Lehr- und Handbücher der Statistik

Herausgegeben von
Universitätsprofessor Dr. Rainer Schlittgen

Bisher erschienene Werke:

Caspary/Wichmann, Lineare Modelle
Chatterjee/Price (Übers. Lorenzen), Praxis der Regressionsanalyse, 2. Auflage
Degen/Lorscheid, Statistik-Aufgabensammlung, 2. Auflage
Harvey (Übers. Untiedt), Ökonometrische Analyse von Zeitreihen, 2. Auflage
Harvey (Übers. Untiedt), Zeitreihenmodelle, 2. Auflage
Heiler/Michels, Deskriptive und Explorative Datenanalyse
Naeve, Stochastik für Informatik
Oerthel/Tuschl, Statistische Datenanalyse mit dem Programmpaket SAS
Pokropp, Lineare Regression und Varianzanalyse
Rinne, Wirtschafts- und Bevölkerungsstatistik
Rüger, Induktive Statistik, 3. Auflage
Schlittgen, Statistik, 5. Auflage
Schlittgen/Streitberg, Zeitreihenanalyse, 6. Auflage

Induktive Statistik

Einführung für Wirtschafts- und Sozialwissenschaftler

Von

Dr. Bernhard Rüger

Professor für Statistik
Universität München

3., überarbeitete Auflage

R. Oldenbourg Verlag München Wien

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Rüger, Bernhard:
Induktive Statistik : Einführung für Wirtschafts- und Sozialwissenschaftler / von Bernhard Rüger. - 3., überarb. Aufl. - München ; Wien : Oldenbourg, 1996
(Lehr- und Handbücher der Statistik)

ISBN 3-486-23543-5

Gesamtherstellung: R. Oldenbourg Graphische Betriebe GmbH, München

ISBN 3-486-23543-5

Inhaltsverzeichnis

Teil I: Wahrscheinlichkeitsrechnung

Teil II: Induktive Statistik

Teil I
Wahrscheinlichkeitsrechnung

Kapitel 1
Grundlagen

1.1 Zufallsexperimente

Im täglichen Leben und in nahezu allen Anwendungsgebieten von Natur- und Sozialwissenschaften werden oft Vorgänge beobachtet oder Versuche durchgeführt, die zufallsabhängig sind oder als vom Zufall abhängig betrachtet werden. Beispiele sind etwa das Wetter, Stauungen im Straßenverkehr, Warteschlangen vor Bedienungsschaltern, das Zustandekommen der Aktienkurse, der Verlauf einer Epidemie, die Durchführung einer Meinungsumfrage, die Qualitätskontrolle einer laufenden Produktion, die Entwicklung der Lage am Arbeitsmarkt, Veränderungen im Altersaufbau der Bevölkerung, usw. Diesen wirklichkeitsnahen und daher sehr komplexen Geschehnissen stehen einfachere Zufallsvorgänge aus der Theorie der Glücksspiele gegenüber, etwa Würfel-, Karten- oder Lottospiele; von besonderer Bedeutung für die Statistik ist das Experiment der zufälligen Ziehung einer Stichprobe aus einer Grundgesamtheit. Allen Beispielen gemeinsam ist, daß ein Vorgang abläuft oder ein Versuch durchgeführt wird, dessen Ergebnis nicht im voraus bestimmbar ist. Solch ein Vorgang bzw. Experiment wird **Zufallsexperiment** genannt; es besteht aus der Festlegung einer Versuchsanordnung und der Menge aller möglichen Ergebnisse, der sogenannten **Ergebnismenge**. Die Ergebnismenge wird mit Ω bezeichnet, jedes Ergebnis ω ist Element von Ω.

In den exakten Naturwissenschaften lassen sich viele Probleme (z.B. in der Physik der freie Fall oder die Bewegung einer Ladung in einem Magnetfeld) durch Experimente untersuchen, die bei genügender Sorgfalt durch Einhaltung konstanter Versuchsbedingungen stets das gleiche (oder doch annähernd gleiche, höchstens im Rahmen von Meßungenauigkeiten schwankende) Ergebnis bei mehrfacher Wiederholung des Versuchs aufweisen. Die Ergebnismenge eines solchen Experimentes besteht aus einem einzigen Element. Die deterministische Betrachtungsweise – *eine* Ursache hat unter fest gegebenen Bedingungen *eine* ganz bestimmte Wirkung – wird der Natur derartiger Vorgänge gerecht und hat für die Entdeckung von Gesetzen einen hohen Wert.

Anders ist die Situation in den Wirtschafts- und Sozialwissenschaften, aber auch in der Psychologie, Medizin und Biologie. In vielen Fällen erlaubt die ungeheure Komplexität der betrachteten Systeme und der darin ablaufenden Vorgänge weder die Einhaltung vorgegebener Versuchsbedingungen noch die Anwendung genauer Meßverfahren, oft läßt sich ein Experiment nicht einmal geplant durchführen, geschweige denn unter konstanten Bedingungen wiederholen. Eine rein deterministische Betrachtungsweise kann das Zusammenwirken ganzer Bündel von Ursachen und Einflußgrößen nicht erfassen und liefert praktisch keinen Erklärungswert für die betrachteten Phänomene. Zweckmäßiger ist es, solche Vorgänge mit Hilfe stochastischer Modelle zu erklären, die Experimente als Zufallsexperimente aufzufassen, deren Ergebnisse neben ihrer Abhängigkeit von einigen hauptsächlichen Faktoren (soweit erkennbar und erklärungsnützlich) *auch* als zufallsabhängig angesehen werden. Wie dabei deterministische von stochastischen Komponenten abgegrenzt werden, ob ein Vorgang überhaupt stochastische Komponenten enthält oder eher als rein deterministisch einzustufen ist, ist in diesem Zusammenhang eine eher pragmatische als erkenntnistheoretische Frage.

Zur Verdeutlichung der Regeln der Wahrscheinlichkeitsrechnung werden wir im folgenden oft sehr einfache Beispiele benutzen, in denen diese Fragen erst gar nicht auftauchen.

Beispiel 1.1

Ein Würfel wird geworfen. Die möglichen Ergebnisse dieses Zufallsexperimentes sind die sechs verschiedenen Augenzahlen, die Ergebnismenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Beispiel 1.2

Zwei (unterscheidbare) Würfel werden auf einmal oder ein Würfel wird zweimal hintereinander geworfen. Jedes Ergebnis ω dieses Zufallsexperimentes besteht aus einem Zahlenpaar (i, j), wobei i und j jeweils eine der sechs Augenzahlen sind. Die Ergebnismenge Ω besteht aus allen $6 \cdot 6 = 36$ solchen Zahlenpaaren. – Allgemein: Das Zufallsexperiment „ein Würfel wird n mal geworfen" besitzt eine Ergebnismenge mit 6^n Elementen, jedes Ergebnis ist ein n-Tupel von Augenzahlen.

Beispiel 1.3

Aus einer Grundgesamtheit von N Elementen werden zufällig und ohne Zurücklegen n Elemente gezogen, $1 \leqq n < N$, wobei es auf die Reihenfolge, in der die Elemente gezogen werden, nicht ankommen soll. (Man kann sich auch vorstellen, daß die n Elemente der Grundgesamtheit mit einem Griff entnommen werden.) Als Ergebnis dieses Zufallsexperimentes kann jede Teilmenge von n Elementen auftreten. In Abschnitt 3.1.1 wird gezeigt, daß es $\binom{N}{n}$ solche Teilmengen gibt. Die Ergebnismenge Ω, die Menge aller dieser Teilmengen, hat also $\binom{N}{n}$ Elemente. – Dazu als Zahlenbeispiel die Ziehung der Lottozahlen: Aus $N = 49$ Kugeln, die mit den Zahlen 1 bis 49 beschriftet sind, werden zufällig und ohne Zurücklegen $n = 6$ Kugeln gezogen; die Reihenfolge spielt keine Rolle. (Die sogenannte Zusatzzahl, die Ziehung einer siebten Kugel, wird außer acht gelassen.) Als Ergebnis kann jede Teilmenge von 6 Elementen, also z.B. $\{3, 11, 15, 24, 30, 42\}$ auftreten. Es gibt

$$\binom{49}{6} = \frac{49!}{6! \cdot 43!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 13983816$$

derartige Teilmengen. (Die „Chance" für „sechs Richtige" ist 1 zu 13983816. Diese Aussage werden wir später mit Hilfe der Laplaceschen Wahrscheinlichkeitsdefinition präzisieren.)

Beispiel 1.4

Eine Münze wird so oft geworfen, bis zum erstenmal „Zahl" erscheint; man interessiert sich für die Anzahl der dazu erforderlichen Würfe. Die Ergebnisse dieses Zufallsexperimentes sind die Zahlen 1, 2, 3, ... usw. Ergebnismenge ist die Menge der natürlichen Zahlen. Man kann Ω durch keine noch so große Zahl n beschränken; denn es ist möglich, daß n mal hintereinander „Kopf" erscheint, auch wenn dies für großes n sehr unwahrscheinlich ist. Im Unterschied zu den ersten drei Beispielen ist Ω hier eine (abzählbar) unendliche Menge.

Beispiel 1.5

Ein Kunde begibt sich mit einem Paket zur nächstgelegenen Poststelle, von der er weiß, daß sie nur einen Paketannahmeschalter hat. Er interessiert sich für seine Wartezeit, die er vor dem Schalter bis zu seiner Bedienung verbringen muß. Bevor er die Post betritt, kennt er die Länge der Schlange vor dem Paketschalter nicht. Seine Wartezeit kann daher zwischen 0 (falls kein Kunde vor dem Schalter steht) und irgendeiner maximalen Zeit T (auf die es uns hier nicht ankommt) schwanken. Mißt er die Wartezeiten in Minuten, so lautet die Ergebnismenge dieses Zufallsexperimentes $\{0, 1, 2, \ldots, T\}$. Man kann aber auch Fragen der Meßgenauigkeit außer acht lassen und die Zeit als kontinuierliche Größe betrachten (Zeitkontinuum), ein Standpunkt, der bei zeitabhängigen Zufallsvorgängen häufig eingenommen wird. In unserem Beispiel ergibt sich dann als Ergebnismenge die Menge aller reellen Zahlen zwischen 0 und T. (Dies ist eine überabzählbar unendliche Menge.)

1.2 Ereignisse

Zu einem Zufallsexperiment mit der Ergebnismenge Ω betrachtet man neben den einzelnen Ergebnissen vor allem Ereignisse: Ein (zufälliges) **Ereignis** ist eine Teilmenge von Ω. Wir bezeichnen Ereignisse mit großen lateinischen Buchstaben A, B usw.. Man sagt, das Ereignis A **tritt ein**, wenn das Zufallsexperiment ein Ergebnis ω liefert, das zu A gehört (Element von A ist).

