinlichkeitsraum betrachtet:

- Es gibt n mögliche Ergebnisse, also $S = \{e_1, \ldots, e_n\}$, $|S| = n$, $n \in \mathbb{N}$.
- $\mathfrak{A} = \mathfrak{P}(S)$
- Jedes Elementarereignis hat dieselbe Wahrscheinlichkeit $P(\{e_i\}) = \frac{1}{n}$

Ferner sei E ein Ereignis, das aus k möglichen Ergebnissen besteht, also |E| = k. Dann gilt, dass die Wahrscheinlichkeit für E gleich der relativen Häufigkeit der Ergebnisse aus E bezogen auf die Gesamtheit ist, also:

$$P(E) = \frac{|E|}{|S|} = \frac{k}{n}$$

Dies ist die **Laplace'sche Formel der Wahrscheinlichkeit** und wird kurz beschrieben als:

$$P(E) = \frac{\text{Anzahl der für E günstigen Fälle}}{\text{Anzahl möglicher Fälle}} = \frac{|E|}{|S|}$$

Die Laplacesche Formel ist einfach eine relative Häufigkeit. Es ist erstaunlich, dass diese Tatsache in Veröffentlichungen meistens nicht erwähnt wird. Ob man diese relative Häufigkeit dann als Wahrscheinlichkeit interpretiert oder nicht, ist für das Rechenverfahren belanglos.

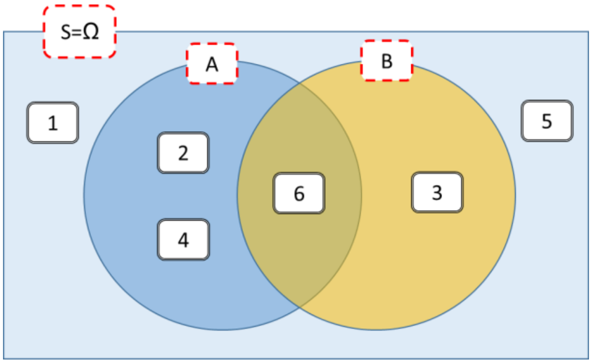
Typische Anwendungsbeispiele sind der Münzwurf, das Würfeln oder die Ziehung der Lottozahlen, sofern man von bestimmten Annahmen ausgeht (siehe Kapitel 8.2).

(6.3.-6) ***Beispiel*** zur bedingten Wahrscheinlichkeit
Einmaliges Würfeln mit einem Würfel, ungültige Würfe werden ignoriert,

$S = \{1, 2, 3, 4, 5, 6\}$, $\mathfrak{A} = \mathfrak{P}(S)$,

jedes Elementarereignis habe die Wahrscheinlichkeit $= \frac{1}{6}$.

Gegeben sind die Ereignisse: $A = \{2, 4, 6\}$; $B = \{3, 6\}$



P(A|B):

Weiß man, dass das Ereignis B eingetreten ist, so gibt es zwei gleichwahrscheinliche Möglichkeiten:

1. Es ist eine 3 gewürfelt worden, A ist also nicht eingetreten
2. Es ist eine 6 gewürfelt worden, A ist also eingetreten

Weiß man, dass das Ereignis B eingetreten ist, so ist demnach die Wahrscheinlichkeit, dass auch das Ereignis A eingetreten ist: $P(A|B) = \frac{1}{2}$.

Wegen $P(A \cap B) = P(\{6\}) = \frac{1}{6}$ und $P(B) = \frac{2}{6}$ ist tatsächlich erfüllt: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}$.

P(B|A):

Weiß man, dass das Ereignis A eingetreten ist, so gibt es drei gleichwahrscheinliche Möglichkeiten:

1. Es ist eine 2 gewürfelt worden, B ist also nicht eingetreten
2. Es ist eine 4 gewürfelt worden, B ist also nicht eingetreten
3. Es ist eine 6 gewürfelt worden, B ist also eingetreten

Weiß man, dass das Ereignis A eingetreten ist, so ist demnach die Wahrscheinlichkeit, dass auch das Ereignis B eingetreten ist: $P(B|A) = \frac{1}{3}$ (nämlich in nur einem von 3 möglichen Fällen)

Wegen $P(B \cap A) = P(\{6\}) = \frac{1}{6}$ und $P(A) = \frac{3}{6}$ ist tatsächlich erfüllt: $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$.

(6.3.-7) ***Beispiel*** zur stochastischen Unabhängigkeit
Einmaliges Würfeln mit einem Würfel, ungültige Würfe werden ignoriert,

$S = \{1, 2, 3, 4, 5, 6\}$, $\mathfrak{A} = \mathfrak{P}(S)$,

jedes Elementarereignis habe die Wahrscheinlichkeit = $\frac{1}{6}$.

Es werden folgende Ereignisse betrachtet:

A = „es wird eine durch 2 teilbare Zahl gewürfelt“, kurz: $A = \{2, 4, 6\}$
B = „es wird eine durch 3 teilbare Zahl gewürfelt“, kurz: $B = \{3, 6\}$
C = „es wird eine durch 4 teilbare Zahl gewürfelt“, kurz: $C = \{4\}$

Es ist:

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

A und B sind also voneinander stochastisch unabhängig
A und B sind auch logisch voneinander unabhängig: wenn eine Zahl durch 2 teilbar ist, so kann sie durch 3 teilbar sein oder nicht; wenn eine Zahl nicht durch 2 teilbar ist, so kann sie durch 3 teilbar sein oder nicht; die Teilbarkeiten durch 2 und durch 3 sind also unabhängig voneinander.

Es ist:

$$P(A \cap C) = P(\{4\}) = \frac{1}{6} \neq \frac{3}{6} \cdot \frac{1}{6} = P(A) \cdot P(C),$$

A und B sind also voneinander stochastisch abhängig
A und B sind auch logisch voneinander abhängig: wenn eine Zahl durch 2 teilbar ist, kann sie durch 4 teilbar sein oder nicht; wenn aber eine Zahl nicht durch 2 teilbar ist, kann sie erst recht nicht durch 4 teilbar sein; die Teilbarkeiten durch 2 und durch 4 sind also abhängig voneinander.

Da nur einmal gewürfelt wird, beeinflussen sich die Ereignisse nicht, da sie alle gleichzeitig eintreten bzw. nicht eintreten.

6.4 Kombinationen von Vorgängen

Bislang haben wir einen Vorgang mit einer Ergebnismenge betrachtet.

In den vorangegangenen Kapiteln wurden Aussagen betrachtet, die durch Verknüpfung von Aussagen entstanden sind, und es wurde ihr gemeinsames Bewertungsverfahren bestimmt. Analog erhält man einen neuen Vorgang, wenn man mehrere Vorgänge, bei denen jeder seine eigene Ergebnismenge hat, kombiniert.

(6.4.-1) ***Beispiel***
Vorgang 1 = einmaliger Wurf eines Würfels. $S_1 = \{1, 2, 3, 4, 5, 6\}$
Vorgang 2 = einmaliger Wurf einer Münze. $S_2 = \{w, z\}$
Die Ergebnismenge des Gesamtvorganges kann dann durch die Menge

$$S = S_1 \times S_2 = \{(e_1; e_2) \mid e_1 \in S_1; e_1 \in S_1\} \text{ mit } |S| = |S_1| \cdot |S_2|$$

dargestellt werden.
Ereignisse bezüglich S sind z. B.:

A = „es wurde eine gerade Zahl gewürfelt und Wappen geworfen“
= $\{(2, w), (4, w), (6, w)\}$
B = „es wurde eine 3 gewürfelt (und das Ergebnis des Münzwurfs ist unerheblich)“

$= \{(3, w), (3, z)\}$

(6.4.-2) ***Beispiel***
Vorgang 1 = einmaliges Würfeln mit einem Würfel. $S_1 = \{1, 2, 3, 4, 5, 6\}$
Vorgang 2 = einmaliges Würfeln mit einem Würfel. $S_2 = \{1, 2, 3, 4, 5, 6\}$

- Wenn die einzelnen Vorgänge hintereinander mit demselben Würfel ausgeführt werden, ist die Zuordnung der Ergebnisse:
  Vorgang 1 – Ergebnis des 1. Wurfes
  Vorgang 2 – Ergebnis des 2. Wurfes.
- Wenn die einzelnen Vorgänge mit verschiedenen Würfeln ausgeführt werden, so müssen die Würfe unterscheidbar sein – z. B. durch die Farbe der Würfel oder durch den Zeitpunkt oder den Ort des Würfelns. Die Zuordnung der Ergebnisse ist dann z. B.:
  Vorgang 1 – Ergebnis des Wurfes mit dem roten Würfel
  Vorgang 2 – Ergebnis des Wurfes mit dem blauen Würfel

In beiden Fällen kann man die Ergebnismenge des Gesamtvorganges wieder darstellen als
$S = S_1 \times S_2 = \{(e_1; e_2) \mid e_1 \in S_1; e_2 \in S_2\}$ mit $|S| = |S_1| \cdot |S_2|$

(6.4.-3) ***Beispiel***
Gibt es logische Abhängigkeiten zwischen den einzelnen Vorgängen, so kann es vorkommen, dass bestimmte Ergebniskombinationen nicht vorkommen.
Ist bei einem Projekt S_i die Ergebnismenge der Projektphase i, also
$S_1 = \{\text{erfolgreich, nicht erfolgreich}\}$
$S_2 = \{\text{erfolgreich, nicht erfolgreich}\}$
so sollte die Kombination $e = (ne_1; e_2) = (\text{nicht erfolgreich, erfolgreich})$ in der Praxis und damit auch in S nicht vorkommen, also $S = \{(e_1; e_2), (e_1; ne_2); (ne_1; ne_2)\}$.

Gegeben seien zwei Wahrscheinlichkeitsräume $(S_i, \mathfrak{A}_i, P_i)$, i = 1,2. Die beiden Bewertungsverfahren seien zusammen widerspruchsfrei, so dass es ein gemeinsames Bewertungsverfahren mit der Wahrscheinlichkeit P gibt. Dann ist für alle Ergebnisse $e_1 \in S_1$ und $e_2 \in S_2$:
$$P(\{e_1\}) = P_1(\{e_1\}) \quad \text{und} \quad P(\{e_2\}) = P_2(\{e_2\})$$
Die Wahrscheinlichkeiten ändern sich durch das Gesamtverfahren mit der übergreifenden Wahrscheinlichkeit nicht.

Dann ist für jedes Ergebnis $(e_1, e_2) \in S$, $S = S_1 \times S_2$:
$$\begin{aligned} P(\{(e_1, e_2)\}) &= P(\text{„}e_1 \text{ ist Ergebnis von Vorgang 1“ und „}e_2 \text{ ist Ergebnis von Vorgang 2“}) \\ &= P(\text{„}e_1 \text{ ist Ergebnis von Vorgang 1“} \mid \text{„}e_2 \text{ ist Ergebnis von Vorgang 2“}) \cdot \\ &\quad\; P(\text{„}e_2 \text{ ist Ergebnis von Vorgang 2“}) \\ &= P(\{e_1\}|(\{e_2\}) \cdot P(\{e_2\}) \end{aligned}$$
Sind die Elementarereignisse $\{e_1\}$ und $\{e_2\}$ stochastisch voneinander unabhängig, dann gilt
$$\begin{aligned} P(\{(e_1, e_2)\}) &= P(\{e_1\}) \cdot P(\{e_2\}) \\ &= P_1(\{e_1\}) \cdot P_2(\{e_2\}) \end{aligned}$$
Man kann also auf die ursprünglichen etwas einfacher gebauten Bewertungsverfahren zurückgreifen.

(6.4.-4) ***Beispiel***
Vorgang 1: einmaliges Würfeln mit einem roten Würfel, ungültige Würfe werden ignoriert,
$S_1 = \{1, 2, 3, 4, 5, 6\}$, $\mathfrak{A}_1 = \mathfrak{P}(S_1)$,

$P_1(\{i\}) = \frac{1}{6}$ (i = 1, 2, 3, 4, 5, 6)

Vorgang 2: einmaliges Würfeln mit einem blauen Würfel, ungültige Würfe werden ignoriert,

$S_2 = \{1, 2, 3, 4, 5, 6\}$, $\mathfrak{A}_2 = \mathfrak{P}(S_2)$,

$P_2(\{1\}) = 0{,}25$; $P_2(\{i\}) = 0{,}15$ (i = 2, 3, 4, 5, 6)

Der blaue Würfel ist auf einer Seite etwas klebrig, daher die ungleichen Wahrscheinlichkeiten.
Dann ist $S = S_1 \times S_2 = \{(1, 1), (1, 2), \ldots, (3, 5), \ldots, (5, 3), \ldots, (6, 5), (6, 6)\}$ mit $|S| = |S_1| \cdot |S_2| = 6 \cdot 6 = 36$.
Das übergreifende Bewertungsverfahren P ist einfach die Vereinigung der beiden Bewertungsverfahren, also:

Wenn der Würfel rot ist, dann gilt P_1, wenn der Würfel blau ist, dann gilt P_2

Wären die Würfel nicht unterscheidbar, würden sich die beiden Bewertungsverfahren widersprechen und man könnte kein gemeinsames Verfahren definieren, das zu beiden Einzelverfahren konform ist.
Geht man davon aus, dass sich die Würfe gegenseitig nicht beeinflussen, so sind die jeweiligen Ereignisse stochastisch unabhängig und es ist z. B.:

$P(\{1, 1)\} = P_1(\{1\}) \cdot P_2(\{1\}) = \frac{1}{6} \cdot 0{,}25 = \frac{1}{24}$.