Beispiele: In Beispiel 1.1 (Werfen eines Würfels) kann das Ereignis A, daß eine Fünf oder Sechs gewürfelt wird, betrachtet werden; $A = \{5, 6\}$. In Beispiel 1.2 (Werfen zweier Würfel) ist das Ereignis, daß ein Pasch gewürfelt wird, mit der Teilmenge $\{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$ identisch.

Auch ein einzelnes Ergebnis ω läßt sich als Ereignis auffassen, nämlich als die einelementige Teilmenge $\{\omega\}$ von Ω. Diese Ereignisse heißen **Elementarereignisse**. Die Ergebnismenge Ω selbst (als unechte Teilmenge von Ω) ist ein Ereignis, das bei jedem Ergebnis des Zufallsexperimentes eintritt, und heißt deshalb das **sichere Ereignis**. Entsprechend heißt die leere Menge $\emptyset$ (als Teilmenge von Ω) das **unmögliche Ereignis**.

Die mengentheoretischen Operationen (Bildung von Vereinigung, Durchschnitt und Komplement) lassen sich unmittelbar in die „ereignistheoretische Sprache" übersetzen. Sind A und B Teilmengen von Ω, so bedeuten

$A \cap B$ das Ereignis „A und B", das genau dann eintritt, wenn sich ein ω ergibt, das zu A und B gehört,

$A \cup B$ das Ereignis „A oder B", das genau dann eintritt, wenn sich ein ω ergibt, das zu A oder B gehört („oder" im nicht ausschließenden Sinn),

$\bar{A}$ das Ereignis „nicht A", das genau dann eintritt, wenn sich ein ω ergibt, das nicht zu A gehört,

$A \setminus B$ das Ereignis „A aber nicht B", das genau dann eintritt, wenn sich ein ω ergibt, das zu A aber nicht zu B gehört.

Zwei Ereignisse A und B, die nicht gemeinsam eintreten können, für die also $A \cap B = \emptyset$ (das unmögliche Ereignis) ist, nennt man **unvereinbar** oder **disjunkt**.

Beispiel 1.6

Wir betrachten in Beispiel 1.2 (Werfen zweier Würfel) die Ereignisse
A: „es wird ein Pasch gewürfelt"
B: „die Summe der erzielten Augen ist größer oder gleich 10"
C: „die Summe der erzielten Augen ist kleiner oder gleich 3".
B und C sind unvereinbare Ereignisse. Weiterhin gilt:
$A \cap B = \{(5,5), (6,6)\}$
$A \cap C = \{(1,1)\}$

$A \cup C = \{(1,1), (1,2), (2,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$
$A \setminus B = \{(1,1), (2,2), (3,3), (4,4)\}$

Jedem Ereignis A eines reinen Zufallsexperimentes soll eine *Wahrscheinlichkeit* zugeordnet werden, welche die Chance für das Eintreten von A beschreibt. Dazu ist es nicht immer zweckmäßig, *jede* Teilmenge von Ω als Ereignis in Betracht zu ziehen. Ist Ω eine überabzählbare Menge, so treten mathematisch tiefer liegende Schwierigkeiten auf, wenn man *jeder* Teilmenge von Ω eine Wahrscheinlichkeit zuordnen will: Ist Ω die Menge der reellen Zahlen oder ein Intervall reeller Zahlen, so könnte man unter Umständen keine stochastischen Modelle bilden, die das Zufallsexperiment adäquat beschreiben, z.B. keine Modelle mit stetigen Verteilungen*). Aus diesem Grund muß man gegebenenfalls neben der Ergebnismenge Ω auch noch eine Klasse (Menge) von Ereignissen festlegen, die dem Zufallsexperiment als zugehörig betrachtet werden soll. In der Wahl einer solchen Klasse ist man nicht frei: Zunächst wird verlangt, daß alle Ereignisse, die in praktischen Anwendungen auftreten können, zur Klasse gehören. Darüber hinaus sollen die genannten mengentheoretischen Operationen innerhalb der Klasse durchführbar sein, das heißt genauer: Mit einem Ereignis A soll stets auch $\bar{A}$ zur Klasse gehören und mit endlich oder abzählbar unendlich vielen Ereignissen sollen stets auch alle daraus herstellbaren Durchschnitte und Vereinigungen zur Klasse gehören. Eine Klasse mit diesen Eigenschaften nennt man eine **Ereignisalgebra** (genauer: Ereignis-σ-Algebra). Ist Ω endlich oder abzählbar unendlich, so kann stets die Menge aller Teilmengen, die sogenannte Potenzmenge von Ω als Ereignisalgebra gewählt werden, ohne daß die oben genannten Schwierigkeiten auftreten. Auch für überabzählbare Ergebnismengen Ω lassen sich im allgemeinen Ereignisalgebren konstruieren, die diese Schwierigkeiten vermeiden: Sie sind in der Regel kleiner als die Potenzmenge von Ω, aber doch groß genug, um alle in Theorie und Praxis benötigten Ereignisse zu enthalten. Ist Ω z.B. die Menge der reellen Zahlen, so läßt sich eine derartige Ereignisalgebra bilden, die alle Zahlenintervalle als Ereignisse enthält. Auf die damit zusammenhängenden Fragen können wir hier nicht eingehen. Wir setzen für das Folgende voraus, daß einem Zufallsexperiment mit seiner Ergebnismenge auch stets eine geeignete Ereignisalgebra zugeordnet ist, ohne diese eigens zu erwähnen.

1.3 Wahrscheinlichkeit

Es ist heute üblich, die Wahrscheinlichkeitstheorie als ein Teilgebiet der Mathematik zu betrachten und die Wahrscheinlichkeit mathematisch formal durch Axiome zu definieren. Diese Vorgangsweise, der wir uns anschließen werden, ist berechtigt, wenn man dabei im Auge behält, daß durch sie nur eine Theorie über den rechnerischen Umgang mit Wahrscheinlichkeiten, also eine Wahrscheinlichkeits*rechnung* und keine Theorie über die Wahrscheinlichkeit selbst zustande kommt. Eine umfassende Wahrscheinlichkeits*theorie* hätte ihren Gegenstand unter den drei folgenden Gesichtspunkten zu betrachten:

- die inhaltliche Bedeutung des Begriffes Wahrscheinlichkeit,
- die mathematisch formale Definition der Wahrscheinlichkeit,
- die numerische Bestimmung von Wahrscheinlichkeiten.

*) Vgl. Abschnitt 2.3

Die Frage nach der Bedeutung des Wahrscheinlichkeitsbegriffes steht in engem Zusammenhang mit dem Problem der Begründung induktiver Schlüsse, dem sogenannten Induktionsproblem. Ein Induktionsschluß ist ein Schluß vom Einzelnen, Besonderen auf etwas Allgemeines, Gesetzmäßiges. Ein solcher Schluß führt über die Aussage der Prämisse hinaus. Ihm kann daher keine absolute Gewißheit, sondern nur ein gewisser Grad an Sicherheit, genannt Wahrscheinlichkeit, zukommen. Das Induktionsproblem, insbesondere auch die Frage, mit welchem Wahrscheinlichkeitsbegriff der Sicherheitsgrad eines induktiven Schlusses ausgedrückt werden soll, führt uns in den Bereich der *Philosophie,* genauer der *induktiven Logik.*

Neben *Aristoteles* (384-322 v. Chr.) waren es in der Antike vor allem die *Epikureer,* orientiert an der Heilkunde, die sich diesem Bereich der Philosophie widmeten*). Ihre Schriften waren jedoch in Vergessenheit geraten und wurden erst im 19. Jahrhundert wiederentdeckt. Nach der Antike befaßte sich erst wieder die Philosophie der Neuzeit mit dem Induktionsproblem. Zu erwähnen sind hier aus der Zeit der Renaissance vor allem *Galileo Galilei* (1564-1642), der die Methode der Induktion in die Physik einführte, und *Francis Bacon* (1561-1626), der als Begründer einer empirisch eingestellten Wissenschaft gilt und den Utilitätsgesichtspunkt des Wissens betonte („Wissen ist Macht"). Ihren entscheidenden Anstoß aber erfuhr die induktive Logik durch *David Hume* (1711-1776), dessen Werk einen Höhepunkt des englischen Empirismus, der philosophischen Gegenbewegung zum Rationalismus, bildet. Er hat als erster die Problematik der induktiven Schlußweise in voller Schärfe aufgezeigt und Induktionsschlüsse als Wahrscheinlichkeitsschlüsse formuliert. In dieser englischen Tradition stehen auch *William Whewell* (1794-1866), der eine Geschichte der induktiven Wissenschaften schrieb, und *John Stuart Mill* (1806-1873), der ein System logischer Regeln für induktive Schlüsse aufstellte und als Begründer der induktiven Logik als formale Wissenschaft angesehen wird. Von den Philosophen unseres Jahrhunderts, die sich dem Induktionsproblem gewidmet haben, sind vor allem *Hans Reichenbach* (1891-1953), *Rudolf Carnap* (1891-1970), *Karl Raimund Popper* (1902-1994) und *Wolfgang Stegmüller* (1923-1991) zu nennen. Daneben haben sich auch innerhalb der *Grundlagenforschung der Statistik* verschiedene Theorien zur Stützung statistischer Schlüsse (im Sinne induktiver Schlüsse, man denke an einen Schluß von einer Stichprobe auf eine Grundgesamtheit) entwickelt, die mit unterschiedlichen Auffassungen des Wahrscheinlichkeitsbegriffes verbunden sind. In erster Linie ist hier der große englische Statistiker *Ronald Aylmer Fisher* (1890-1962) zu nennen, der den Anstoß und entscheidende Anregungen für derartige statistische Theorien gab.

Dieses Gebiet einer Wahrscheinlichkeitstheorie, das teils im Bereich der induktiven Logik, teils in der Grundlagenforschung der Statistik seine Heimat hat, können wir hier nicht behandeln. Nur andeutungsweise sollen in Abschnitt 1.3.4 einige Wahrscheinlichkeitsbegriffe skizziert und in Teil II verschiedene Konzepte für die Stützung statistischer Induktionsschlüsse vorgestellt werden. Dem Leser, dessen Neugier auf eine philosophische oder wissenschaftstheoretische Behandlung des Induktionsproblems geweckt wurde, werden vor allem die Arbeiten von *Carnap* [1945, 1952, 1967], *Carnap* und *Jeffrey* [1971], *Carnap* und *Stegmüller* [1972], *Essler* [1970], *Hacking* [1984], *Kneale* [1966], *Lakatos* [1968], *Reichenbach* [1932, 1935, 1983], *Stegmüller* [1973] und *de Wright* [1965] empfohlen. Auch die wichtigsten Werke der oben genannten Philosophen und Grundlagenforscher

*) Vgl. dazu Essler [1970]

auf dem Gebiet der Wahrscheinlichkeitsrechnung und Statistik sind im Literaturverzeichnis angegeben.

Der zweite Gesichtspunkt, die mathematisch formale Definition der Wahrscheinlichkeit, hat die Entwicklung von Rechenregeln und Gesetzen zum Ziel, denen die Wahrscheinlichkeiten gehorchen. Von dieser Wahrscheinlichkeitsrechnung wollen wir in Teil I die wichtigsten und einfachsten Bausteine kennenlernen. Ihre axiomatische Grundlage wurde 1933 durch den russischen Mathematiker *Andrej Nikolajewitsch Kolmogoroff* (1903-1987) geschaffen. Die Axiome werden in Abschnitt 1.3.3 vorgestellt.

Auf den dritten Gesichtspunkt, die Frage nach der numerischen Bestimmung von Wahrscheinlichkeiten, gehen wir in Abschnitt 1.3.5 ein.

In der geschichtlichen Entwicklung der Wahrscheinlichkeitstheorie sind diese drei Gesichtspunkte stets präsent, treten aber nur selten begrifflich voneinander getrennt auf. Insbesondere wurde in die mathematische Definition der Wahrscheinlichkeit die Art ihrer numerischen Bestimmung mit einbezogen. Die beiden auch heute noch wichtigsten Definitionen dieser Art gehen auf *Laplace* und *von Mises* zurück.

1.3.1 Die Laplacesche Wahrscheinlichkeitsdefinition

Der französische Mathematiker und Astronom *Pierre Simone de Laplace* (1749-1827) hat die zu seiner Zeit bekannten wahrscheinlichkeitstheoretischen Begriffe und Gesetze lehrbuchmäßig zusammengefaßt. Sein Hauptwerk „Théorie analytique des probabilités" erschien 1812, zwei Jahre später folgte „Essai philosophique sur les probabilités".

Als Ausgangspunkt der Wahrscheinlichkeitsrechnung gilt allgemein der Briefwechsel zwischen den französischen Mathematikern *Blaise Pascal* (1623-1662) und *Pierre de Fermat* (1601-1665) im Jahr 1654. Ausgelöst wurde er durch *Chevalier de Méré*, der *Pascal* einige Fragen über Glücksspiele stellte, unter anderem eine Frage nach der gerechten Aufteilung der Einsätze bei Abbruch eines Spiels*). Der holländische Physiker und Mathematiker *Christiaan Huygens* (1629-1695) muß die Ergebnisse dieses Briefwechsels**) kennengelernt haben. Er erweiterte sie zu dem 1657 veröffentlichten ersten Buch über Wahrscheinlichkeitsrechnung „De ratiociniis in ludo aleae". Dieses Buch regte *Jakob Bernoulli* (1654-1705), einen Baseler Mathematikprofessor, an, sich mit der Wahrscheinlichkeitsrechnung zu befassen. Sein Werk „Ars conjectandi", das 1713 posthum veröffentlicht wurde, enthält bereits so wichtige Dinge wie die Kombinatorik, die Binomialverteilung (Bernoulliverteilung) und das spezielle Gesetz der großen Zahlen (Theorem von Bernoulli). Schließlich sind vor Laplace noch zu erwähnen: der nach England geflüchtete Hugenotte *Abraham de Moivre* (1667-1754), dessen Buch „The doctrin of chances" [1718] in einer späteren Auflage zum erstenmal die Normalverteilung (als Grenzverteilung der Binomialverteilung) enthält, und der englische Pfarrer *Thomas Bayes* (1702-1761) mit seinem Aufsatz „An essay towards solving a problem in the doctrine of chances", 1763 posthum veröffentlicht, der das bekannte Bayessche Theorem zum Inhalt hat.

---

*) Diese Fragen sind Gegenstand einiger Übungsaufgaben am Ende des Kapitels.

**) Die Briefe sind in den „Œuvres complétes de Blaise Pascal" [1963] und den „Œuvres de Pierre de Fermat" [1891-1922] wiedergegeben. Man vergleiche auch Hald [1990], eine ausführliche Darstellung der Entwicklung der Wahrscheinlichkeitsrechnung von Cardano (1501-1576) bis de Moivre (1667-1754).

Alle diese Wahrscheinlichkeitstheoretiker bestimmten die Wahrscheinlichkeit (Chance) für das Eintreten eines Ereignisses A, indem sie die Anzahl der für A günstigen Fälle (der Fälle, in denen A eintritt) mit der Anzahl aller möglichen Fälle verglichen, wobei angenommen wurde, daß jeder Fall gleichmöglich (im Sinne von gleichwahrscheinlich) ist. Ein „Fall" ist gleichzusetzen mit einem Ergebnis des betreffenden Zufallsexperimentes. Die Kunst der Berechnung von Wahrscheinlichkeiten bestand (und besteht zum Teil heute noch) aus der Kunst des richtigen Abzählens. Dabei wird „richtig" gezählt, wenn man das Zufallsexperiment mit seiner Ergebnismenge so formuliert, daß alle Ergebnisse als gleichwahrscheinlich angesehen werden können. *Laplace* erhob diese Berechnungsart zur Definition der Wahrscheinlichkeit. Wir legen zunächst fest:

Ein **Laplace-Experiment** ist ein Zufallsexperiment mit einer endlichen Ergebnismenge Ω, bei dem jedes Ergebnis gleichwahrscheinlich ist.

Damit lautet die **Laplacesche Wahrscheinlichkeitsdefinition**:

In einem Laplace-Experiment ist die Wahrscheinlichkeit P(A) eines Ereignisses A definiert als

$$P(A) = \frac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}}$$

$$= \frac{\text{Anzahl der Elemente von A}}{\text{Anzahl der Elemente von }\Omega}$$

Die Wahrscheinlichkeit jedes einzelnen Ergebnisses ω, genauer jedes Elementarereignisses $\{\omega\}$, ist demnach in einem Laplace-Experiment

$$P(\{\omega\}) = \frac{1}{N} \quad , \quad N = \text{Anzahl der Elemente von } \Omega.$$

Beispiel 1.7

Wir betrachten das Zufallsexperiment „Werfen zweier Würfel", dessen Ergebnismenge Ω aus allen 36 Zahlenpaaren (i, j) besteht, wobei i und j jeweils eine der sechs Augenzahlen ist (vgl. Beispiel 1.2). Man kann annehmen, daß jedes dieser Ergebnisse gleichwahrscheinlich ist, so daß es sich hier um ein Laplace-Experiment handelt:

$$P(\{(i,j)\}) = \frac{1}{36} \quad \text{für jedes Ergebnis } (i,j).$$

Dabei wird für $i \neq j$ zurecht zwischen den Ergebnissen (i, j) und (j, i) unterschieden. Das Ereignis

A_{ij}: „Es werden die Augenzahlen i und j geworfen"

kann nämlich für $i \neq j$ auf zweierlei Weise zustandekommen: Der eine Würfel ergibt die Augenzahl i, der andere j oder umgekehrt. Also ist für $i \neq j$

$A_{ij} = \{(i, j), (j, i)\}$ kein Elementarereignis. Es gilt:

$$P(A_{ij}) = \frac{2}{36} = \frac{1}{18} \quad \text{für } i \neq j, \qquad \text{während sich für } i = j$$

$$P(A_{ii}) = P(\{(i,i)\}) = \frac{1}{36} \quad \text{ergibt.}$$

Hätte man die A_{ij} auch für $i \neq j$ als Elementarereignisse aufgefaßt und dementsprechend eine Ergebnismenge aus nur 21 Elementen konstruiert, so wäre die Gleichwahrscheinlichkeitsannahme verletzt und die Laplacesche Wahrscheinlichkeitsdefinition nicht anwendbar. Der Übung halber bestimme man die Wahrscheinlichkeiten der in Beispiel 1.6 angegebenen Ereignisse A: „es wird ein Pasch geworfen", B: „die Summe der erzielten Augen ist größer oder gleich 10" und C: „die Summe der erzielten Augen ist kleiner oder gleich 3". Sie lauten:

$$P(A) = \frac{6}{36}, \quad P(B) = \frac{6}{36} \quad \text{und} \quad P(C) = \frac{3}{36}.$$

Beispiel 1.8

Aus einer Urne mit fünf Kugeln, zwei weißen und drei schwarzen, werden mit einem Griff zufällig zwei Kugeln gezogen. Es kann angenommen werden, daß dabei jede Teilmenge von zwei aus fünf Kugeln die gleiche Wahrscheinlichkeit hat, ausgewählt zu werden. Um dieses Zufallsexperiment als Laplace-Experiment betrachten zu können, hat man jede solche Teilmenge als ein mögliches Ergebnis aufzufassen. Zur Formulierung der Ergebnisse müssen auch gleichfarbige Kugeln unterscheidbar gemacht werden. Wir numerieren dazu die Kugeln in der Urne mit den Zahlen 1 bis 5, wobei die weißen Kugeln gerade und die schwarzen ungerade Nummern erhalten. Die Ergebnismenge Ω besteht dann aus den folgenden 10 gleichwahrscheinlichen Ergebnissen:

$$\begin{array}{lllll} \omega_1 = \{1,2\} & \omega_2 = \{1,3\} & \omega_3 = \{1,4\} & \omega_4 = \{1,5\} & \omega_5 = \{2,3\} \\ \omega_6 = \{2,4\} & \omega_7 = \{2,5\} & \omega_8 = \{3,4\} & \omega_9 = \{3,5\} & \omega_{10} = \{4,5\} \end{array}$$

In der Regel sind hier nur die drei Ereignisse A_k: „Unter den beiden gezogenen Kugeln sind k weiße", $k = 0, 1, 2$, von Interesse. Deren Wahrscheinlichkeiten sind:

$$P(A_0) = P(\{\omega_2, \omega_4, \omega_9\}) = \frac{3}{10}$$

$$P(A_1) = P(\{\omega_1, \omega_3, \omega_5, \omega_7, \omega_8, \omega_{10}\}) = \frac{6}{10}$$

$$P(A_2) = P(\{\omega_6\}) = \frac{1}{10}.$$

1.3.2 Die von Misessche Wahrscheinlichkeitsdefinition

Etwa hundert Jahre nach Laplace entwickelte der österreichische Mathematiker *Richard von Mises* (1883-1953) eine neue Definition der Wahrscheinlichkeit (*von Mises* [1919, 1928]), die sich auf die relative Häufigkeit in langen Versuchsreihen stützt. Bevor wir darauf eingehen, wollen wir einen Blick auf die Zeit zwischen *Laplace* und *von Mises* werfen. Was war in dieser Zeit geschehen? Die Antwort kann kurz gefaßt werden: Innerhalb der Wahrscheinlichkeitsrechnung viel, in der Entwicklung ihrer Grundlagen wenig. Die Aufmerksamkeit galt der bereits von *Laplace* eingeleiteten naturwissenschaftlichen Anwendung.