7 Diskrete Wahrscheinlichkeitsfunktionen

Zusammenfassung
Wahrscheinlichkeitsräume lassen sich durch die Einführung von Zufallsvariablen drastisch vereinfachen, da nur noch einige wenige Ereignisse in der Betrachtung relevant sind. Dabei wird der Begriff Zufallsvariable als Variable und nicht wie üblich als Abbildung definiert.
Mit Hilfe von Zufallsvariablen können Kombinationen bestimmter gleichartiger Vorgänge durch diskrete Wahrscheinlichkeitsfunktionen, wie der hypergeometrischen Verteilung oder der Binomialverteilung, beschrieben werden.
* * *

7.1 Diskrete Zufallsvariable

Es wird wieder eine endliche Ergebnismenge S zugrunde gelegt.

Bei den bisherigen Beispielen wurde deutlich, dass man bei einem Vorgang und dem betrachteten Teilaspekt je nach Interessenlage unterschiedliche Ergebnismengen definieren kann. Beschreiben zwei Ereignisse auf Basis verschiedener Ergebnismengen denselben Sachverhalt, so müssen die jeweiligen Bewertungsverfahren so aufeinander angepasst sein, dass die Wahrscheinlichkeiten gleich sind. In den Fällen, die in diesem Kapitel behandelt werden, ist die Situation:

- Ergebnismenge S_1: viele mögliche Ergebnisse und Ereignisse, aber einfach zu ermittelnde Wahrscheinlichkeiten.
- Ergebnismenge S_2: wenige mögliche Ergebnisse und Ereignisse, aber schwierig zu ermittelnde Wahrscheinlichkeiten.

Der Wahrscheinlichkeitsraum $(S_1, \mathfrak{A}_1, P_1)$ wird benutzt, um die eigentlich interessierenden Wahrscheinlichkeiten von Ereignissen in $(S_2, \mathfrak{A}_2, P_2)$ zu bestimmen.

(7.1.-1) ***<u>Beispiel</u>***
Interessieren beim einmaligen Würfeln mit zwei unterscheidbaren Würfeln – oder gleichwertig: beim zweimaligen Würfeln mit einem Würfel – die genauen Würfelergebnisse, so ist

$$S_1 = \{(a, b) \mid a, b \in \{1, 2, 3, 4, 5, 6\}\} ; \quad |S_1| = 36$$

Interessiert dagegen wie bei manchen Würfelspielen nur die Summe der beiden Augenzahlen, so kann man wählen

$$S_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} ; \quad |S_2| = 11$$

Jeder Teilmenge von S_2 entspricht eindeutig eine Teilmenge von S_1, die Ereignisse von S_2 sind äquivalent zu bestimmten Ereignissen von S_1. So gelten z. B. die Äquivalenzen:

$$\{5\} \subset S_2 \quad \leftrightarrow \quad \{(1, 4), (2, 3), (3, 2), (4, 1)\} \subset S_1$$
$$\{3, 11\} \subset S_2 \quad \leftrightarrow \quad \{(1, 2), (2, 1), (5, 6), (6, 5)\} \subset S_1$$

Äquivalente Mengen beschreiben denselben Sachverhalt, nur die Darstellungsweise ist unterschiedlich. Vereinigung, Durchschnitt und Komplement von Teilmengen von S_2 sind äquivalent zu Vereinigung, Durchschnitt und Komplement der entsprechenden Teilmengen von S_1.
Zu jeder Teilmenge aus S_2 gibt es genau eine äquivalente Teilmenge aus S_1, aber nicht umgekehrt, wie man schon an der Anzahl der möglichen Teilmengen sieht:

Anzahl Ereignisse = Teilmengen von S_1: $2^{36} =$ 68.719.476.736

Anzahl Ereignisse = Teilmengen von S_2: 2^{11} = 2.048

Zum Beispiel gibt es nur zu den Elementarereignissen {(1, 1)} und {(6, 6)} aus S_1 ein äquivalentes Ereignis aus S_2, zu allen anderen Elementarereignissen aus S_1 gibt es kein äquivalentes Ereignis aus S_2. Betrachtet man S_2 statt S_1, so hat man eine drastische Vereinfachung vorgenommen und dabei

$$1 - \frac{2^{11}}{2^{36}} = 99{,}999997\ \%$$

der auf Basis von S_1 möglichen Ereignisse ausgeschlossen.

Sind $A_1 \in \mathfrak{A}_1$ und $A_2 \in \mathfrak{A}_2$ zwei äquivalente Ereignisse, d. h., beschreiben sie denselben Sachverhalt, so sind die Wahrscheinlichkeiten gleich, wenn man von derselben Einstellung der bewertenden Person ausgeht: $P_1(A_1) = P_2(A_2)$.

(7.1.-2) ***Beispiel***

Es seien S_i und P_i wie im vorangegangenen Beispiel mit $\mathfrak{A}_i = \mathfrak{P}(S_i)$. Geht man davon aus, dass die Wahrscheinlichkeiten für jedes Elementarereignis in S_1 gleich sind, so ist für jedes Elementarereignis {(a, b)} in S_1:

$$P_1(\{(a, b)\}) = \frac{1}{36}.$$

Damit ergibt sich für die Elementarereignisse in S_2:

$$P_2(\{2\}) = P_1(\{(1, 1)\}) = \frac{1}{36}$$

$$P_2(\{3\}) = P_1(\{(1, 2), (2, 1)\}) = P_1(\{(1, 2)\}) + P_1(\{(2, 1)\}) = \frac{2}{36}$$

…

$$P_2(\{6\}) = \ldots = \frac{5}{36}$$

$$P_2(\{7\}) = \ldots = \frac{6}{36}$$

$$P_2(\{8\}) = \ldots = \frac{5}{36}$$

…

$$P_2(\{11\}) = \ldots = \frac{2}{36}$$

$$P_2(\{12\}) = \ldots = \frac{1}{36}$$

allgemein:

$$P_2(\{i\}) = \frac{6-|7-i|}{36}$$

Daraus kann man dann die Wahrscheinlichkeit jedes beliebigen Ereignisses in S_2 leicht ableiten. S_1 wird nicht mehr weiter benötigt.

Am Beispiel sieht man auch, dass alle Anforderungen aus der Definition der Wahrscheinlichkeit für P_2 inklusive der Subadditivität erfüllt sind. Es handelt sich also tatsächlich um eine Wahrscheinlichkeit.

(7.1.-3) ***Definition***

Eine Abbildung $g: S_1 \rightarrow S_2$ heißt **surjektiv**, wenn es zu jedem $a_2 \in S_2$ ein $a_1 \in S_1$ mit $g(a_1) = a_2$ gibt. Jedes Element aus S_2 taucht also als Bild eines Elementes aus S_1 auf.

(7.1.-4) ***Satz***

Gegeben sind zwei Vorgänge mit den endlichen Ergebnismengen S_1 und S_2 und dem Wahrscheinlichkeitsraum $(S_1, \mathfrak{A}_1, P_1)$, sowie eine surjektive Abbildung $g: S_1 \rightarrow S_2$.

Dann ist $(S_2, \mathfrak{A}_2, P_2)$ mit

$S_2 = g(S_1) = \{a_2 \mid \text{es gibt ein } a_1 \in S_1 \text{ mit } g(a_1) = a_2\}$

$\mathfrak{A}_2 = g(\mathfrak{A}_1) = \{A_2 \mid \text{es gibt ein } A_1 \in \mathfrak{A}_1 \text{ mit } g(A_1) = A_2\}$

$P_2 = P_1 \circ g^{-1} \iff P_2(A_2) = P_1(g^{-1}(A_2))$ für jedes $A_2 \in \mathfrak{A}_2$

ein Wahrscheinlichkeitsraum.

Beweis:

Es seien $A_2, B_2 \in \mathfrak{A}_2$.

Dann gibt es eindeutig bestimmte Mengen $A_1, B_1 \in \mathfrak{A}_1$ mit $A_2 = g(A_1)$ und $B_2 = g(B_1)$.

(1) Zu zeigen: $\mathfrak{A}_2 = g(\mathfrak{A}_1)$ ist eine σ-Algebra
$A_2 \cup B_2 = g(A_1) \cup g(B_1) = g(A_1 \cup B_1)$. Wegen $A_1 \cup B_1 \in \mathfrak{A}_1$ ist dann auch $A_2 \cup B_2 \in \mathfrak{A}_2$.
Analog für Durchschnitt und Komplement.

(2) Zu zeigen: $0 \leq P_2(A_2) \leq 1$
Das gilt wegen $0 \leq P_1(A_1) \leq 1$ und $P_1(A_1) = P_1(g^{-1}(A_2)) = P_2(A_2)$

(3) Zu zeigen: $P_2(A_2 \cup B_2) = P_2(A_2) + P_2(B_2) - P_2(A_2 \cap B_2)$

$$\begin{aligned} P_2(A_2 \cup B_2) &= P_1 \circ g^{-1}(A_2 \cup B_2) = P_1(A_1 \cup B_1) = P_1(A_1) + P_1(B_1) - P_1(A_1 \cap B_1) \\ &= P_2(A_2) + P_2(B_2) - P_2(A_2 \cap B_2) \end{aligned}$$

(7.1.-5) ***Definition***

Gegeben sind zwei Vorgänge mit den endlichen Ergebnismengen S_1 und S_2 und dem Wahrscheinlichkeitsraum $(S_1, \mathfrak{A}_1, P_1)$, sowie eine surjektive Abbildung $g: S_1 \rightarrow S_2$.
Ist $s \in S_1$ die unabhängige Variable, so heißt die abhängige Variable $X = g(s) \in S_2$ eine **diskrete Zufallsvariable**. Im Spezialfall $S_2 \subset \mathbb{R}$ spricht man auch von einer **reellen diskreten Zufallsvariablen**.

Mithilfe einer Zufallsvariablen X kann man statt einer Wahrscheinlichkeit eines Ereignisses einfacher die Wahrscheinlichkeit einer Aussage formulieren:

$P_2(\text{Aussage bezüglich } X) = P_2(\{y \in S_2 \mid \text{die Aussage ist für } y \text{ wahr}\})$

(7.1.-6) ***Beispiel***

Im obigen Beispiel (7.1.-2) ist

$s = (a, b) \in S_1$,

$g(s) = g((a, b)) = a + b \in S_2$,

die Abbildung g ordnet also den Augenzahlen zweier Würfel die Summe zu. Die zugehörige Zufallsvariable ist die abhängige Variable:

$X = a + b$.

Die Elementarereignisse in S_2 sind die Mengen, die genau eine Realisierung von X enthalten. Die Elementarereignisse in S_2 mit den zugehörigen Wahrscheinlichkeiten wurden oben bereits aufgelistet.
Beispiele:

$P_2(X = 3) = P_2(\{y \in S_2 \mid y = 3\}) = P_2(\{3\}) = \frac{2}{36}$

$P_2(X \leq 8) = P_2(\{y \in S_2 \mid y \leq 8\}) = P_2(\{2, 3, 4, 5, 6, 7, 8\}) = \frac{1+2+3+4+5+6+5}{36} = \frac{26}{36}$

(7.1.-7) ***Hinweise***

1. Da g surjektiv ist, ist $|S_2| \leq |S_1|$ und $|\mathfrak{A}_2| \leq |\mathfrak{A}_1|$

Beziehen sich beide Ergebnismengen S_i auf den gleichen Vorgang, so kann durch g eine vereinfachte Beschreibung der Ereignisse und Wahrscheinlichkeiten erzeugt werden – siehe Beispiel oben und in den folgenden Kapiteln (7.3.) und (7.4.).

2. Der Vorteil bei der Verwendung der Variablen X anstelle von g(s) ist, dass man die Urbilder der Abbildung g ausblenden kann, wenn sie nicht von Bedeutung sind.