In seiner Theorie der Meß- und Beobachtungsfehler begründete *Carl Friedrich Gauß* (1777-1855) die Bedeutung der Normalverteilung (Gauß-Verteilung), verbesserte den von *de Moivre* gefundenen Grenzübergang und entwickelte die Methode der kleinsten Quadrate, die er als erster in einen Zusammenhang mit der Wahrscheinlichkeitsrechnung brachte. *Siméon Denis Poisson* (1781-1840) verallgemeinerte das Theorem von Bernoulli, schuf dafür den Namen „loi des grands nombres" („Gesetz der großen Zahlen") und entwickelte die nach ihm benannte Poisson-Verteilung (*Poisson* [1837]). Auf *Auguste Bravais* (1811-1863) gehen die Anfänge der Korrelationsrechnung zurück. Darüber hinaus wurde die Entwicklung der Statistik vor allem in England und Rußland vorangetrieben.

Die englische Schule stand unter dem Einfluß der revolutionären Entwicklung der Biologie durch *Charles Robert Darwin* (1809-1882). Sie konzentrierte sich auf Anwendungen der Wahrscheinlichkeitsrechnung und vor allem Statistik im biologischen Bereich. Als Vater dieser Schule kann *Sir Francis Galton* (1822-1911) gelten, ein Anhänger und Vetter *Darwins*. Er entwickelte die Regressionsrechnung und führte die Methode der kleinsten Quadrate zur Berechnung der Regressionskurve ein. Auf ihn geht übrigens der Name Normalverteilung zurück. Sein engster Mitarbeiter und Nachfolger als Haupt der Schule war *Karl Pearson* (1857-1936), der Begründer der modernen Biometrie. Er entdeckte die χ^2-Verteilung wieder, die vor ihm schon bei *Friedrich Robert Helmert* (1843-1917) aufgetaucht war, entwickelte den χ^2-Anpassungstest, die Gamma- und Betaverteilung und führte die Korrelations- und Regressionsrechnung fort. Weiterhin sollen genannt werden: *George Udny Yule* (1871-1951) mit seinen Arbeiten über Vierfeldertafeln zu Beginn dieses Jahrhunderts und *William Sealy Gosset* (1876-1937), ein Mathematiker und Bierbrauer, der 1908 seine Arbeiten über die t-Verteilung unter dem Pseudonym *Student* veröffentlichte. Den Höhepunkt der angelsächsischen Schule bilden der bereits erwähnte *R. A. Fisher* sowie *Egon Sharpe Pearson* (1895-1980) und *Jerzey Neyman* (1894-1981). Neben zahlreichen Verfahren auf dem Gebiet der Schätz- und Testtheorie verdanken wir Fisher vor allem die statistische Theorie der Versuchsplanung und grundlegende Konzepte zur Begründung statistischer Schlüsse, darunter insbesondere das Likelihood-Konzept (vgl. Abschnitt 7.3). *E. S. Pearson,* ein Sohn Karl Pearsons, und *J. Neyman* haben die heutige Testtheorie (vgl. Kapitel 9) und das Konzept der Konfidenzschätzung (vgl. Abschnitt 8.2) entwickelt. *Fisher, Neyman* und *Pearson* sowie der Begründer der statistischen Entscheidungstheorie, der Österreicher *Abraham Wald* (1902-1950) haben – jeweils verschiedene – Grundpfeiler der modernen (induktiven) Statistik geschaffen und dürfen als ihre geistigen Väter angesehen werden.

Die russische Schule war der reinen Mathematik wesentlich enger verbunden als die englische. Sie behandelte die Wahrscheinlichkeitsrechnung als ein Teilgebiet der Mathematik: Ihr Gründer ist der Mathematiker *Pafnuti Lwowitsch Tschebyscheff* (1821-1894), der vor allem durch die nach ihm benannte Tschebyscheff-Ungleichung (vgl. Abschnitt 2.6.7) und einer Verallgemeinerung des Poissonschen Gesetzes der großen Zahlen bekannt ist. Seine bedeutendsten Schüler waren *Andrej Andrejewitsch Markoff* (1856-1922), auf den die Theorie der Markoff-Ketten, einem Teilgebiet der stochastischen Prozesse, zurückgeht, und *Alexander Michailowitsch Ljapunoff* (1857-1918), der als erster den zentralen Grenzwertsatz (vgl. Abschnitt 4.2) in seiner heutigen Form bewies. Unter den jüngeren Vertretern sind *Alexander Jakowlewitsch Chintschin* (1894-1959) und der bereits genannte *A. N. Kolmogoroff* hervorzuheben.

Im Gegensatz zur Laplaceschen ist die **von Misessche Wahrscheinlichkeitsdefinition** *empirisch* orientiert. Sie lautet:

> Die Wahrscheinlichkeit P (A) eines Ereignisses A ist gleich dem Grenzwert der relativen Häufigkeiten des Auftretens von A in einer unendlichen Folge unabhängiger Wiederholungen des betreffenden Zufallsexperimentes.

Dazu ergeben sich sofort zwei Fragen: Existiert der genannte Grenzwert überhaupt, d.h. konvergiert die Folge der relativen Häufigkeiten gegen eine feste Zahl? Und: Was soll man unter einer Folge unabhängiger Wiederholungen verstehen? Beide Fragen versuchte *von Mises* durch den Begriff des Kollektivs zu beantworten.

Zur Erläuterung dieses Begriffes schreiben wir in einer Folge von Wiederholungen jedesmal eine 1 oder 0 auf je nachdem, ob bei der betreffenden Wiederholung A eintritt oder nicht. Auf diese Weise läßt sich jede Folge von Wiederholungen durch eine Folge von Nullen und Einsen darstellen. Eine solche Folge heißt nach *von Mises* ein **Kollektiv**, wenn sie zwei Forderungen erfüllt: Erstens müssen die relativen Häufigkeiten der in ihr auftretenden Einsen gegen eine feste Zahl, sagen wir p, konvergieren, und zweitens muß diese Konvergenzaussage (mit demselben Grenzwert p) gültig bleiben bezüglich jeder Teilfolge, die man aus der ursprünglichen Null-Eins-Folge auswählen kann, sofern die Auswahl unabhängig vom Ausgang der jeweiligen Wiederholung, also nur im Hinblick auf die Stelle der Wiederholung erfolgt (Stellenauswahl). Die zweite Forderung nennt *von Mises* das **Regellosigkeitsaxiom**. Es besagt, daß eine Null-Eins-Folge keinem Bildungsgesetz gehorchen darf, wenn sie ein Kollektiv sein soll. Mit anderen Worten: Ein Kollektiv ist eine *zufällige Folge*.

Der v. Misessche Aufbau der Wahrscheinlichkeitsrechnung stellt den interessanten Versuch dar, den *Zufallsbegriff* als Ausgangspunkt zu setzen und daraus den Wahrscheinlichkeitsbegriff abzuleiten.

Beispielsweise wird man die Null-Eins-Folge

$$1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad . \; . \; . \; ,$$

in der sich die Eins und Null stets genau abwechseln (Bildungsgesetz), nicht als zufällige Folge ansehen. Sie bildet auch kein Kollektiv im v. Misesschen Sinne: Die Folge der relativen Häufigkeiten der Eins, nämlich

$$\frac{1}{1} \quad \frac{1}{2} \quad \frac{2}{3} \quad \frac{2}{4} \quad \frac{3}{5} \quad \frac{3}{6} \quad \frac{4}{7} \quad \frac{4}{8} \quad \frac{5}{9} \quad \frac{5}{10} \cdots$$

konvergiert zwar gegen eine feste Zahl, nämlich gegen $p = 1/2$; wählt man aber aus der Null-Eins-Folge jedes Glied, das an ungerader Stelle steht, aus, so erhält man eine Folge, die aus lauter Einsen besteht, für welche die relativen Häufigkeiten der Eins mithin sämtlich gleich 1 sind, also gegen 1 (und nicht gegen 1/2) konvergieren. Das Regellosigkeitsaxiom ist nicht erfüllt.

Gegen die v. Misessche Definition der Wahrscheinlichkeit auf der Grundlage des Kollektivbegriffes werden vor allem die folgenden Einwände erhoben:

1) Der Kollektivbegriff ist in sich widersprüchlich: Die Eigenschaft der Regellosigkeit ist mit der Konvergenz der relativen Häufigkeiten unvereinbar.
2) In der Realität können immer nur endlich viele Wiederholungen durchgeführt, also nur ein erster Abschnitt von n Gliedern der Null-Eins-Folge beobachtet werden. Wie kann man damit überprüfen, ob dieser *endliche* Abschnitt der Anfang einer *unendlichen* Folge ist, welche die beiden Kollektiveigenschaften besitzt?
3) Selbst wenn man voraussetzt, daß die ersten n Glieder den Anfangsabschnitt eines Kollektivs bilden, bleibt die Frage offen, wie groß n mindestens sein muß, um P (A) mit einer vorgegebenen Genauigkeit aus den relativen Häufigkeiten zu bestimmen.

Sofern man sich auf die v. Misesschen Arbeiten selbst beruft, sind diese Einwände gegen die Definition der Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten auch zutreffend. In der nachfolgenden Entwicklung auf diesem Gebiet, die sich vor allem mit der Frage nach einer tragfähigen Definition einer Zufallsfolge befaßte, konnten diese Einwände jedoch entkräftet werden. Die wesentlichen Stationen auf diesem Weg sind die Aufsätze von *A. H. Copeland* [1932], *A. Wald* [1937], *J. Ville* [1939], *A. Church* [1940], *A. N. Kolmogoroff* [1965], *P. Martin-Löf* [1966, 1970] und *C. P. Schnorr* [1969, 1971], dessen Monographie „Zufälligkeit und Wahrscheinlichkeit" einen gewissen Abschluß dieser Entwicklung darstellt. Diese Arbeiten sind jedoch nur einem mathematisch fortgeschrittenen Leser verständlich.