3. Liegen die Werte einer reellen diskreten Zufallsvariablen so dicht beieinander, dass man sie mit akzeptablem Fehler durch eine stetige Variable approximieren kann, so heißt die stetige Variable eine **stetige reelle Zufallsvariable**. Stetige reelle Zufallsvariable werden im Folgenden nicht betrachtet.

4. Solange man sich nur in S_2 mit P_2 bewegt, kann der Index weggelassen werden, und so findet man das auch im Allgemeinen in Veröffentlichungen.

5. Der Begriff „Zufall“ in „Zufallsvariable“ ist irreführend, da eine Zufallsvariable höchstens in speziellen Beispielen etwas mit Zufall zu tun hat.

6. Achtung: in Veröffentlichungen wird üblicherweise die Abbildung g als Zufallsvariable bezeichnet, statt g der Buchstabe X verwendet und dann z. B. im Falle $S_2 = \mathbb{R}$ Ausdrücke wie „X = 9“ benutzt. Damit handelt man sich aber Probleme ein, denn eine Abbildung kann nicht eine Zahl sein, Funktion und Funktionswert sind zwei verschiedene Dinge. Die Autoren müssen sich folglich immer verrenken, um eine nachvollziehbare Interpretation von „X = 9“ zu liefern. Gemeint ist aber in jedem Falle immer das, was oben in Definition (7.1.-5) beschrieben ist, nämlich die abhängige Variable, nicht die Abbildung selbst.
   Außerdem steht diese oft verwendete Definition der Zufallsvariablen im Widerspruch zu den üblichen Definitionen in der Mathematik, wie das folgende Beispiel zeigt:
   Es sei f: $\mathbb{R} \rightarrow \mathbb{R}$, mit y = f(x) und y = x^2. Dann heißt
   - f eine Abbildung (nicht eine Variable)
   - x die unabhängige Variable
   - y die abhängige Variable
   - x = 3 eine Realisierung von x
   - y = 9 die zugehörige Realisierung von y

   „f = 9“ statt „y = 9“ ist formal falsch und so etwas kommt in der Analysis auch nicht vor.

7.2 Gleichverteilung

(7.2.-1) ***Definition***
Es wird ein Vorgang mit dem folgenden Wahrscheinlichkeitsraum betrachtet:
- Es gibt n mögliche Ergebnisse, also $S = \{e_1, \ldots, e_n\}$, $|S| = n$, $n \in \mathbb{N}$.
- $\mathfrak{A} = \mathfrak{P}(S)$
- Jedes Elementarereignis hat dieselbe Wahrscheinlichkeit, also
  $P(\{e_i\}) = \frac{1}{n}$ für i = 1, …, n.

In diesem Fall heißt P eine **Gleichverteilung** auf S.

Die Wahrscheinlichkeit eines Ereignisses E ist einfach gleich der relativen Häufigkeit mit der die Elemente von E in S vorkommen:

$$P(E) = \frac{|E|}{|S|},$$

und das ist die klassische Laplacesche Definition der Wahrscheinlichkeit.

(7.2.-2) ***<u>Definition</u>***

Gegeben sei S_1 mit der Wahrscheinlichkeit P_1, g: $S_1 \rightarrow S_2$ surjektiv und X die zugehörige Zufallsvariable. Ist die Wahrscheinlichkeit $P_2 = P_1 \circ g^{-1}$ eine Gleichverteilung auf S_2, so sagt man auch: die **Zufallsvariable X ist gleichverteilt**.

Ist $S_2 = \{x_j | j = 1, \ldots, m\}$, so ist für j = 1, …, m:

$$P_2(\{x_j\}) = P_2(X = x_j) = \frac{1}{m}$$

Die Eigenschaften von $(S_1, \mathfrak{A}_1, P_1)$ spielen keine Rolle.

7.3 Hypergeometrische Verteilung

Ausgangspunkt sind die im Kapitel 3.3 bereits behandelten Fälle:

Fall 1: Variation (= Reihenfolge wird berücksichtigt) ohne Wiederholung
Fall 3: Kombination (= Reihenfolge wird nicht berücksichtigt) ohne Wiederholung

Die Frage lautet: „Wie groß ist die Wahrscheinlichkeit, dass unter n zufällig aus G ausgewählten Elementen genau i die Eigenschaft E haben?“ Es handelt sich dabei also um eine A-priori-Wahrscheinlichkeit.

Im Folgenden wird immer nur „Variation“ benutzt, die Ergebnisse für „Kombination“ sind gemäß Kapitel 3.3 dieselben.

Um die Formeln aus Kapitel 3.3 anwenden zu können, muss unterstellt werden, dass jede Variation von n Elementen die gleiche Wahrscheinlichkeit hat, ausgewählt zu werden. Das ist z. B. erfüllt, wenn die Elemente nacheinander ausgewählt werden und bei jedem Schritt für jedes Element dieselbe Wahrscheinlichkeit besteht, ausgewählt zu werden. Dabei verändert sich die Wahrscheinlichkeit von Schritt zu Schritt, da ohne Wiederholung ausgewählt wird.
In der Bewertungstabelle für die Wahrscheinlichkeit kommen also nur die Variationen mit ihren relativen Häufigkeiten vor, es werden keine weiteren Einflussfaktoren berücksichtigt.

(7.3.-1) ***<u>Definition</u>***

Um die Frage nach der Wahrscheinlichkeit konkret beantworten zu können, wird definiert:

G Grundgesamtheit
E eine Eigenschaft von gewissen Elementen der Grundgesamtheit
N = Anzahl der Elemente von G, N = |G|
M = Anzahl der Elemente von G mit der Eigenschaft E $(0 \leq M \leq N)$
n = Anzahl der ausgewählten Elemente
i = Anzahl der ausgewählten Elemente mit der Eigenschaft E $(0 \leq i \leq n)$

Da Wiederholungen bei der Auswahl nicht erlaubt sind, gilt:

- $0 \leq n \leq N$
- $0 \leq i \leq M$ und $0 \leq n-i \leq N-M$

(7.3.-2) ***Definition***

Zur Ermittlung der gesuchten Wahrscheinlichkeit wird definiert:

Der Wahrscheinlichkeitsraum $(S_1, \mathfrak{A}_1, P_1)$ mit

$S_1 = G\times\ldots\times G = G^n$ ist die Menge der Variationen von n beliebigen Elementen aus der Grundgesamtheit G. $|S_1| = \binom{N}{n}\cdot n!$

$e = (e_1, \ldots, e_n) \in S_1$; $e_j \in G$ $(j = 1, \ldots, n)$ sind die Komponenten von e

$\mathfrak{A}_1 = \mathfrak{P}(S_1)$

$P_1(\{e\}) = \frac{1}{|S_1|} = \frac{1}{\binom{N}{n}\cdot n!}$, da Gleichverteilung vorausgesetzt ist.

$K(i) = \{e \mid e = (e_1, \ldots, e_n) \in S_1$ und genau i Komponenten von e haben die Eigenschaft $E\} \subset S_1$

Der Wahrscheinlichkeitsraum $(S_2, \mathfrak{A}_2, P_2{=}P)$ mit

$S_2 = \{0, 1, 2, \ldots, m\}$

$\mathfrak{A}_2 = \mathfrak{P}(S_2)$

$P = P_1\circ g^{-1}$

mit

g = Abbildung $S_1 \rightarrow S_2$ mit g(e) = Anzahl Komponenten von e mit Eigenschaft E.
Da ohne Wiederholen ausgewählt wird, ist $m = \min(n, M)$
Es ist $g(K(i)) = \{i\}$

X = Zufallsvariable mit $X = g(e)$

Die Zufallsvariable X ist **hypergeometrisch** verteilt oder H(N; M; n)-verteilt.

(7.3.-3) ***Formel***

Mit den Bezeichnungen aus den Definitionen (7.3.-1) und (7.3.-2) gilt: die Wahrscheinlichkeit, dass unter n zufällig aus G ausgewählten Elementen genau i die Eigenschaft E haben, beträgt

$$\mathbf{P(X = i) = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}}}$$

Beweis:

Nach Kapitel 3.3 ist:

$$P(X = i) = P(\{i\}) = P_1\circ g^{-1}(\{i\}) = P_1(K(i)) = \frac{|K(i)|}{|S_1|} = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}\cdot n!}{\binom{N}{n}\cdot n!} = \frac{\binom{M}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}}$$

Die hypergeometrische Verteilung hat also zunächst nichts mit Wahrscheinlichkeiten zu tun, sondern stellt eine relative Häufigkeit dar. Nur durch eine entsprechende Interpretation wird aus der relativen Häufigkeit eine Wahrscheinlichkeit. Rechnerisch muss man sich also auch in diesem Fall nicht mit Wahrscheinlichkeiten befassen, die relativen Häufigkeiten tun's auch.

(7.3.-4) ***Beispiel***

Das Beispiel aus Kapitel 3.3 wird dann umformuliert zu:

$$P(X = 2) = \frac{\binom{3}{2}\cdot\binom{7-3}{3-2}}{\binom{7}{3}} = \frac{12}{35} \approx 0{,}343$$

Mit Hilfe eines <u>Baumdiagrammes</u> ergibt sich dasselbe Ergebnis. Dabei wird die Darstellung gewählt, dass die 3 Kugeln nacheinander ohne Zurücklegen gezogen werden.

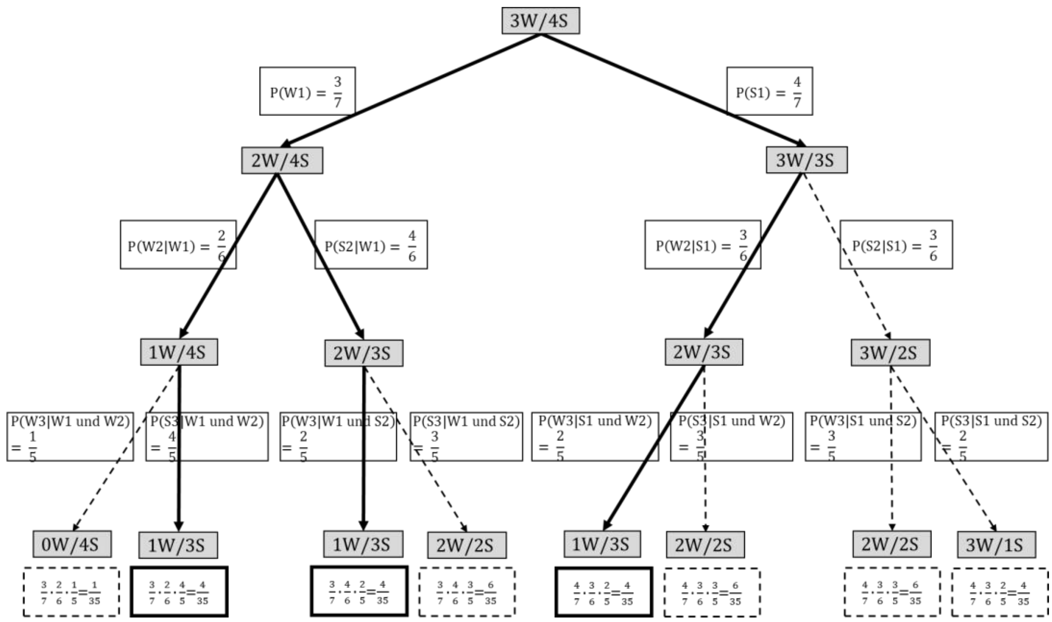
Im Diagramm bedeutet zum Beispiel:

2W/3S	Es sind 2 weiße und 3 schwarze Kugeln in der Urne
P(S2|W1)	Die Wahrscheinlichkeit, dass im 2. Zug eine schwarze Kugel gezogen wurde unter der Bedingung, dass im 1. Zug eine weiße Kugel gezogen wurde

Zieht man 2 weiße und 1 schwarze Kugel, so verbleiben 1 weiße und 3 schwarze Kugeln in der Urne. Es müssen dann die Wahrscheinlichkeiten am Ende der Pfade, bei denen 1 weiße und 3 schwarze Kugeln in der Urne verblieben sind (also 1W/3S), addiert werden. Diese Pfade und die zugehörigen Endwahrscheinlichkeiten sind durch fette Pfeile bzw. Umrandungen gekennzeichnet.

Es ergibt sich mit dieser Darstellung ebenfalls:

$$P(X = 2) = \frac{4}{35} + \frac{4}{35} + \frac{4}{35} = \frac{12}{35} \approx 0{,}343$$



7.4 Binomialverteilung

Ausgangspunkt ist der im Kapitel 3.3 bereits behandelten Fall:

Fall 2: Variation (= Reihenfolge wird berücksichtigt) mit Wiederholung

Auch hier kann dieselbe Frage wie bei der hypergeometrischen Verteilung gestellt werden: „Wie groß ist die Wahrscheinlichkeit, dass unter n zufällig aus G ausgewählten Elementen genau i die Eigenschaft E haben?“ Es handelt sich dabei also um eine A-priori-Wahrscheinlichkeit.