Gegenüber der auf den Kolmogoroffschen Axiomen aufbauenden Wahrscheinlichkeitsrechnung hat die auf *von Mises* zurückgehende Theorie der Wahrscheinlichkeit den Vorteil, auch für den *Begriff des Zufalls* eine Erklärung anzubieten. Warum hat sie sich dennoch nicht durchsetzen können? Wohl aus zwei Gründen: Erstens wegen des hohen Grades an Komplexität, der bereits in ihren Grundbegriffen vorhanden ist, und zweitens, weil sie keine reichere, sondern wohl eher eine ärmere Theorie als die Kolmogoroffsche darstellt, wenn man als Kriterium die Menge der innerhalb einer Theorie gültigen Aussagen (beweisbaren Sätze) heranzieht.

1.3.3 Die axiomatische Grundlage der Wahrscheinlichkeitsrechnung

Der v. Misessche Versuch einer Begründung der Wahrscheinlichkeitsrechnung hat sich als ungeeignet herausgestellt. Die Laplacesche Definition der Wahrscheinlichkeit ist aus logischen Gründen nicht haltbar: Die zu definierende Größe „Wahrscheinlichkeit" wird innerhalb der Definition in der Annahme gleich-„wahrscheinlicher" Ergebnisse bereits vorausgesetzt. Auch andere, weniger bekannte Versuche (z.B. *J. Venn* [1866] oder *J. v. Kries* [1886]) waren nicht erfolgreicher, so daß noch bis weit ins 20. Jahrhundert hinein eine tragfähige Grundlage der Wahrscheinlichkeitsrechnung fehlte. Auf diesen Umstand machte auch der bedeutende Mathematiker *David Hilbert* (1862-1943) in einem im Jahre 1900 gehaltenen Vortrag über ungelöste Probleme der Mathematik aufmerksam. Seit Hilbert hat sich die axiomatische Methode zur Begründung einer (mathematischen) Theorie durchgesetzt. Danach werden die grundlegenden Begriffe nicht mehr explizit definiert, sondern nur noch Eigenschaften von ihnen und Beziehungen zwischen ihnen durch Axiome festgelegt. Nach Vorarbeiten anderer Mathematiker*) gelang es *A. N. Kolmogoroff* im Jahre 1933, ein solches Axiomensystem für die Wahrscheinlichkeitsrechnung aufzustellen, das seiner *Einfachheit* und *Allgemeinheit* wegen heute fast ausnahmslos als Begründung dieser Theorie anerkannt wird.

Welche Eigenschaften können nach Kolmogoroff zur Charakterisierung der Wahrscheinlichkeit herangezogen werden? Zur Beantwortung dieser Frage halten wir uns vor Augen, daß nur solche Eigenschaften in Betracht kommen, die auch die Laplacesche und die v. Misessche Auffassung der Wahrscheinlichkeit aufweisen; denn die Kolmogoroffschen Axiome sollen nicht im Gegensatz zu diesen historischen Begriffen stehen; vielmehr sollen sie die in ihnen enthaltenen intuitiven Vorstellungen über die Wahrscheinlichkeit in eine logisch einwandfreie Definition einbetten. An den oben angegebenen „Definitionen" nach *Laplace* und *von Mises* erkennt man nun, daß in beiden Fällen die Größe P(A), die als Wahrscheinlichkeit eines Ereignisses A erklärt wurde, die folgenden Eigenschaften besitzt:

Erstens ist $P(A)$ eine nichtnegative reelle Zahl. Zweitens ist für das sichere Ereignis $P(\Omega) = 1$. Und drittens gilt für zwei disjunkte Ereignisse A und B stets: $P(A \cup B) = P(A) + P(B)$.

Die ersten beiden Eigenschaften sind unmittelbar klar. Die dritte Eigenschaft ergibt sich bei *Laplace,* weil für zwei disjunkte Ereignisse A und B die Anzahl der Elemente von $A \cup B$ gleich der Summe aus der Anzahl der Elemente von A und derjenigen von B ist. Bei *von Mises* erhält man sie folgendermaßen: Bezeichnet man der Reihe nach mit m, m_1 und m_2 die Anzahl der Versuche, in denen $A \cup B$ bzw. A bzw. B unter n Versuchswiederholungen auftreten, so gilt, da wegen der Disjunktheit A und B immer nur getrennt auftreten können: $m = m_1 + m_2$, also auch für die relativen Häufigkeiten $\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n}$.

Diese drei Eigenschaften bilden den wesentlichen Kern der Kolmogoroffschen Axiome. Zu ihrer Formulierung rufen wir uns noch einmal den Ausgangspunkt, insbesondere die Betrachtung über Ereignisalgebren gegen Ende von Abschnitt 1.2, in Erinnerung:

In einem Zufallsexperiment mit der Ergebnismenge Ω werden Ereignisse (Teilmengen von Ω) betrachtet, denen eine Wahrscheinlichkeit zugeordnet werden soll. Die Menge aller dieser Ereignisse soll eine Ereignisalgebra bilden, eine Forderung, die man bei *Kolmogoroff* in Form eines eigenen Axioms findet. Wir wollen sie, wie oben dargelegt, stets stillschweigend als erfüllt voraussetzen und daher nicht als eigenes Axiom erwähnen. (Strenggenommen beziehen sich die folgenden Aussagen auf Ereignisse einer zugrunde gelegten Ereignisalgebra; mit „für alle Ereignisse" o.ä. ist stets gemeint „für alle Ereignisse, die zu der Ereignisalgebra gehören".)

*) Broggi [1907], ein Schüler Hilberts, und Reichenbach [1932]

Die **Axiome der Wahrscheinlichkeitsrechnung** lauten:

Axiom 1:	Jedem Ereignis A ist eine nichtnegative reelle Zahl P (A), genannt die **Wahrscheinlichkeit von A**, zugeordnet.
Axiom 2:	*(Normierung)* $P(\Omega) = 1$.
Axiom 3a:	*(Additivität)* Für zwei disjunkte Ereignisse A und B gilt stets: $P(A \cup B) = P(A) + P(B)$.
Axiom 3b:	*(Vollständige Additivität)* Für abzählbar unendlich viele, paarweise disjunkte Ereignisse $A_1, A_2, A_3, \ldots$ gilt stets $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$

Die Axiome 1, 2 und 3a entsprechen den oben erwähnten drei Eigenschaften. Für Zufallsexperimente mit endlichem Ω kommt man damit auch aus. Das gegenüber Axiom 3a stärkere Axiom 3b (aus der vollständigen Additivität folgt die bloße Additivität, Axiom 3a hätte daher auch weggelassen werden können) ist nur für unendliche Ergebnismengen Ω von Bedeutung.

Diese Axiome legen die Wahrscheinlichkeit nur in formaler Hinsicht fest. Sie enthalten keine Hinweise auf die Beantwortung der Fragen, welche inhaltliche Bedeutung der Wahrscheinlichkeit zukommt und wie man die Wahrscheinlichkeit eines Ereignisses bestimmen kann. Dies ist aber nur scheinbar ein Mangel, denn die Offenheit der Axiome gegenüber verschiedenen Interpretationen und Bestimmungsmethoden hat den Grad ihrer allgemeinen Anerkennung nur erhöht.

Eine Rechtfertigung erfahren diese Axiome zum einen durch die Tatsache, daß sie mit den vorhandenen intuitiven Auffassungen der Wahrscheinlichkeit nach Laplace und v. Mises vereinbar sind, zum anderen durch die Vielfalt der Resultate, die man aus ihnen in Form beweisbarer Sätze gewinnen kann. Darunter spielt das *Gesetz der großen Zahlen* (vgl. Abschnitt 4.1) eine besondere Rolle: Es stellt den v. Misesschen Zusammenhang zwischen der Wahrscheinlichkeit und den relativen Häufigkeiten in Form eines beweisbaren Satzes dar.

1.3.4 Objektive, subjektive und logische Wahrscheinlichkeit

Auf die Frage, wie man den Begriff „Wahrscheinlichkeit" verstehen und interpretieren soll, gibt es keine eindeutige Antwort. Man unterscheidet heute drei verschiedene Auffassungen der Wahrscheinlichkeit. Alle drei haben die geschichtliche Entwicklung der Wahrscheinlichkeitstheorie von Anfang an begleitet. Diese drei Wahrscheinlichkeits*begriffe* sollen hier kurz (und ohne Anspruch auf wissenschaftstheoretische Exaktheit) beschrieben werden.

a) Objektive Wahrscheinlichkeit

Eine weitverbreitete Betrachtungsweise ist die Auffassung der Wahrscheinlichkeit als eine objektive Größe: Einem zufälligen Ereignis A ist „seine" Wahrscheinlichkeit P (A) als *objektives Attribut* zugeordnet, man sagt auch, A besitzt die Wahrscheinlichkeit P(A). Innerhalb eines Zufallsexperimentes ist P(A) ein „tatsächlicher", vom jeweiligen Betrachter unabhängiger Wert. Charakteristisch für diesen objektivistischen Wahrscheinlichkeitsbegriff ist die *Häufigkeitsinter-*

pretation der Wahrscheinlichkeit: Bezeichnet man mit $r_n(A)$ die relative Häufigkeit, mit der A unter n unabhängigen Durchführungen des Zufallsexperimentes eintritt, so ist P(A) derjenige Wert, in dessen Nähe sich die Folge der $r_n(A)$ mit wachsendem n „stabilisiert". Man spricht deswegen auch von einer *frequentistischen* Auffassung der Wahrscheinlichkeit. Diese Interpretation, die sich durch das Gesetz der großen Zahlen rechtfertigen läßt, stellt auch eine Meßvorschrift dar, nach der eine unbekannte Wahrscheinlichkeit P(A) bestimmt wird. Die Anhänger dieser Wahrscheinlichkeitsauffassung, man nennt sie kurz Objektivisten, dürfen daher strenggenommen nur dann von der Wahrscheinlichkeit eines Ereignisses reden, wenn dieses zu einem beliebig oft wiederholbaren Zufallsexperiment gehört. (Die Wahrscheinlichkeit dafür, daß es morgen regnet, ist für einen Objektivisten eine sinnlose Größe.) Die entschiedensten Vertreter dieses Wahrscheinlichkeitsbegriffes, der wohl zum erstenmal bei *Abraham de Moivre* auftaucht, sind *Richard v. Mises* und *Abraham Wald* sowie *Egon S. Pearson* und *Jerzey Neyman*. Da sich aus den Axiomen Kolmogoroffs das Gesetz der großen Zahlen herleiten läßt, können diese als mathematische Formalisierung der objektiven Wahrscheinlichkeit dienen. Man sagt dazu kurz: Die objektive Wahrscheinlichkeit erfüllt die Kolmogoroffschen Axiome. Damit ist für diese Auffassung der Wahrscheinlichkeit eine tragfähige mathematische Theorie vorhanden.