Der Unterschied liegt lediglich darin, dass bei der Binomialverteilung mit Wiederholung ausgewählt wird, während bei der hypergeometrischen Verteilung ohne Wiederholung ausgewählt wird.

Zunächst wird aber ein allgemeinerer Fall betrachtet, der die Frage dann als Spezialfall beantwortet.

Man beobachtet n Vorgänge, die sich nicht gegenseitig beeinflussen. Der Ablauf und das Ergebnis eines Vorgangs hat keinerlei Auswirkungen auf den Ablauf und das Ergebnis eines anderen Vorganges.

Zu jedem Vorgang gibt es einen Wahrscheinlichkeitsraum $(S_k, \mathfrak{A}_k, P_k)$ und ein Ereignis $A_k \in \mathfrak{A}_k$, sodass für alle k gilt:
$\mathfrak{A}_k = \{\emptyset, A_k, \overline{A_k}, S\}$
Es gibt eine feste von k unabhängige Zahl p mit $p = P_k(A_k)$

Es sei

$S = S_1 \times \ldots \times S_n$
$\mathfrak{A} = \mathfrak{A}_1 \times \ldots \times \mathfrak{A}_n$
$A \in \mathfrak{A}$
$P(A) = P_1(A_1) \cdot \ldots \cdot P_n(A_n) = p^n$

Der Wahrscheinlichkeitsraum $(S^*, \mathfrak{A}^*, P^*)$ sei definiert durch

$S^* = \{0, 1, \ldots, n\}$
$\mathfrak{A}^* = \mathfrak{P}(S^*)$
$P^* = P \circ g^{-1}$

mit

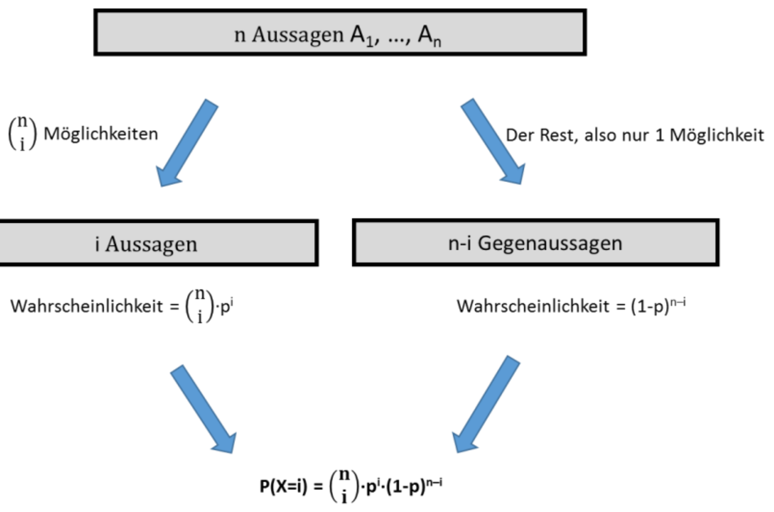
$e = (e_1, \ldots, e_n) \in S$
g = Abbildung $S \rightarrow S^*$ mit g(e) = Anzahl Komponenten von e mit $e_k \in A_k$
X = Zufallsvariable mit X = g(e)
X zählt also, wie viele der Ereignisse A_k eingetreten sind.

Die Zufallsvariable X ist **binomial** verteilt oder B(n; p)-verteilt oder $B_{n;p}$-verteilt.

(7.3.-3) ***Formel***
Mit den Bezeichnungen aus den Definitionen (7.3.-1) und (7.3.-2) gilt: die Wahrscheinlichkeit, dass genau i der Ereignisses A_k eingetreten sind, ist

$$\mathbf{P(X = i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} \;; \quad i = 0, 1, \ldots, n}$$

Beweis analog zu Fall 3 in Kapitel 3.3:



(7.4.-2) ***Beispiel***
Für jedes k = 1, …, n ist

$$P(A_k) = p \text{ und } P(\overline{A_k}) = 1-p.$$

Es sein n = 5. Dann ist z. B.

$$P(A_1 \text{ und } \overline{A_2} \text{ und } \overline{A_3} \text{ und } A_4 \text{ und } \overline{A_5}) = P_1(A_1)\cdot P_2(\overline{A_2})\cdot P_3(\overline{A_3})\cdot P_4(A_4)\cdot P_5(\overline{A_5}) = p^2\cdot(1-p)^3$$

Es gibt insgesamt $\binom{5}{2}$ = 10 Möglichkeiten für den Fall, dass 2 Ereignisse eingetreten und 3 Ereignisse nicht eingetreten sind und für alle diese Möglichkeiten ergibt sich dieselbe Wahrscheinlichkeit. Also ist

$$P(X = 2) = \binom{5}{2}\cdot p^2\cdot(1\text{-}p)^3$$

Meistens wird folgender Sonderfall betrachtet:

(7.4.-3) ***Sonderfall***

- Man beobachtet n gleichartige Vorgänge. Ob diese Vorgänge gleichzeitig oder hintereinander passieren, ist egal.
- Es wird ein für alle Vorgänge gleiches Ereignis A vorgegeben.
- Die Vorgänge beeinflussen sich gegenseitig nicht, sodass die Wahrscheinlichkeit für den Eintritt des Ereignisse A bei jedem Vorgang unverändert gleich p = P(A) ist.
- Die Zufallsvariable X ist definiert als:

  X = Häufigkeit des Eintretens von A

  X kann also alle ganzzahligen Werte von 0, …, n annehmen.

(7.4.-5) ***Beispiel***
Es wird mit 5 Würfeln gleichzeitig einmal gewürfelt. Es werden nur gültige Ergebnisse berücksichtigt. Für die jeweiligen Einzelwürfe wird dasselbe Ereignis A, dass eine 6 gewürfelt wird – A = {6} – betrachtet.

Unter der Annahme, dass die Wahrscheinlichkeit für jedes einzelne Würfelergebnis unabhängig von den anderen Würfelergebnissen = $\frac{1}{6}$ ist, soll die Wahrscheinlichkeit ausgerechnet werden, dass genau 3-mal eine „6“ gewürfelt wird.

Da die Wahrscheinlichkeit aller Ereignisse gleich ist, kann die Formel der Binomialverteilung benutzt werden, also:

$n = 5$
$i = 3$
$p = P(A) = \frac{1}{6}$

Dann ist

$$P(X = 3) = \binom{5}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 = \frac{250}{7776} = 0{,}03215.$$

Genauso gut könnte man auch mit einem Würfel 5-mal hintereinander würfeln. Das zugehörige Baumdiagramm hat $2^5 = 32$ Enden, von denen $\binom{5}{3} = 10$ das Ende eines Pfades mit genau drei A’s (es wird eine 6 gewürfelt) und 2 $\overline{A}$’s (es wird keine 6 gewürfelt) sind.

(7.4.-6) ***Hinweis***

Wie in Kapitel 3.3 bereits gezeigt wurde, nähert sich die hypergeometrische Verteilung der Binomialverteilung unter gewissen Bedingungen immer mehr an.
In Büchern findet man daher oft Beispiele, bei denen in Ermangelung der Größe der Grundgesamtheit einfach die Binomialverteilung genommen wird, obwohl „ohne Zurücklegen“ ausgewählt wird. Es muss dann aber zumindest erwähnt werden, dass die Aufgabe nicht exakt lösbar ist, sondern nur näherungsweise über die Binomialverteilung. Dieser Hinweis fehlt aber oft!
Typisch ist folgendes Beispiel.

(7.4.-7) ***Beispiel***

76,8% der Bevölkerung in Deutschland nutzte 2014 das Internet. Wie groß ist die Wahrscheinlichkeit, dass 2 von 3 zufällig daraus ausgewählten Personen Internetbenutzer sind?
Jede Person wird sinnvollerweise nur einmal ausgewählt, es wird also „ohne Wiederholen“ ausgewählt, und damit muss die hypergeometrische Verteilung benutzt werden. Da aber die Grundgesamtheit nicht vorgegeben ist, ist die Aufgabe so nicht lösbar.

- Ansatz 1: Man beschafft sich N
  Geht man zusätzlich von einer Bevölkerung von 80.000.000 aus, so ergibt sich
  $N = 80.000.000$
  $M = 0{,}768 \cdot 80.000.000 = 61.440.000$
  $n = 3$
  $i = 2$
  $$P(X = 2) = \frac{\binom{61.440.000}{2} \cdot \binom{80.000.000 - 61.440.000}{3-2}}{\binom{80.000.000}{3}} \approx 41{,}05\%$$
- Ansatz 2: Näherung durch Binomialverteilung
  Dann ist:
  $p = 0{,}768$
  $n = 3$
  $i = 2$
  $$P(X = 2) = \binom{3}{2} \cdot 0{,}768^2 \cdot 0{,}232^1 \approx 41{,}05\% .$$

8 Einzelthemen

Zusammenfassung
In diesem Kapitel werden mehrere Einzelthemen behandelt. Zunächst wird gezeigt, dass die Verwendung des Begriffs Wahrscheinlichkeit in den Kolmogoroffschen Axiomen irreführend ist. Dann wird erläutert, warum das weit verbreitete Beispiel des idealen oder fairen Würfels ebenfalls irreführend ist. Und schließlich wird das Indifferenzprinzip betrachtet und gewisse damit zusammenhängende Paradoxien gelöst.
Abschließend werden die Kapitel 5 bis 7 auf einer Seite zusammengefasst.
* * *

8.1 Die Wahrscheinlichkeit nach Kolmogoroff

Der russische Mathematiker Andrei Kolmogoroff veröffentlichte 1933 die „Grundbegriffe der Wahrscheinlichkeitsrechnung“ in [KA]. Die darin enthaltenen Axiome wurden zur Grundlage der Wahrscheinlichkeitstheorie als Teilgebiet der Mathematik und werden heute in etwas veränderter Form benutzt. Dieser Ansatz heißt auch „axiomatische Definition der Wahrscheinlichkeit“, kurz: „axiomatische Wahrscheinlichkeit“.

Im Folgenden sind die relevanten Ausschnitte aus der Originalarbeit zitiert:

(8.1.-1) ***Endlichkeit***
Voraussetzung ist, dass die Ausgangsmenge E, die der Menge S in Definition (6.3.-1) entspricht, endlich ist. Für unendliche Mengen E gibt es als 6. Axiom das Stetigkeitsaxiom, das hier nicht betrachtet wird.
[KA], Seite 14:
„Bei der Beschreibung irgendwelcher wirklich beobachtbarer zufälliger Prozesse kann man nur endliche Wahrscheinlichkeitsfelder erhalten. Unendliche Wahrscheinlichkeitsfelder erscheinen nur als idealisierte Schemata reeller zufälliger Prozesse“

(8.1.-2) ***Mengenkörper***
[KA], Seite 2:
„Ein Mengensystem heißt ein Körper, wenn Summe Durchschnitt und Differenz von zwei Mengen des Systems wieder dem System angehören“

(8.1.-3) ***§1. Axiome***
[KA], Seite 2:
„Es sei eine Menge von Elementen $\xi, \eta, \zeta, \ldots$, welche man elementare Ereignisse *nennt, und $\mathfrak{F}$ eine Menge aus Teilmengen von E; die Elemente der Menge $\mathfrak{F}$ werden weiter* zufällige Ereignisse *genannt.*

- *I.* *$\mathfrak{F}$ ist ein Mengenkörper.*
- *II.* *$\mathfrak{F}$ enthält die Menge E.*
- *III.* *Jeder Menge A aus $\mathfrak{F}$ ist eine nichtnegative reelle Zahl P(A) zugeordnet. Diese Zahl P(A) nennt man die Wahrscheinlichkeit des Ereignisses A.*

IV. P(E) = 1.
V. Wenn A und B disjunkt sind, so gilt
P(A + B) = P(A) + P(B).

(8.1.-4) ***Interpretation***
Kolmogoroff hat bereits darauf hingewiesen, dass durch die Axiome nicht die konkrete Bedeutung in der Realität geklärt wird. „Wahrscheinlichkeit“ ist nur ein (bedeutungsloser) Name für einen bestimmten zu untersuchenden Gegenstand.
[KA], Seite 2:
„Die Wahrscheinlichkeitstheorie als mathematische Disziplin soll und kann genau in diesem Sinne axiomatisiert werden wie die Geometrie oder Algebra. Das bedeutet, dass, nachdem die Namen der zu untersuchenden Gegenstände und ihrer Grundbeziehungen sowie die Axiome, denen diese Grundbeziehungen zu gehorchen haben, angegeben sind, die ganze weitere Darstellung sich ausschließlich auf diese Axiome gründen soll und keine Rücksicht auf die jeweilige konkrete Bedeutung dieser Gegenstände und Beziehungen nehmen darf“.
Siehe auch [AT], Kapitel 3.1.

(8.1.-5) ***Konkrete Bedeutung***
[KA], Seite 1:
Ferner hatte Kolmogoroff in seiner Arbeit darauf hingewiesen, dass es ganz unterschiedliche reale Objekte geben kann, die die Axiome erfüllen, obwohl sie nichts mit dem umgangssprachlichen Begriff der Wahrscheinlichkeit zu tun haben:
„Jede axiomatische (abstrakte) Theorie lässt bekanntlich unbegrenzt viele konkrete Interpretationen zu. In dieser Weise hat auch die mathematische Wahrscheinlichkeitstheorie neben derjenigen ihrer Interpretationen, aus der sie aufgewachsen ist, auch zahlreiche andere. Wir kommen so zu Anwendungen der mathematischen Wahrscheinlichkeitstheorie auf Untersuchungsgebiete, die mit den Begriffen des Zufalls und der Wahrscheinlichkeit im konkreten Sinne dieser Begriffe nichts zu tun haben.“

Es ist in der Wissenschaft durchaus üblich, Begriffe aus der Umgangssprache zu benutzen, obwohl es keinerlei inhaltlichen Zusammenhang zwischen der Verwendung in der Wissenschaft und in der Umgangssprache gibt. Es könnte in diesen Fällen also genauso gut ein beliebiger anderer Begriff in der Wissenschaft benutzt werden. Beispiele:

- „Geschlecht“ in der Topologie hat nichts mit dem biologischen oder grammatikalischen Geschlecht zu tun.
- „Ring“ und „Körper“ in der Algebra haben nichts mit den umgangssprachlichen Begriffen zu tun.
- „Flavour“/“flavor“ und „beauty“ in der Quantentheorie haben nichts mit den umgangssprachlichen Begriffen zu tun.

Daraus folgt:
Die axiomatische Wahrscheinlichkeit nach Kolmogoroff hat nichts mit der umgangssprachlichen oder der praktischen Wahrscheinlichkeit zu tun.

Wer das Gegenteil behauptet (und dazu zähle auch ich), muss es beweisen oder zumindest plausibel machen. Wie bei jedem anderen Zusammenhang zwischen Realität und mathematischem Modell braucht man dafür drei Dinge:

1. Eine exakte Definition des abstrakten Objektes in der Mathematik. Das ist hier die axiomatische Wahrscheinlichkeit und die findet man bei [KA] und in vielen Stochastikbüchern.
2. Eine exakte Definition des konkreten realen Objektes. Das ist hier die „praktische Wahrscheinlichkeit“ und die findet man im vorliegenden Buch in Kapitel 5.6. zusammen mit dem Verfahren zur Bestimmung einer Wahrscheinlichkeit in Kapitel 5.8 und im Regelfall nicht in Stochastikbüchern.
3. Einen plausiblen Zusammenhang zwischen beiden Definitionen. Das ist die folgende Tabelle.
   K steht für Kolmogoroff und R für Realität, wobei auch das Kapitel 5.9.1 über die Subadditivität einbezogen wurde.

(8.1.-6) ***<u>Tabelle</u>***

Axiomatische Wahrscheinlichkeit (K)	**Praktische Wahrscheinlichkeit (R)**	**Vergleich**
Basis: Mengensysteme	Basis: Aussagen; Mengensysteme als Sonderfall	R allgemeiner als K
Menge E (endlich; mit VI. Stetigkeitsaxiom auch unendlich)	Menge S (endlich)	bei Mengensystemen: K allgemeiner als R
Elementare Ereignisse	Ergebnisse	gleichwertig
Zufällige Ereignisse	Ereignisse	gleichwertig
$\mathfrak{F}$ ist eine Menge von Teilmengen von E I. $\mathfrak{F}$ ist ein beliebiger Mengenkörper II. $\mathfrak{F}$ enthält E III. $P(A) \geq 0$ IV. $P(E) = 1$ V. A, B disjunkt, so gilt: $P(A + B) = P(A) + P(B)$	$\mathfrak{A}$ ist eine Menge von Teilmengen von S $\mathfrak{A}$ ist σ-Algebra $\mathfrak{A}$ enthält S $P(A) \geq 0$ $P(S) = 1$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$	gleichwertig
Praktische Interpretation von „Wahrscheinlichkeit“ fehlt, es sind unterschiedliche Interpretationen möglich	Eindeutige praktische Interpretation der Wahrscheinlichkeit ist enthalten	R kann unmittelbar praktisch angewendet werden, K nicht
Verfahren zur Ermittlung einer Wahrscheinlichkeit fehlt. Zu einem Mengenkörper gibt es unendlich viele verschiedene Wahrscheinlichkeiten.	Verfahren zur Ermittlung einer Wahrscheinlichkeit ist enthalten. Das Verfahren ist konkret, lässt aber persönliche Spielräume.	

Daraus folgt:

> **Es gibt einen gewissen Zusammenhang zwischen axiomatischer und praktischer Wahrscheinlichkeit, aber die beiden Definitionen sind nicht 1:1 aufeinander abbildbar.**

Man hat also grundsätzlich drei Möglichkeiten:

<u>1. Möglichkeit</u>

Die Axiome bilden die Grundlage der Betrachtung.

Damit bleibt man im Reich der Mathematik, also auf der abstrakten Modellebene. Dann ist der Begriff „Wahrscheinlichkeit“ genauso willkürlich und praktisch bedeutungslos wie die Begriffe „Körper“ oder „Ring“ in der Algebra oder der Begriff „Geschlecht“ in der Topologie. Das bedeutet, dass in der axiomatischen Definition und in allen darauf aufbauenden Sätzen ohne Substanzverlust der Begriff „Wahrscheinlichkeit“ z. B. durch „Blubb“ ersetzt werden könnte. Der Satz „die Wahrscheinlichkeit, eine 3 zu würfeln, beträgt $\frac{1}{6}$“ wird dann zu „die Blubb, eine 3 zu würfeln, beträgt $\frac{1}{6}$“ und das ist genauso aussagekräftig wie vor der Ersetzung. Und aus Wahrscheinlichkeitstheorie wird Blubbtheorie. Der Begriff Blubbtheorie ist in diesem Fall besser als Wahrscheinlichkeitstheorie, da damit vermieden wird, dass bewusst oder unbewusst ein Zusammenhang zur Realität und zum umgangssprachlichen oder praktischen Begriff Wahrscheinlichkeit hergestellt wird. Man kann natürlich statt Blubb auch den in der Maßtheorie üblichen Begriff „normiertes Maß“ nehmen (siehe z. B. [HR], Seite 6 und 22), und dann entsteht die Aussage „das normierte Maß, eine 3 zu würfeln, beträgt $\frac{1}{6}$“, aber das bringt einen auch nicht wirklich weiter.

2. Möglichkeit

Man will Wahrscheinlichkeiten in realen praktischen Situationen untersuchen. Dann nimmt man die Definition der praktischen Wahrscheinlichkeit aus Kapitel 5.6, die Methode zur Bestimmung einer Wahrscheinlichkeit in Kapitel 5.8 und weitere Darstellungen. Eine Aussage wie „die Wahrscheinlichkeit, eine 3 zu würfeln, beträgt $\frac{1}{6}$“ ist auf dieser Basis praktisch sinnvoll interpretierbar. Die Axiome nach Kolmogoroff werden nicht benötigt.

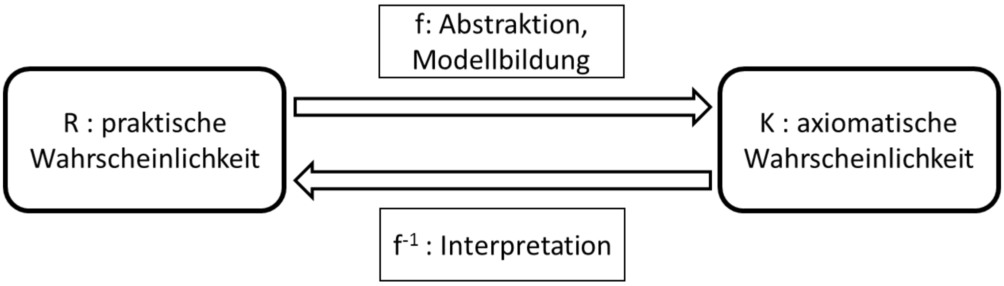
3. Möglichkeit

Man will die erste und zweite Möglichkeit kombinieren, also Aussagen über die Realität machen und dabei die mathematischen Erkenntnisse, die auf Basis der Kolmogoroffschen Axiome gewonnen wurden, nutzen. Dann muss man aber einen plausiblen Zusammenhang zwischen dem Begriff bei Kolmogoroff (mag er nun Blubb oder Wahrscheinlichkeit oder normiertes Maß oder sonst wie heißen) und einem adäquaten umgangssprachlichen Begriff (mag er nun relativer Anteil oder Wahrscheinlichkeit oder sonst wie heißen) herstellen. Das liefert die Tabelle oben bezüglich des axiomatischen und des praktischen Begriffs Wahrscheinlichkeit.

Kritisch sind folgende Formulierungen, die man sinngemäß in vielen Büchern findet

- „Ich kann nicht genau erklären, was Wahrscheinlichkeit in der Realität konkret bedeutet. (…) Die Wahrscheinlichkeit, eine 3 zu würfeln, beträgt $\frac{1}{6}$“. Was das jetzt genau aussagt, bleibt rätselhaft.
- Oder es werden die objektive, die subjektive und ggf. weitere Arten von Wahrscheinlichkeiten erklärt und ab dann rechnerisch die axiomatische Wahrscheinlichkeit benutzt, ohne dass klar wird, welcher Zusammenhang dazwischen besteht.
- Oft wird beschrieben, welche Probleme es bei der konkreten Interpretation des Begriffs Wahrscheinlichkeit gibt und dass man heute deswegen den axiomatischen Zugang wählt, d. h., man verdrängt das Problem, statt es zu lösen. Man muss sich also nicht mit der praktischen Wahrscheinlichkeit beschäftigen, da sie ja mit mathematischen Modellen abgebildet wird. Das ist aber ein Widerspruch in sich, denn was man praktisch nicht kennt, kann man auch nicht theoretisch modellieren. Von was genau sind denn diese mathematischen Modelle ein Abbild? Diese Frage bleibt offen.
- Wieder andere gehen den einfachsten Weg: da in den Axiomen der Begriff Wahrscheinlichkeit vorkommt, wird das schon richtig sein, so dass man die Axiome auch auf praktische Probleme anwenden kann. Wie oben gezeigt wurde, ist das ein Trugschluss.

Und jetzt nochmal das Problem mehr mathematisch formuliert:
Hat man eine Abbildung f: R → K und kennt man die Menge K mit ihren Eigenschaften sehr gut, dann kann man diese Erkenntnisse nur dann auf R anwenden, wenn man f und R vorher eindeutig und klar definiert hat – sonst nicht. R steht für Menge der Informationen zur praktischen Wahrscheinlichkeit, K steht für Menge der Informationen zur axiomatischen Wahrscheinlichkeit, f ist die Abstraktion oder Modellbildung und f^{-1} ist die praktische Interpretation.



Oder eine mehr künstlerische Beschreibung des Problems: „Ich weiß zwar nicht genau, wie die Person aussieht, aber das macht nichts, denn ich habe hier ja ein Modell dieser Person erstellt. Und da das Modell einen Lockenkopf hat, hat die reale Person auch einen Lockenkopf“. Dieser Logik würde vermutlich niemand folgen wollen.

Es würde also weniger Verwirrung um den Begriff der Wahrscheinlichkeit geben, wenn man in den Axiomen den Begriff „Wahrscheinlichkeit“ durch einen neutralen Begriff wie „normiertes Maß“ ersetzen würde, denn dann passen mathematisches Modell und die praktische Interpretation „relativer Anteil“ besser zusammen – siehe auch [HR], Kapitel 3.1.
Auch die Visualisierung durch Mengen-Diagramme macht deutlich, dass „relativer Anteil“ der bessere Begriff ist, denn die Mengen-Diagramme stellen relative Flächenanteile dar.
Die Wahrscheinlichkeit kommt erst dann ins Spiel, wenn man berücksichtigt, dass relative Anteile zur Bestimmung einer praktischen Wahrscheinlichkeit beitragen können (siehe Kapitel 5.8).

Und noch ein – zugegebenermaßen nicht zwingender – Grund, warum „Wahrscheinlichkeit“ in den Axiomen eine unglückliche Wortwahl ist: Würde man in der axiomatischen Definition das Wort Wahrscheinlichkeit durch „?“ ersetzen, so würde vermutlich niemand auf die Idee kommen, dass es sich um ein Modell des Begriffs Wahrscheinlichkeit handelt – es sei denn, man wusste es sowieso vorher schon.