b) Subjektive Wahrscheinlichkeit

Kaum weniger verbreitet ist die Auffassung der Wahrscheinlichkeit als eine subjektive Größe: Einem Ereignis A wird vom jeweiligen Betrachter eine Wahrscheinlichkeit P(A) zugeordnet, die ausdrücken soll, welche Eintrittschance der Betrachter dem Ereignis beimißt. P(A) ist ein vom Betrachter, insbesondere von dessen Kenntnisstand, abhängiges und damit *subjektives Attribut* von A. Verschiedene Personen können demselben Ereignis unterschiedliche Wahrscheinlichkeiten zuweisen. Charakteristisch für diesen subjektivistischen Wahrscheinlichkeitsbegriff ist die folgende *Interpretation der Wahrscheinlichkeit als größter Wettquotient.*

Der Person, die dem Ereignis A eine Wahrscheinlichkeit P(A) zuordnen soll, wird eine Wette auf das Eintreten von A angeboten, mit der Auszahlung a (einem Geldbetrag, den sie erhält, wenn A eintritt) und dem Einsatz e (den sie auf das Eingehen der Wette zahlen muß). Als Entscheidungskriterium wird die *Gewinnerwartung* zugrunde gelegt: Ist der erwartete Gewinn der Wette größer als Null, so geht die Person die Wette ein, ist er kleiner als Null, lehnt sie die Wette ab, und ist er gleich Null, so ist sie der Wette gegenüber neutral. Mit der Wahrscheinlichkeit P(A) gewinnt die Person den Betrag a − e, mit der Wahrscheinlichkeit $P(\bar{A}) = 1 - P(A)$ *) verliert sie den Betrag e. Der erwartete Gewinn (Verlust = negativer Gewinn) ist also gleich**)

$$(a - e)\,P(A) + (-e)\,P(\bar{A}) = aP(A) - e.$$

*) Vgl. Regel (1.3) in Abschnitt 1.4.
**) Näheres über den Erwartungswert findet man in Abschnitt 2.6.4.

Er ist genau dann nichtnegativ, wenn gilt $P(A) \geqq \frac{e}{a}$. Daher geht die Person genau dann auf die Wette ein, wenn der sogenannte *Wettquotient* $\frac{e}{a}$ kleiner als $P(A)$ ist. Wird nun die Wette (bei $e = 0$ beginnend) mit wachsenden Einsätzen und konstant bleibender Auszahlung a wiederholt angeboten, so folgt: *Die subjektive Wahrscheinlichkeit* $P(A)$ *ist gleich dem größten Wettquotienten,* zu dem die Person gerade noch bereit ist, die Wette auf das Eintreten von A einzugehen (genauer: sie nicht abzulehnen).

Diese Festlegung der subjektiven Wahrscheinlichkeit bedarf einer wesentlichen Modifikation. Sie beruht auf der Erfahrungstatsache, daß eine Person ihre Entscheidung, auf eine angebotene Wette einzugehen, nicht ausschließlich vom Wert des Wettquotienten abhängig macht: Hat z.B. die Person eine subjektive Wahrscheinlichkeit von $P(A) = 0.1$ und wird die Wette auf A mit der Auszahlung $a = 1$ DM angeboten, so ist sie in der Regel bereit, einen Betrag von $e = 10$ Pf einzusetzen; wird aber die Wette auf A mit $a = 1000$ DM angeboten, so ist es fraglich, ob die Person einen Einsatz von $e = 100$ DM dagegen setzt, obwohl in beiden Fällen der Wettquotient gleich ist. Dieser Umstand wird berücksichtigt, wenn man statt der reinen Geldbeträge den *Nutzen* betrachtet, den die Person dem Gewinn eines bestimmten Betrages beimißt (bzw. den *Schaden,* den sie durch den Verlust eines Betrages erfährt). Grundlage der Betrachtung wird dann die (subjektive) *Nutzenfunktion* $u = u(x)$ der Person: $u(x)$ bezeichnet den Nutzen, der dem Besitz (Kassenbestand) von x DM beigemessen wird (wobei vorausgesetzt wird, daß der Nutzen durch eine reelle Zahl ausdrückbar ist). In der Regel kann man sich $u(x)$ als eine monoton wachsende Funktion vorstellen, deren Zuwächse (Steigungen) mit größer werdendem x kleiner werden (Gesetz vom fallenden Grenznutzen). Typisches Beispiel einer solchen Funktion ist $u(x) = \log x$. Als Entscheidungskriterium dient nun der *Erwartungswert des Nutzens*: Die Wette wird genau dann eingegangen, wenn der Erwartungswert des Nutzenzuwachses durch die Wette größer als Null ist.

Wir erkennen, daß man auf diese Weise zugleich mit der subjektiven Wahrscheinlichkeit auch den Begriff des Nutzens erklären muß: Grundlage einer subjektiven Wahrscheinlichkeitstheorie ist ein *gemeinsames Axiomensystem* für „Wahrscheinlichkeit" und „Nutzen". In einem solchen System wird das rationale Verhalten einer Person in Entscheidungssituationen beschrieben. „Subjektivisten" bringen daher die Wahrscheinlichkeitstheorie in mehr oder weniger enge Verbindung mit der Entscheidungstheorie.

Eine subjektive Wahrscheinlichkeitsauffassung findet man schon bei *Huygens* und *Laplace.* Die wichtigsten modernen Vertreter sind neben *Frank Plumpton Ramsey* (1903-1930) vor allem *Bruno de Finetti* (1906-1986) und *Leonhard Jimmie Savage* (1917-1971). Die von ihnen entwickelten Axiomensysteme, die in der Sprache hypothetischer Wetten formuliert sind, lassen den Nachweis zu, daß die subjektive Wahrscheinlichkeit die Kolmogoroffschen Axiome erfüllt. (Uneingeschränkt gilt dies nur für die Axiome 1, 2 und 3a; das Axiom 3b ist bei Subjektivisten umstritten, kann aber ohne starke Einschränkung mit hinzugenommen werden; vergleiche dazu auch *Richter* [1972].) Damit liegt auch für diese Auffassung der Wahrscheinlichkeit eine tragfähige mathematische Theorie vor, und zwar dieselbe wie für den objektivistischen Begriff. Zwischen subjektiver und objektiver Wahrscheinlichkeit besteht in mathematisch formaler Hinsicht kein Unterschied.

c) Logische Wahrscheinlichkeit

Die *logische* oder *induktive* Wahrscheinlichkeit wird vornehmlich im Bereich der Philosophie (induktiven Logik) betrachtet. Es handelt sich dabei um die am Anfang dieses Abschnitts erwähnte Wahrscheinlichkeit, die als *Bestätigungsgrad des Schlusses von einer Prämisse auf eine Conclusio* aufgefaßt wird. Die logische Wahrscheinlichkeit ordnet daher nicht einem Ereignis eine Wahrscheinlichkeit zu, wie wir es bisher gewohnt waren, sondern stellt eine *Relation zwischen zwei Aussagen* dar. Bezeichnen wir die Conclusio mit A und die Prämisse mit B, so läßt sich diese Relation in der Form P (A; B) schreiben mit der folgenden Interpretation: P (A; B) = p bedeutet „der Bestätigungsgrad durch die Prämisse B für die Conclusio A ist gleich p". Im Gegensatz zur objektivistischen und subjektivistischen Auffassung stellt die logische Wahrscheinlichkeit eine *zweistellige* Relation dar. Dabei spielt in ihrer Anwendung als Stützung induktiver Schlüsse die Conclusio A die Rolle einer *Hypothese* in Form eines allgemeinen Gesetzes, während die Prämisse B das Beobachtungsdatum in Form einzelner Fälle, kurz eine *Stichprobe* darstellt. Damit wird „P (A; B) = p" gleichbedeutend mit „der Bestätigungsgrad durch die Stichprobe B für die Hypothese A ist gleich p". Man erkennt die herausragende Bedeutung des logischen Wahrscheinlichkeitsbegriffes für die induktive Statistik.

An einer mathematischen Formalisierung dieses Begriffes, der zum erstenmal bei *Gottfried Wilhelm Leibniz* (1646-1716) auftaucht, haben *Johannes v. Kries* (1853-1928), *John Maynard Keynes* (1883-1946) – vor seiner Hinwendung zur Nationalökonomie –, *Friedrich Waismann* (1896-1959) und die bereits genannten *R. Carnap* und *K. R. Popper* gearbeitet. Gleichwohl gibt es bis heute noch keine allgemein anerkannte mathematisch operationale Theorie für die logische Wahrscheinlichkeit – ihr fehlt eine Wahrscheinlichkeits*rechnung* – so daß sie trotz ihrer Bedeutung in begrifflicher Hinsicht bisher keine Anwendung zur Begründung statistischer Schlüsse finden konnte. Wir werden daher den logischen Wahrscheinlichkeitsbegriff innerhalb dieses Lehrbuches ausklammern müssen.