(8.1.-7) ***Andere Interpretation***
Kolmogoroff hatte – wie oben bereits erwähnt – darauf hingewiesen, dass seine Axiome auch ganz andere praktische Interpretationen zulassen.
Versucht man die Axiome praktisch zu interpretieren, so sieht man, dass die Axiome durch beliebige relative Anteile erfüllt werden:

- E ist das Ganze
- Die Elemente von E sind atomare Teile des Ganzen
- Die Teilmengen von E, also die Ereignisse, sind Teile des Ganzen
- Die Wahrscheinlichkeit ist der relative Anteil am Ganzen

(8.1.-8) ***Beispiel***

Bezeichnungen wieder wie bei den Kolmogoroffschen Axiomen.

- E sei eine Torte.
- Die Elemente („Elementare Ereignisse“) sind die Moleküle oder die Krümel; sie treten nicht explizit in Erscheinung.
- A sei die untere Hälfte und B die linke Hälfte der Torte, also „zufällige Ereignisse“.
- P misst den relativen Anteil am Volumen oder am Preis oder an den Kalorien oder an der Anzahl der Blaubeeren oder ….

Alle Axiome sind erfüllt und es gilt:

$P(E) = 1$
$P(A) = P(B) = 0{,}5$
$P(A \cup B) = 0{,}75$
$P(A \cap B) = 0{,}25$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Statt zu sagen „das ist eine halbe Torte“ könnte man gemäß den Axiomen also auch sagen „die Wahrscheinlichkeit dieses zufälligen Ereignisses ist 0,5“, was aber zu Kommunikationsschwierigkeiten mit dem/r Bäckereifachverkäufer/in führen dürfte.

Die Kolmogoroffschen Axiome beschreiben aber auch völlig unsinnige reale Darstellungen. Wenn sich alle einig sind, dass es sich bei A und B tatsächlich um halbe Anteile handelt und jemand kommt hinzu und behauptet:

$P(A) = 0{,}9$
$P(B) = 0{,}1$
$P(A \cup B) = 0{,}95$
$P(A \cap B) = 0{,}05$,

so ist das mit den Kolmogoroffschen Axiome konform, aber inhaltlich falsch. Die Kolmogoroffschen Axiome bilden also auch beliebigen Unsinn ab, sofern nur die Axiome erfüllt sind.

Die Kolmogoroffschen Axiome beschreiben also auch relative (richtige oder falsche) Anteile eines Ganzen, relative Anteile sind Beispiele für Wahrscheinlichkeiten, kurz: ein relativer Anteil ist eine Wahrscheinlichkeit, insbesondere ist eine relative Häufigkeit eine Wahrscheinlichkeit.

Auch hier wird ein Unterschied zur Umgangssprache und zur Realität deutlich: in der Umgangssprache ist ein relativer Anteil kein Beispiel für eine Wahrscheinlichkeit, sondern etwas Objektives, jeder kommt grundsätzlich zu demselben Ergebnis. Wahrscheinlichkeiten haben jedoch auch eine subjektive Komponente. Bei Fragestellungen aus der Realität können relative Anteile neben anderen Einflussfaktoren dazu benutzt werden, um Wahrscheinlichkeiten zu ermitteln (siehe Kapitel 5.8).
Hat man 1000-mal gewürfelt und alle Ergebnisse notiert, so kann man fragen: „wie hoch war die relative Häufigkeit der Augenzahl 3?“. Grundsätzlich wird jeder dieselbe Antwort geben und die hat mit Wahrscheinlichkeiten nichts zu tun. Würde man aber die Kolmogoroffschen Axiome nutzen, so würde in Frage und Antwort der Begriff „Wahrscheinlichkeit“ statt „relative Häufigkeit“ vorkommen und das würde inhaltlich nicht passen. In der Praxis kommt die Wahrscheinlichkeit erst ins Spiel, wenn man z. B. fragt: „Wie hoch ist die Wahrscheinlichkeit, beim nächsten Wurf eine 3 zu würfeln?“ oder „Ich wähle zufällig eines der notierten Ergebnisse aus. Wie groß ist die Wahrscheinlichkeit für eine 3?“.
Also:

- Axiomatisch: relativer Anteil ist ein Beispiel für Wahrscheinlichkeit
- Praktisch: relativer Anteil ist kein Beispiel für Wahrscheinlichkeit, sondern ein eigenständiger Begriff. Relative Anteile können einen Beitrag dazu leisten, eine Wahrscheinlichkeit zu ermitteln.

Und schließlich: die Kolmogoroffschen Axiome lassen selbst bei dem trivialen Beispiel des Münzwurfs unendliche viele Wahrscheinlichkeiten zu, also z. B.:

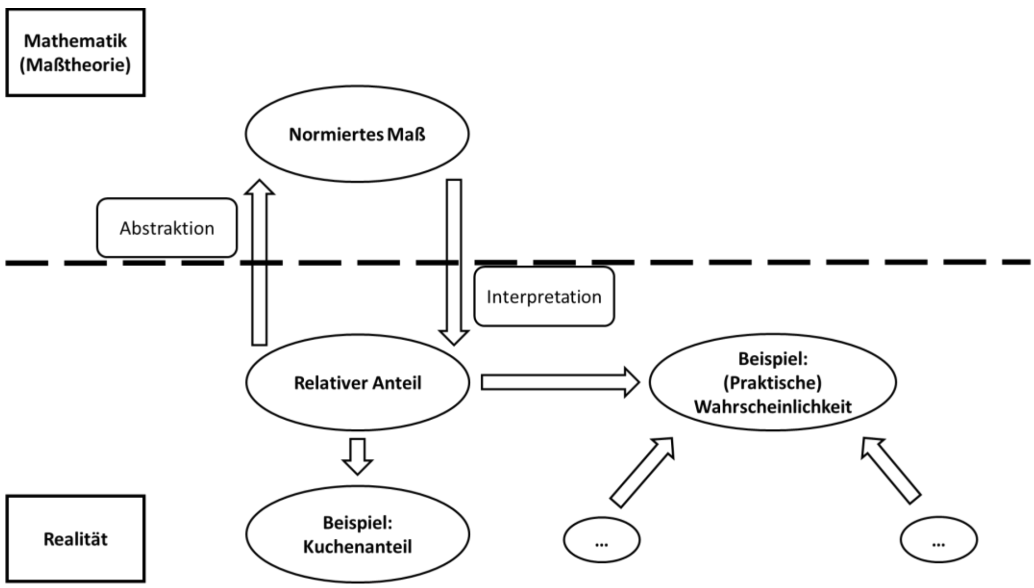
P(„Kopf“) = 0,1 und P(„Zahl“) = 0,9

oder umgekehrt oder etwas ganz Anderes. Es gibt keinen Hinweis, wie man eine Wahrscheinlichkeit ermitteln kann.

(8.1.-9) ***Grafik***

Die folgende Grafik fasst die Erläuterungen auf einen Blick zusammen:

- Wahrscheinlichkeiten lassen sich unter anderem aus relativen Anteilen ermitteln
- Relative Anteile lassen sich durch normierte Maße modellieren
- Für relative Anteile gibt es viele Beispiele



(8.1.-10) ***Ergebnis***

Es gibt drei gewichtige Gründe, warum die Axiome nach Kolmogoroff – im Gegensatz zur praktischen Wahrscheinlichkeit – für die Beschreibung der Realität nicht direkt sinnvoll nutzbar sind:

1. Die Axiome beschreiben nicht, was Wahrscheinlichkeit konkret in der Realität bedeutet
2. Die Axiome liefern keinen Hinweis darauf, wie man eine Wahrscheinlichkeit ermittelt
3. Die Axiome lassen viele Interpretationen zu, auch Interpretationen, die nichts mit praktischen Wahrscheinlichkeiten zu tun haben oder völlig unsinnig sind.

Der Begriff Wahrscheinlichkeit sollte also – um Verwirrungen zu vermeiden – im Rahmen der Maßtheorie durch „normiertes Maß“ oder ein beliebiges anderes Wort ersetzt und ansonsten nur in praktischen Zusammenhängen benutzt werden.

8.2 Der „ideale Würfel“

Die Ausdrücke „idealer Würfel“, „fairer Würfel“, „symmetrischer Würfel“, „regelmäßiger Würfel“, „echter Würfel“ oder „perfekter Würfel“ – im Folgenden zusammenfassend nur „idealer Würfel“ – werden benutzt, um deutlich zu machen, dass jede Augenzahl dieselbe Wahrscheinlichkeit hat.

Wer so argumentiert, berücksichtigt nicht, dass es in der Stochastik um Ergebnisse eines Vorgangs geht – ein Würfel ist aber kein Vorgang. Es muss also ein Vorgang definiert werden und nicht nur ein beim Vorgang benutzter Gegenstand.

Der Begriff „idealer Würfel“ ist also ungeeignet, denn tatsächlich ist es leicht, mit einem idealen Würfel bestimmte Augenzahlen bevorzugt zu würfeln. Man kann das Würfelergebnis mit etwas Übung in einer gewünschten Weise beeinflussen, indem man z. B. …

- … als Unterlage tiefen Sand wählt, also z. B. am Strand oder in einer Sandkiste würfelt
- … den Würfel aus geringer Höhe auf die Unterlage fallen lässt
- … einen durchsichtigen Würfelbecher benutzt
- … einen Würfelbecher benutzt, der so klein ist, dass sich der Würfel darin nicht bewegen kann
- … geeignete Systeme einsetzt, die ein Gebläse so steuern, dass der Würfel mit der gewünschten Seite nach oben liegen bleibt
- … (der Phantasie sind keine Grenzen gesetzt)

Am sichersten und einfachsten ist die Kombination aus mehreren dieser Methoden, also z. B. den idealen Würfel aus geringer Höhe in tiefen Sand fallen lassen oder so …
Es wäre vielleicht ein interessantes Thema für ein Seminar, in dem die Studierenden ihre Ideen, wie man mit einem idealen Würfel gezielt bestimmte Augenzahlen erzielen kann, präsentieren können.

Will man praktisch untersuchen, ob beim Würfeln tatsächlich keine Augenzahl bevorzugt wird, so muss der gesamte Würfelvorgang betrachtet werden und zumindest Folgendes gewährleistet sein:

a. Der Würfel ist ideal, d. h., er ist vollkommen homogen, die Seiten, Kanten und Ecken sind gleichartig gestaltet und haben die gleiche Oberflächenbeschaffenheit. Es gibt keine Unwucht, keine Seite ist klebrig. Die Oberflächenbeschaffenheit muss vor jedem Wurf erneut überprüft werden. Der Würfel darf nicht magnetisch sein.

b. Ecken und Kanten des Würfels sind abgerundet, damit sich der Würfel leicht überschlägt. Ein idealer Würfel im mathematischen Sinne ist ungünstig.

c. Die Unterlage ist eben, fest und rau, kein tiefer Sand und keine wellige Tischdecke, so dass sichergestellt ist, dass sich der Würfel mehrmals überschlägt und dass eindeutig bestimmbar ist, welche Augenzahl oben liegt.

d. Vor jedem Wurf ist die Hand, die den Würfel nimmt, auf Klebrigkeit (z. B. durch Chips oder Schokolade oder fettige Pommes) zu überprüfen und ggf. zu reinigen.

e. Die Würfelbewegung muss mit Schwung durchgeführt werden (also nicht einfach senkrecht fallen lassen), so dass sich der Würfel mehrmals überschlägt. Oder man würfelt mit einem nicht durchsichtigen und ausreichend großen Würfelbecher.

f. Der Würfelvorgang darf nicht durch äußere Kräfte wie Magnetfelder oder Luftbewegungen beeinflusst werden.

g. Es müssen physikalische Kräfte wirken, so dass der Würfel auf der Unterlage landet und dort verbleibt:

- An Land: der Würfel darf nicht eine geringere Dichte als die Luft haben, also z. B. nicht mit Gas gefüllt sein, sonst schwebt er nach oben weg.
- Unter Wasser: Das Würfeln könnte problematisch werden, insbesondere wenn der Würfel aus Holz ist. Für das Würfeln in der Badewanne muss es also bestimmte Zusatzregeln geben.
- Im Weltraum / in einer Raumstation: auch hier kann der Würfel leicht wegschweben, wie das Foto des Astronauten Reid Wiseman von der ISS zeigt.

h. Es werden nur Versuche, bei denen der Würfel eindeutig ablesbar auf der Unterlage liegt, gewertet. Verbleibt der Würfel nicht auf der Unterlage, sondern fällt z. B. vom Tisch oder bleibt wider Erwarten schräg liegen („brennt"), so wird der Wurf nicht gezählt. Anderenfalls hätte man 7 mögliche Versuchsergebnisse und die Wahrscheinlichkeiten für die einzelnen Augenzahlen wären zwar gleich, aber $< \frac{1}{6}$.
Es gibt auch Würfelspiele, bei denen solche Versuche berücksichtigt werden, wie z. B. das Würfelspiel „Zehntausend".

i. … und dann darf auch nicht mit einem Eiswürfel auf einer heißen Herdplatte als Unterlage gewürfelt werden oder mit einem Suppenwürfel über einer nassen Oberfläche oder andere lustige Experimente. Auch hier ist wieder Kreativität gefordert.

Im Allgemeinen wird nur „idealer Würfel" vorausgesetzt, also nur das, was unter Teil a. erläutert wurde. Auch im Mathematikabitur in Niedersachsen im Jahre 2016, Aufgabe 2A, ist nur „idealer Würfel" vorausgesetzt.

Wenn man „idealer Würfel" zulässt, so könnte man stattdessen genauso gut auch schreiben
- Würfeln auf einer idealen Unterlage
- Würfeln mit einem idealen Würfelbecher
- Würfeln mit einer idealen Armbewegung
- …

Statt „idealer Würfel" muss es also zumindest heißen: „ideales Würfeln" oder „ideales Würfelspiel". Wenn wie allgemein üblich in Büchern steht, dass ein Versuch unter gleichen Bedingungen mehrfach wiederholt wird, dann sind alle wesentlichen Bedingungen damit gemeint und nicht nur die eine Bedingung „idealer Würfel".

Aus der Liste der Bedingungen oben folgt, dass es in der Praxis kaum möglich ist, festzustellen, ob ideal gewürfelt wird. Es ist daher wesentlich realistischer, wenn man nur voraussetzt, dass man jedem Würfelergebnis von vorneherein dieselbe Wahrscheinlichkeit beimisst. Woher diese Einschätzung kommt, kann offen bleiben.

Die Qualität der Annahmen beim Würfeln (und analog bei Vorgängen wie dem Münzwurf, der Ziehung der Lottozahlen oder der Ziehung von Kugeln aus einer Trommel) kann also zusammenfassend folgendermaßen charakterisiert werden:
1. Optimal: es wird vorausgesetzt, dass jedes Ergebnis dieselbe Wahrscheinlichkeit hat. Woher diese Annahme kommt, kann offen bleiben.
2. Mittelmäßig sinnvoll: ideales Würfeln bzw. ideales Würfelspiel. Ob das Würfeln ideal ist, ist aber praktisch nur schwer feststellbar.
3. Nicht sinnvoll: die übliche Formulierung „Würfeln mit idealem Würfel".

8.3 Das Indifferenzprinzip

In Hinweis (5.8.-06) wurde bereits dargestellt: kennt man bei der Bestimmung der Wahrscheinlichkeit einer Aussage A Gründe, die für die Richtigkeit der Aussage sprechen, aber keine Gründe, die dagegen sprechen, so kann man die Aussage als eher wahr einstufen, also $P(A) > 0{,}5$.

In manchen Situationen gilt folgender Sonderfall, der als Indifferenzprinzip bezeichnet wird: hat man Gründe, warum zwei Aussagen die gleiche Plausibilität haben, aber keine Gründe, warum sie nicht die gleiche Plausibilität haben sollten, so nimmt man ihre Wahrscheinlichkeiten als gleich an. Genauer müsste man sagen: die Wahrscheinlichkeit, dass die beiden Wahrscheinlichkeiten gleich sind, ist größer als 0,5. Woher diese Einschätzung kommt, spielt rechnerisch keine Rolle.

(8.3.-1) ***Hinweise***

- Das Indifferenzprinzip ist nur sinnvoll, bevor man die Wahrscheinlichkeiten berechnet hat, denn ansonsten folgt automatisch eine der drei Relationen $P(A) > P(B)$, $P(A) < P(B)$, $P(A) = P(B)$.
- Das Indifferenzprinzip kann auf verschiedene Weisen zur Ermittlung von Wahrscheinlichkeiten genutzt werden:
  - Unter den Voraussetzungen des Indifferenzprinzips folgt aus der Subadditivität:
    $$P(A) = P(B) = \frac{P(A \text{ oder } B) + P(A \text{ und } B)}{2}$$
    Schließen sich zusätzlich A und B gegenseitig aus, so ergibt sich vereinfacht
    $$P(A) = P(B) = \frac{P(A \text{ oder } B)}{2}$$
    Das lässt sich leicht auf n Aussagen verallgemeinern:
    Schließen sich die n Aussagen A_i ($i = 1, \ldots, n$) paarweise aus, so ist unter den Bedingungen des Indifferenzprinzips für jedes i:
    $$P(A_i) = \frac{P(A_1 \text{ oder } \ldots \text{ oder } A_n)}{n}$$
    Dieses Ergebnis ist analog zu Beispiel (5.7.3.-2).
  - Kennt man nur $P(B)$, nicht aber $P(A)$, so ist unter den Bedingungen des Indifferenzprinzips $P(A) = P(B)$, man kann also $P(A)$ ermitteln.

(8.3.-2) ***Hinweis***

Manchmal wird nur der zweite Teil der Voraussetzungen des Indifferenzprinzips genannt (siehe z. B. [TW], Seite 12): „Gibt es keinen Grund, A für wahrscheinlicher oder unwahrscheinlicher zu halten als B, dann gilt: $P(A) = P(B)$“, formal: „Gibt es keinen Grund für $P(A) \neq P(B)$, so folgt $P(A) = P(B)$“.
Das Problem ist dann aber, dass es sein kann, dass man über die Plausibilität der Aussagen nichts weiß und damit ergeben sich die drei Varianten:

- Variante 1: „es gibt keinen Grund für $P(A) > P(B)$ oder $P(A) < P(B)$, also ist $P(A) = P(B)$“
- Variante 2: „es gibt keinen Grund für $P(A) > P(B)$ oder $P(A) = P(B)$, also ist $P(A) < P(B)$“
- Variante 3: „es gibt keinen Grund für $P(A) = P(B)$ oder $P(A) < P(B)$, also ist $P(A) > P(B)$“

Es gibt keinen Grund, warum man ausgerechnet Variante 1 den Vorzug geben sollte. In der Konsequenz heißt das, dass mit dieser Methode keine Aussagen über die Wahrscheinlichkeit möglich sind.

(8.3.-3) ***Beispiel***
Eine Reißzwecke hat auf harten ebenen Boden zwei mögliche Lagen: entweder zeigt die Spitze nach oben oder schräg nach unten. Die Auswahl der beiden Lagen erfolgt wie üblich durch einen Wurf (siehe Beispiel (5.7.3.-9)). Gegeben seien die beiden Aussagen
A = „Spitze zeigt nach oben“
B = „Spitze zeigt schräg nach unten“
Wenn nichts weiter bekannt ist, so kann man nichts über die Wahrscheinlichkeiten aussagen. Dann sind aber alle drei Varianten im vorangegangenen Hinweis (8.3.-2) möglich, man kann also mit dieser Methode keine Aussage über die Wahrscheinlichkeit machen.
Aber man kann natürlich wie in Beispiel (5.7.3.-9) irgendwelche Werte für P(A) und P(B) mit P(A) + P(B) = 1 annehmen.

(8.3.-4) ***Beispiel***
Es wird gewürfelt und außer der allgemeinen Tatsache, wie Würfeln funktioniert, weiß man nichts über diesen Prozess. Dann kann man folgende Meinungen vertreten:

- Optimist: „Ich weiß nichts, also gehe ich davon aus, dass beim nächsten Wurf die Augenzahl bevorzugt wird, die mir nützt bzw. dem Gegner schadet“.
- Realist: „Ich weiß nichts, also gehe ich davon aus, dass beim nächsten Wurf keine Augenzahl bevorzugt wird, jede Augenzahl hat dieselbe Wahrscheinlichkeit“.
- Pessimist: „Ich weiß nichts, also gehe ich davon aus, dass beim nächsten Wurf die Augenzahl bevorzugt wird, die mir schadet bzw. dem Gegner nützt“.

In allen Fällen werden Unterstellungen gemacht, denen man folgen kann oder nicht. Die Annahmen sind problemlos und können als Faktoren in die jeweiligen unterschiedlichen Bewertungen einfließen – aber natürlich nicht in dieselbe Bewertung.
Meistens wird aber in der Literatur der Fall „Realist“ angenommen:

- A-priori: da der Würfel symmetrisch ist, scheint keine Seite beim Würfeln bevorzugt zu werden,
- A-posteriori: aus Erfahrung weiß ich, dass jede Augenzahl annähernd gleich häufig auftritt, wenn man oft würfelt,

… also sind die Wahrscheinlichkeiten für jedes Ergebnis gleich.

(8.3.-5) ***Beispiel***
In manchen Veröffentlichungen wird folgendes Buch-Paradoxon genannt (z. B. [TW2]; [AE], Kapitel 1.3):
Ich suche ein bestimmtes Buch in einer Bibliothek, weiß aber nicht, welche Farbe der Umschlag hat. Wenn ich mir nur die Frage stelle, ob das Buch blau oder nicht blau ist, dann ist nach dem verkürzten Indifferenzprinzip gemäß Hinweis (8.3.-2), Variante 1:

$$P(\{b\}) = P(\{nb\}) = 0{,}5.$$

Dasselbe gilt aber auch für grün und rot, also

$$P(\{g\}) = P(\{ng\}) = 0{,}5,$$
$$P(\{r\}) = P(\{nr\}) = 0{,}5.$$

Da sich die Ereignisse {b}, {g} und {r} gegenseitig ausschließen, ist die Wahrscheinlichkeit, dass das Buch blau oder grün oder rot ist:

$$P(\{b, g, r\}) = P(\{b\}) + P(\{g\}) + P(\{r\}) = 1{,}5$$

und das ist ein Widerspruch.
Das Paradoxon entsteht aber nur dadurch, dass man bei den verschiedenen Farben unterschiedliche Wahrscheinlichkeitsräume zugrunde legt.
Der Widerspruch löst sich sofort auf, wenn man die korrekte Methode zur Bestimmung einer Wahrscheinlichkeit gemäß (6.3.-4) anwendet, also:

- erst S bestimmen,
- dann die Elementarwahrscheinlichkeiten mit demselben Bewertungsverfahren bestimmen,
- dann die Wahrscheinlichkeit für das Ereignis ausrechnen

Wendet man das auf das Buch-Paradoxon an, so ergeben sich unter der Annahme, dass die einzelnen Wahrscheinlichkeiten gleich sind, folgende Varianten:

- $S_1 = \{b, nb\}$, $\mathfrak{A}_1 = \mathfrak{P}(S_1)$, $P_1(\{b\}) = P_1(\{nb\}) = \frac{1}{2}$.
- $S_2 = \{g, ng\}$, $\mathfrak{A}_2 = \mathfrak{P}(S_2)$, $P_2(\{g\}) = P_2(\{ng\}) = \frac{1}{2}$.
- $S_3 = \{r, nr\}$, $\mathfrak{A}_3 = \mathfrak{P}(S_3)$, $P_3(\{r\}) = P_3(\{nr\}) = \frac{1}{2}$.
- $S_4 = \{b, g, r\}$, $\mathfrak{A}_4 = \mathfrak{P}(S_4)$, $P_4(\{b\}) = P_4(\{g\}) = P_4(\{r\}) = \frac{1}{3}$.

Natürlich bekommt man Probleme, wenn man die einzelnen Wahrscheinlichkeiten bei unterschiedlichen Bedingungen ermittelt. Das wäre so, als ob Stiftung Warentest unterschiedliche Mozzarellasorten vergleichen und dabei je nach Sorte unterschiedliche Bewertungsverfahren anwenden würde. Ein solcher Vergleich wäre unseriös.

(8.3.-6) ***Hinweis***
Eine andere Formulierung des Indifferenzprinzips findet sich bei [AE], Kapitel 1.3: „… das Prinzip der Indifferenz … besagt, dass wir alle Ereignisse als gleichwahrscheinlich betrachten sollten, solange wir nicht wissen, mit welcher Wahrscheinlichkeit eines von mehreren Ereignissen eintreten wird."
In dieser Formulierung kann das Prinzip falsch sein: auch wenn man z. B. die Wahrscheinlichkeiten im Lotto nicht kennt, so weiß dennoch jeder, dass die Wahrscheinlichkeit, 3 Richtige zu haben, größer ist, als die Wahrscheinlichkeit, 6 Richtige zu haben, die Wahrscheinlichkeiten können also nicht gleich sein. Oder dass die Wahrscheinlichkeit, dass „1" gezogen wird, größer ist, als die Wahrscheinlichkeit, dass „3, 24, 49" gezogen wird. In diesen Fällen kennt man die Wahrscheinlichkeiten nicht, man weiß aber trotzdem, dass sie nicht gleich sein können.