1.3.5 Die Bestimmung von Wahrscheinlichkeiten

Schließlich haben wir noch auf den dritten Gesichtspunkt einzugehen, unter dem man den Wahrscheinlichkeitsbegriff betrachten muß, auf die Frage nämlich, wie man die Wahrscheinlichkeit P(A) eines Ereignisses A numerisch bestimmen kann (wenn P (A) nicht aus den Wahrscheinlichkeiten anderer Ereignisse herleitbar ist). Auf diese Frage gibt es im wesentlichen drei Antworten. Alle drei sind oben bereits angesprochen worden und sollen hier nur noch einmal zusammengestellt werden.

a) Empirische Bestimmung

Das Zufallsexperiment, bei dem A eintreten kann, wird n mal unabhängig voneinander wiederholt. Die *relative Häufigkeit* r_n (A), mit der unter den n Versuchen A eingetreten ist, dient als *Näherungswert* für P (A). Diese Bestimmungsmethode, die *von Mises* zu einer Definition von P (A) erheben wollte, liefert einen vom Betrachter unabhängigen Wert für P (A) und geht daher von P (A) als einer *objektiven Wahrscheinlichkeit* aus. Zu ihrer Anwendung muß man voraussetzen können, daß das betreffende Zufallsexperiment beliebig oft und unabhängig voneinander wiederholt werden kann. Bei dieser Methode wird von einer Beobachtung (Stichprobe) auf die unbekannte Wahrscheinlichkeit geschlossen. Es handelt sich daher um eine *empirisch induktive* Methode. Derartige Verfahren sind Gegenstand der induktiven Statistik und werden in Teil II ausführlich behandelt. Wir werden dort r_n (A) als Schätzung für P(A) betrachten und kennenlernen, wie „gut" diese Schätzung ist, das heißt wie nahe r_n (A) bei P (A) liegt. Insbesondere werden wir ermitteln, wie groß n mindestens sein muß, damit man P(A) mit einer vorgegebenen Genauigkeit bestimmen kann.

b) Subjektive Bestimmung

In vielen Fällen kann man aus ökonomischen oder ethischen Gründen von einem Zufallsexperiment nur eine relativ kleine Anzahl von Wiederholungen durchführen oder beobachten, oft gibt es – gerade im wirtschaftlichen oder sozialen Bereich – Zufallsvorgänge, die eine gewisse Einmaligkeit besitzen. Beispiele dafür sind etwa die wirtschaftliche Lage im kommenden Jahr, die Nachfrage nach einem neuen Produkt, der Konkurs einer Aktiengesellschaft, der Ausgang einer Bundestagswahl, das Ausmaß einer Katastrophe, usw. Die empirische Bestimmung von Wahrscheinlichkeiten liefert dann entweder zu ungenaue Ergebnisse oder ist völlig unmöglich. In solchen Situationen kann man die unbekannte Wahrscheinlichkeit P (A) eines Ereignisses A als eine *subjektive Wahrscheinlichkeit* auffassen und ihren Wert aufgrund sachkundiger Einschätzung durch einen oder mehrere Experten ermitteln. Zur zahlenmäßigen Festlegung einer nur mehr oder weniger unbewußt vorhandenen subjektiven Wahrscheinlichkeit kann man, wie oben ausführlich dargestellt, das Instrument *hypothetischer Wetten* benutzen. Bei der Inanspruchnahme eines Experten-Gremiums sollten die Richtlinien der sogenannten Delphi-Methode beachtet werden, die man bei *Kirsch* u.a. [1973] findet.

c) Bestimmung nach Laplace

Grundlage dieser klassischen Bestimmung von Wahrscheinlichkeiten ist die sogenannte *Gleichwahrscheinlichkeitsannahme,* die besagt, daß das betrachtete Zufallsexperiment eine endliche Ergebnismenge Ω besitzt, in der jedes Ergebnis gleichwahrscheinlich ist (Laplace-Experiment, s.o.). In einem solchen Fall muß aufgrund der Kolmogoroffschen Axiome für die Wahrscheinlichkeit P (A) eines Ereignisses $A \subset \Omega$ gelten:

$$P(A) = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega}$$

Man kann daher P (A) nach der Laplaceschen Wahrscheinlichkeits*definition* als Quotient aus „Anzahl der für A günstigen Fälle" und „Anzahl aller möglichen Fälle" *bestimmen.* Diese Bestimmung ist *nicht empirisch* orientiert, sie geschieht *vor* jeder Beobachtung von Ergebnissen des Experimentes, ist also eine *a priori* Methode. (Im Gegensatz dazu stellt die empirische Bestimmung eine *a posteriori* Methode dar.) Zur Begründung der Gleichwahrscheinlichkeitsannahme wird das schon von Laplace benutzte *Prinzip vom unzureichenden* Grund herangezogen: Ist kein zureichender Grund ersichtlich, der irgendein Ergebnis vor einem anderen bevorzugt erscheinen läßt, so betrachte man jedes Ergebnis als gleichwahrscheinlich. In diesem Prinzip, auf dessen Problematik wir nicht eingehen, können objektive und subjektive Elemente enthalten sein. *Objektive* Elemente kommen zur Geltung, wenn man die sachlichen Versuchsbedingungen als Kriterium für die Nichtbevorzugung eines jeden Ergebnisses heranzieht. Dabei spielen *Symmetrie-* und *Mischungseigenschaften* des Experimentes eine besondere Rolle. (Für einen symmetrischen Würfel ist jede Augenzahl gleichwahrscheinlich. Bei einem Kartenspiel hat jede Verteilung der Karten an die Spieler die gleiche Wahrscheinlichkeit, wenn man vorher die Karten gut mischt.) *Subjektive* Elemente kommen zum Tragen, wenn man den Wissensstand des Betrachters zum Kriterium macht: Ist dem Betrachter kein Grund für die Bevorzugung irgendeines Ergebnisses bekannt, so ist er berechtigt, jedem Ergebnis die gleiche Wahrscheinlichkeit zuzuordnen. Dementsprechend kann die Bestimmung der Wahrscheinlichkeit nach Laplace sowohl subjektivistisch als auch objektivistisch orientiert sein.

Zusammenfassend halten wir als unsere Grundlage fest:

> Wir behandeln die Wahrscheinlichkeitsrechnung auf der Grundlage der Kolmogoroffschen Axiome. Dabei lassen wir die Auffassung der Wahrscheinlichkeit als eine objektive oder subjektive Größe zu. Je nach konkretem Anwendungsfall ist die eine oder die andere Betrachtung zweckmäßiger. Die Rechenregeln und Gesetze sind für beide Standpunkte dieselben. Die numerische Bestimmung einer Wahrscheinlichkeit erfolgt – situationsbedingt – entweder a priori nach Laplace oder subjektiv oder a posteriori empirisch.

1.4 Einige Rechenregeln

Wir wollen hier einige Regeln für den Umgang mit Wahrscheinlichkeiten aus den zugrunde gelegten Axiomen herleiten. Diese Regeln werden für uns genauso selbstverständlich werden wie die Axiome selbst. Betrachtet wird ein Zufallsexperiment mit der Ergebnismenge Ω und beliebigen Ereignissen A, B usw. Die beiden ersten Regeln lauten:

(1.1) $\qquad P(A) \leqq 1$

(1.2) $\qquad P(\emptyset) = 0$

Mit ihnen lassen sich die Axiome 1 und 2 abrunden zu:

Für jedes beliebige Ereignis A gilt	$0 \leqq P(A) \leqq 1$
Für das sichere Ereignis Ω ist	$P(\Omega) = 1$
Für das unmögliche Ereignis $\emptyset$ ist	$P(\emptyset) = 0$

Beweise zu (1.1) und (1.2):

Da A und $\bar{A}$ disjunkt sind, gilt nach Axiom 3a: $\quad P(A \cup \bar{A}) = P(A) + P(\bar{A})$
Wegen $A \cup \bar{A} = \Omega$ erhält man aus Axiom 2: $\quad P(A \cup \bar{A}) = 1$
Es ergibt sich:

(*) $\qquad P(A) + P(\bar{A}) = 1$

Nach Axiom 1 ist $P(\bar{A}) \geq 0$; daher folgt aus (*): $P(A) \leq 1$. Regel (1.2) folgt aus (*), indem man $A = \Omega$, $\bar{A} = \emptyset$ einsetzt.

Unmittelbar aus (*) folgt:

(1.3) $$\boxed{P(\bar{A}) = 1 - P(A)}$$

Zur Herleitung der nächsten Regel gehen wir von der Situation $A \subset B$ aus. In diesem Fall ist $B = A \cup (\bar{A} \cap B)$, so daß nach Axiom 3a folgt: $P(B) = P(A) + P(\bar{A} \cap B)$. Daraus ergibt sich: $P(B) \geq P(A)$. Wir haben also bewiesen:

(1.4) $$\boxed{\text{Aus } A \subset B \text{ folgt: } P(A) \leqq P(B)}$$

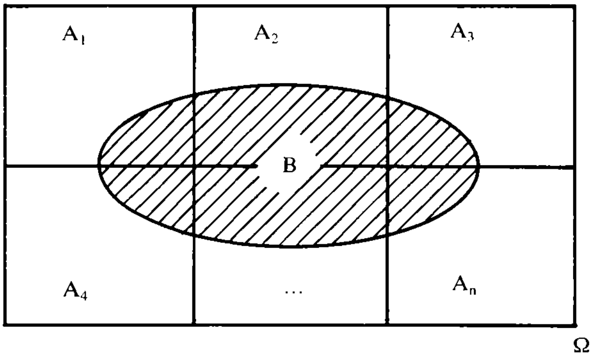

Abb. 1

Eine weitere Regel betrifft eine Zerlegung von Ω: Man sagt, die Ereignisse A_1, $A_2, \ldots, A_n$ bilden eine (endliche) **Zerlegung** von Ω, wenn sie paarweise disjunkt sind und wenn $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$ ist. Für irgendein weiteres Ereignis B gilt (vgl. Abbildung 1): $B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$. Daraus erhält man mit Hilfe von Axiom 3:

(1.5) Bilden $A_1, \ldots, A_n$ eine Zerlegung von Ω, so gilt für jedes Ereignis B:

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$

Schließlich wollen wir noch den **Additionssatz** für zwei beliebige (nicht notwendig disjunkte) Ereignisse herleiten. Er lautet:

(1.6) $$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Beweis:

Mit Hilfe von Abbildung 2 erkennt man:

$$\begin{aligned}
A \cup B &= (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B) \\
A &= (A \cap \bar{B}) \cup (A \cap B) \\
B &= \phantom{(A \cap \bar{B}) \cup {}} (A \cap B) \cup (\bar{A} \cap B)
\end{aligned}$$

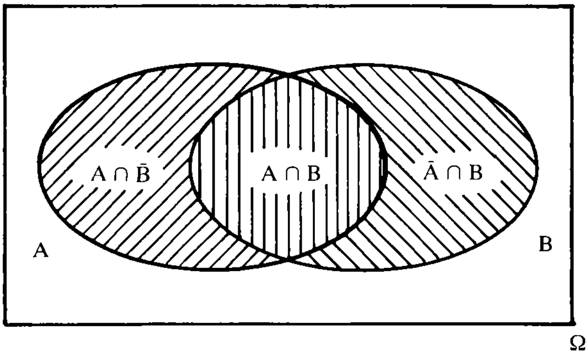

Abb. 2

Dabei werden jeweils auf den rechten Seiten Vereinigungen disjunkter Ereignisse gebildet. Nach Axiom 3 folgt:

$$\begin{aligned} P(A \cup B) &= P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) \\ P(A) &= P(A \cap \bar{B}) + P(A \cap B) \\ P(B) &= \phantom{P(A \cap \bar{B}) + {}} P(A \cap B) + P(\bar{A} \cap B) \end{aligned}$$

Ein Vergleich dieser drei Gleichungen liefert den Satz (1.6).