8.4 Zusammenfassung: vom Mozzarella zum Pasch

In den folgenden Stufen wird der lange Weg von einer beliebigen quantitativen Bewertung bis zur Zufallsvariablen zusammengefasst. Die axiomatische Definition der Wahrscheinlichkeit kommt dabei nicht vor, da sie nicht benötigt wird.

Stufe 1: Quantitative Bewertungen
Es gibt vermutlich nichts, was nicht quantitativ bewertet werden kann. Dabei werden Objekte bezüglich einer gewissen Eigenschaft mit Zahlen bewertet. Auf dieser Basis lässt sich eine Rangordnung definieren und Durchschnitte bilden, weitere Berechnungen sind im Allgemeinen nicht sinnvoll.
Beispiel: Quantitative Bewertung der Qualität von Mozzarella durch Stiftung Warentest

Stufe 2: Subadditive Bewertungen
Manchmal ist es sinnvoll, Objekte zu kombinieren und dann diese Kombination bezüglich derselben Eigenschaft quantitativ zu bewerten. In einigen dieser Fälle ergibt sich als Bewertungsregel von kombinierten Objekten die Subadditivität.

Stufe 3: Wahrscheinlichkeiten von Aussagen

Die Wahrscheinlichkeit von Aussagen ist ein Beispiel einer subadditiven Bewertung. Die Definition und das Verfahren sind auf beliebige Aussagen anwendbar, können aber im Einzelfall schwierig anzuwenden sein. Relative Anteile können ein Teil dieser Bewertung sein.
Statt Wahrscheinlichkeit der Aussage schreibt man P(Aussage).

Stufe 4: Ereignisse als Sonderfälle von Aussagen
Ereignisse lassen sich als Mengen von Ergebnissen darstellen. Um festzustellen, ob ein Ereignis eingetreten ist, muss nur eine einfache Checkliste abgearbeitet werden. Mengen sind erheblich einfacher zu behandeln als Aussagen. Die Anzahl möglicher Ereignisse wächst exponentiell mit der Anzahl der Ergebnisse: n Ergebnisse → bis zu 2^n Ereignisse.
Statt P(Aussage) schreibt man P(Ereignis) = P(Menge von Ergebnissen).

Stufe 5: Auswahl interessierender Ereignisse über Zufallsvariable
Aus der großen Menge der möglichen Ereignisse werden nur solche mit ganz bestimmten Eigenschaften ausgewählt und über Zufallsvariable beschrieben. Damit kann man eine drastische Reduzierung an Komplexität erzielen, denn die uninteressanten Ereignisse werden ausgeblendet. Man erhält so eine neue stark vereinfachte Ergebnismenge mit einer neuen vereinfachten Wahrscheinlichkeitsbewertung.
Zufallsvariable sind Variable, die aber im Allgemeinen nichts mit Zufall zu tun haben.
Statt P(Ereignis) schreibt man P(Bedingungen an Zufallsvariable).
Beispiel: Beim Würfeln mit 2 Würfeln sei nur Pasch interessant. Über die Zufallsvariable werden nur $\frac{1}{6} \approx 16{,}7$ % der Elementarereignisse und $\frac{2^6}{2^{36}} \approx 10^{-9}$ = 0,0000001 % der Ereignisse berücksichtigt.

9 Quellen

[AE]
Eckhart Arnold: Vorlesungsskript. Grundlagen des Entscheidens I, auf: https://eckhartarnold.de/papers/2009_Vorlesung_Entscheidungstheorie/Vorlesung_Entscheidungstheorie.html [03.02.2017]

[AI]
Amnesty International, „Refugees Welcome Survey 2016 - The Results“, https://www.amnesty.org/en/latest/news/2016/05/refugees-welcome-survey-results-2016 [29.10.2016]

[AT]
Thomas Augustin, „Wahrscheinlichkeitsverteilungen psychologischer Merkmale als Meßergebnisse: Ein Beitrag zur probabilistischen Meßtheorie“. Inaugural-Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultät II (Psychologie und Pädagogik) der Universität Regensburg, 2002

[BE]
Ehrhard Behrends, „Elementare Stochastik“, Springer Spektrum, 2012

[BF] Allgemeine deutsche Real-Encyklopädie für die gebildeten Stände. Conversations-Lexikon. Leipzig: F. A. Brockhaus – 1868

[BG]
Günther Bourier, „Beschreibende Statistik“, 11. Auflage, Springer Gabler 2013

[BK]
Gerd Bosbach / Jens-Jürgen Korff, „Lügen mit Zahlen“, 2. Auflage, Heyne, 2012

[BP]
Bundeszentrale für Politische Bildung, „Die soziale Situation in Deutschland – Arbeitslose und Arbeitslosenquote“, http://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61718/arbeitslose-und-arbeitslosenquote [25.10.2016]

[DA]
Firma dacadoo, Gesundheitsindex, https://info.dacadoo.com/de/ [29.10.2016]

[DH]
Hans-Hermann Dubben / Hans-Peter Beck-Bornholdt, „Der Hund, der Eier legt“, 7. Auflage, rororo, 2013

[DU]
Dudenonline, „Stochastik“, http://www.duden.de/suchen/dudenonline/stochastik [29.10.2016]

[EP]
Peter P. Eckstein, „Repetitorium Statistik“, 5. Auflage, Gabler Verlag

[ER]
Sprachniveaustufen nach dem Gemeinsamen Europäischen Referenzrahmen, http://www.europaeischer-referenzrahmen.de/sprachniveau.php
[FE]

Fertigpackungsverordnung, http://www.gesetze-im-internet.de/bundesrecht/fertigpackv_1981/gesamt.pdf [15.10.2016]

[GG]
World Economic Forum, „Global Gender Gap Report 2016“, http://reports.weforum.org/global-gender-gap-report-2015/rankings [29.10.2016]

[GH]
Hans-Otto Georgii, „Stochastik“, 5. Auflage, de Gruyter, 2015

[HD]
Dirk Helbing u. a., „Digitale Demokratie statt Datendiktatur“, http://www.spektrum.de/news/wie-algorithmen-und-big-data-unsere-zukunft-bestimmen/1375933 [29.10.2016]

[HR]
Robert Hable, „Einführung in die Stochastik“, Springer Spektrum 2015

[IF]
ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e. V., ifo-Geschäftsklimaindex, http://www.cesifo-group.de/de/ifoHome/facts/Survey-Results/Business-Climate.html [29.10.2016]

[IP]
Intergovernmental Panel on Climate Change (IPCC) – Zwischenstaatlicher Ausschuss für Klimaänderung, „Klimaänderung 2013 – Zusammenfassung für politische Entscheidungsträger“, Seite 2, Anmerkung 2, https://www.ipcc.ch/pdf/reports-nonUN-translations/deutch/ar5-wg1-spm.pdf [29.10.2016]

[JE]
Edwin Thompson Jaynes, „Probability Theory: The Logic of Science“, edited by G. Lerry Bretthorst, Cambridge University Press, 2003

[KA]
Andrei Nikolajewitsch Kolmogoroff, „Grundbegriffe der Wahrscheinlichkeitsrechnung“, Ergebnisse der Mathematik und ihrer Grenzgebiete, Verlag von Julius Springer, Berlin, 1933

[KJ]
Jürgen Kuri, „Herrschaft der Algorithmen – Die Welt bleibt unberechenbar”, http://www.faz.net/aktuell/feuilleton/herrschaft-der-algorithmen-die-welt-bleibt-unberechenbar-1996485.html [15.10.2016]

[KK]
Karl-Rudolf Koch, „Einführung in die Bayes-Statistik”, Springer, 2000

[KO]
Gesetz zur Kontrolle und Transparenz im Unternehmensbereich (KonTraG) vom 27. April 1998, veröffentlicht im Bundesgesetzblatt Teil I, Nr. 24, vom 30.04.1998, Seite 786-794;
http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=//*%255B@attr_id='bgbl198s0786.pdf'%255D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl198s0786.pdf%27%5D__1477774700736 [29.10.2016]

[ME] Meyers enzyklopädisches Lexikon, Band 22 (Sn – Sud), 9. Auflage, korrigierter Nachdruck, Bibliographisches Institut, 1979

[MK] Meyers Konversationslexikon, Band 18 (Schöneberg – Sternbedeckung), 6. Auflage, Bibliographisches Institut, 1909

[MP]
Max-Planck-Institut für Bildungsforschung (mpib), „Unstatistik des Monats", https://www.mpib-berlin.mpg.de/de/presse/dossiers/unstatistik-des-monats [23.10.2016]

[OE]
OECD, „Besser leben – wie und wo?", http://www.oecdbetterlifeindex.org/de/ [29.10.2016]

[OM]
Markus Oestreich, Oliver Romberg, „Keine Panik vor Statistik!", 5. Auflage, Springer Spektrum, 2014

[PR]
Institut für musterbasierte Prognosetechnik, „Near repeat prediction method", Software precobs, http://www.ifmpt.de/ [29.10.2016]

[PV]
Dialika Neufeld, „Partnervermittlung – it(wert1==wert2){", Der Spiegel, Ausgabe 14/2012, http://www.spiegel.de/spiegel/print/d-84631754.html [29.10.2016]

[RU]
R+V Versicherungen, „Die Ängste der Deutschen", https://www.ruv.de/presse/aengste-der-deutschen/grafiken-die-aengste-der-deutschen-2015 [29.10.2016]

[SB] Statistisches Bundesamt, „Finanzen und Steuern – Jährliche Einkommensteuerstatistik – Sonderthema: Einkünfte aus Kapitalvermögen – Veranlagungsjahr 2009", Tabelle 8, Seite 19, Wiesbaden 2013, https://www.destatis.de/DE/Publikationen/Thematisch/FinanzenSteuern/Steuern/LohnEinkommensteuer/Einkommensteuerstatistik2140711097004.pdf?__blob=publicationFile [25.10.2016]

[SD]
D. Selzer-McKenzie, „Mode-Prognosen aus dem Computer", https://groups.google.com/forum/#!topic/de.etc.finanz.boerse/9K1WPBgsMgo [15.10.2016]

[SJ]
Josef Schira, „Statistische Methoden der VWL und BWL", 4. Auflage, Pearson, 2012

[SM]
Michael Sachs, „Wahrscheinlichkeitsrechnung und Statistik", 4. Auflage, Fachbuchverlag Leipzig im Carl Hanser Verlag

[SR]
Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, „Jahresgutachten 15/16", http://www.sachverstaendigenrat-wirtschaft.de/fileadmin/dateiablage/gutachten/jg201516/wirtschafts-gutachten/jg15_ges.pdf [29.10.2016]

[SU]

Survey Monkey, https://www.surveymonkey.de/mp/customer-satisfaction-surveys [29.10.2016]

[SZ]
Süddeutsche Zeitung Online vom 14.06.2016, http://www.sueddeutsche.de/sport/fifa-blatter-so-wird-bei-auslosungen-manipuliert-1.3032493 [25.01.2017]

[TP]
Philip E. Tetlock / Dan Gardner, „Superforecasting“, Seite 67, S. Fischer Verlag, 2016

[TW]
Wolfgang Tschirk, „Statistik: Klassisch oder Bayes“, Springer Spektrum, 2014

[TW2]
Wolfgang Tschirk, „The Principle of Indifference Does Not Lead to Contradictions“, International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

[VH]
Institute for Economics and Peace (IEP), Vision of Humanity, Global Peace Index, http://www.visionofhumanity.org [29.10.2016]

[WH]
„World Happiness Report“, http://worldhappiness.report [29.10.2016]

[WI]
Deutsches Wanderinstitut, „Wie wird ein Wandererlebnis messbar?“, http://www.wanderinstitut.de/deutsches-wandersiegel/kriterien [29.10.2016]

[WK]
Wikipedia, Stichwort „Wahrscheinlichkeit“, https://de.wikipedia.org/wiki/Wahrscheinlichkeit [29.10.2016]

[WM]
Max C. Wewl „Statistik im Bachelor-Studium der BWL und VWL“, 2. Auflage, Pearson, 2011

[WU]
Stiftung Warentest, „Wundpflaster“, https://www.test.de/Wundpflaster-Die-besten-Helfer-fuer-Ihre-Hausapotheke-1379513-1379677 [29.10.2016]

[ZN]
Nils Zurawski, „Big Data – Wider die Herrschaft der Algorithmen“, http://www.deutschlandradiokultur.de/big-data-wider-die-herrschaft-der-algorithmen.1005.de.html?dram:article_id=345175 [15.10.2016]