Beispiel 1.9

Wir betrachten das Zufallsexperiment „Zweimaliges Werfen eines Würfels", dessen Ergebnismenge Ω aus allen 36 Zahlenpaaren (i, j) besteht, die als gleichwahrscheinlich angesehen werden (Laplace-Experiment). A sei das Ereignis „Sechs im ersten Wurf" und B das Ereignis „Sechs im zweiten Wurf". Dann ist $A \cup B$ das Ereignis „mindestens eine Sechs unter den beiden Würfen". Wir wissen:

$P(A) = P(B) = \frac{1}{6}$ und $P(A \cap B) = P(\{(6,6)\}) = \frac{1}{36}$. Nach Satz (1.6) erhalten wir daraus:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$

1.5 Bedingte Wahrscheinlichkeit

Wir betrachten die bedingte Wahrscheinlichkeit zunächst im Spezialfall eines Laplace-Experimentes. Die Ergebnismenge Ω habe N Elemente. Darin seien A und B zwei Ereignisse: A besitze k Elemente, B besitze l Elemente und $A \cap B$ besitze m Elemente. Dann ist

$$P(A) = \frac{k}{N}, \quad P(B) = \frac{l}{N} \text{ und } P(A \cap B) = \frac{m}{N}$$

Wir betrachten nun die Wahrscheinlichkeit von B mit dem zusätzlichen Wissen, daß A eingetreten ist bzw. unter der Annahme, daß A eintritt. Wir nennen dies die bedingte Wahrscheinlichkeit von B unter der Bedingung A und bezeichnen sie mit $P(B|A)$. Die Kenntnis, daß A eingetreten ist, reduziert die ursprüngliche Ergebnismenge Ω auf die Teilmenge A; von den N möglichen Fällen verbleiben nur noch k; darunter gibt es m Fälle, in denen B eintritt, nämlich die Elemente von $A \cap B$. Unter der Bedingung, daß A eingetreten ist, gibt es also m für B günstige Fälle unter k (gleichwahrscheinlichen) möglichen Fällen. Daher gilt: $P(B|A) = \frac{m}{k}$. Mit der Umformung

$$\frac{m}{k} = \frac{m/N}{k/N} = \frac{P(A \cap B)}{P(A)}$$

erhalten wir in Laplace-Experimenten den Zusammenhang

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

Diese Beziehung wird benutzt, um P (B|A) für beliebige Zufallsexperimente zu *definieren*:

Mit P (B|A) bezeichnen wir die **bedingte Wahrscheinlichkeit von B unter der Bedingung A** (d.h. unter dem Wissen oder der Annahme, daß A eingetreten ist bzw. eintritt). Sie ist für $P(A) \neq 0$*) definiert als

$$\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)}} \tag{1.7}$$

Aus dieser Definition ergibt sich unmittelbar der sogenannte **Multiplikationssatz** für *zwei beliebige* Ereignisse A und B:

$$\boxed{P(A \cap B) = P(B|A)\, P(A)} \tag{1.8}$$

Beispiel 1.10

Aus einer Urne, in der sich 5 Kugeln befinden, davon 3 schwarze und 2 weiße, werden nacheinander und ohne Zurücklegen zufällig 2 Kugeln entnommen. Man kann davon ausgehen, daß bei jedem Zug jede in der Urne befindliche Kugel mit gleicher Wahrscheinlichkeit gezogen wird. Wie groß ist die Wahrscheinlichkeit, daß 2 schwarze Kugeln entnommen werden? Wir betrachten die Ereignisse A_i „beim i-ten Zug erscheint eine schwarze Kugel", $i = 1,2$. Dann ist $P(A_1) = \frac{3}{5}$. Wenn A_1 eingetreten ist, befinden sich vor der zweiten Ziehung noch 4 Kugeln in der Urne, darunter 2 schwarze und 2 weiße. Daher gilt: $P(A_2|A_1) = \frac{2}{4}$. Wir erhalten somit mit Hilfe von (1.8) die Antwort auf unsere Frage:

$P(A_1 \cap A_2) = P(A_2|A_1)\, P(A_1) = \frac{2}{4} \cdot \frac{3}{5} = \frac{3}{10}$.

Man kann den Multiplikationssatz auch auf mehr als zwei Ereignisse ausdehnen. Für drei Ereignisse A_1, A_2, A_3 erhält man durch zweimalige Anwendung von (1.8):

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3|A_1 \cap A_2)\, P(A_1 \cap A_2) = \\ &= P(A_3|A_1 \cap A_2)\, P(A_2|A_1)\, P(A_1) \quad \text{d.h.:} \end{aligned}$$

*) Der in der Definition stehende Nenner soll für jede im folgenden auftretende bedingte Wahrscheinlichkeit von Null verschieden sein.

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

Auf analoge Weise erhält man für n Ereignisse $A_1, A_2, \ldots, A_n$ den **allgemeinen Multiplikationssatz:**

(1.10) $$\boxed{P(A_1 \cap \ldots \cap A_n) = P(A_1)\,P(A_2|A_1) \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})}$$

Beispiel 1.11

Aus einer Urne mit N Kugeln, darunter M schwarze und N−M weiße, werden nacheinander und ohne Zurücklegen zufällig n Kugeln entnommen ($1 \leq n \leq M$). Wie groß ist die Wahrscheinlichkeit, dabei n schwarze Kugeln zu ziehen? Bei jedem Zug haben alle noch in der Urne befindlichen Kugeln die gleiche Wahrscheinlichkeit, gezogen zu werden. A_i sei das Ereignis „beim i-ten Zug erscheint eine schwarze Kugel", $i = 1, 2, \ldots, n$. Dann gilt:

$$P(A_1) = \frac{M}{N}, \quad P(A_2|A_1) = \frac{M-1}{N-1}, \quad P(A_3|A_1 \cap A_2) = \frac{M-2}{N-2} \cdot \quad \ldots$$

und $P(A_n|A_1 \cap \ldots \cap A_{n-1}) = \dfrac{M-(n-1)}{N-(n-1)}$, so daß gemäß (1.10) folgt:

Die Wahrscheinlichkeit dafür, daß alle n gezogenen Kugeln schwarz sind, ist

$$P(A_1 \cap \ldots \cap A_n) = \frac{M}{N} \cdot \frac{M-1}{N-1} \cdots \frac{M-(n-1)}{N-(n-1)}$$

Wendet man (1.8) auf die in der Summe von (1.5) auftretenden Wahrscheinlichkeiten $P(A_i \cap B)$ an, so erhält man den sogenannten **Satz von der totalen Wahrscheinlichkeit**:

(1.11)

Bilden die Ereignisse $A_1, \ldots, A_n$ eine Zerlegung von Ω, so gilt für irgendein weiteres Ereignis B: $$P(B) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i)$$

Beispiel 1.12

In einem Betrieb wird ein Massenartikel auf drei Maschinen M_1, M_2 und M_3 hergestellt. Die Maschinen sind an der Gesamtproduktion wie folgt beteiligt: M_1 mit 50%, M_2 mit 40% und M_3 mit 10%. Die Maschinen arbeiten nicht fehlerfrei. Der Ausschußanteil beträgt bei M_1 3%, bei M_2 6% und bei M_3 11%. Wie groß ist der Ausschußanteil der Gesamtproduktion?

(Wird ein Element aus einer Grundgesamtheit zufällig ausgewählt, so gilt: Die *Wahrscheinlichkeit,* daß dieses Element eine bestimmte Eigenschaft hat, *ist gleich dem Anteil* der Elemente mit dieser Eigenschaft in der Grundgesamtheit.)

Wir betrachten die Ereignisse:

A_i: Ein zufällig aus der Gesamtproduktion gewähltes Stück wurde von Maschine M_i hergestellt. ($i = 1, 2, 3$)

B: Ein zufällig aus der Gesamtproduktion gewähltes Stück gehört zum Ausschuß.

Aus den angegebenen Daten liest man ab:

$P(A_1) = 0.5$ und $P(B|A_1) = 0.03$
$P(A_2) = 0.4$ und $P(B|A_2) = 0.06$
$P(A_3) = 0.1$ und $P(B|A_3) = 0.11$

Da A_1, A_2, A_3 eine Zerlegung bilden, erhalten wir nach (1.11):

$$P(B) = P(B|A_1)\,P(A_1) + P(B|A_2)\,P(A_2) + P(B|A_3)\,P(A_3) = \\ = 0.03 \cdot 0.5 + 0.06 \cdot 0.4 + 0.11 \cdot 0.1 = 0.05$$

Der Ausschußanteil der Gesamtproduktion beträgt 5%.

Sachverhalt und Berechnung lassen sich gut durch ein sogenanntes *Baumdiagramm* darstellen (Abbildung 3):

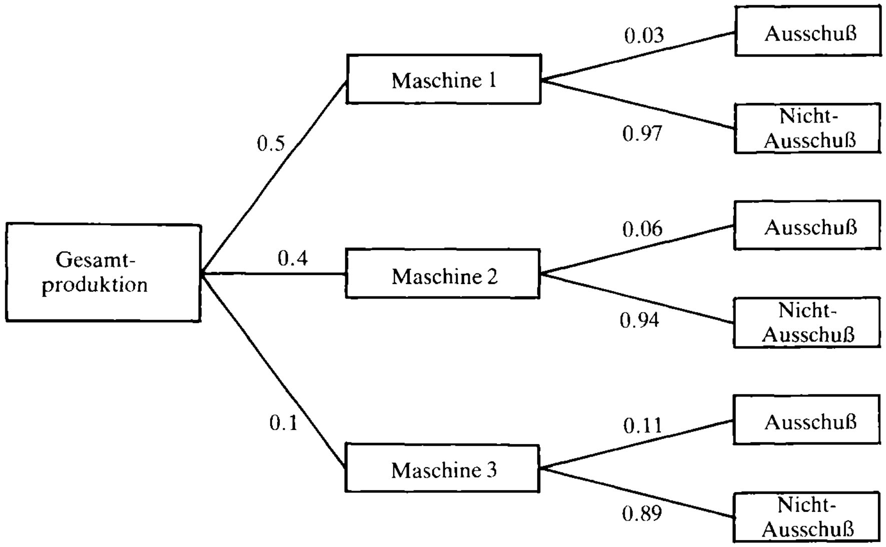

Abb. 3 Aufteilung der Gesamtproduktion auf die einzelnen Maschinen und deren Produktionen auf Ausschuß und Nicht-Ausschuß. Der Ausschußanteil der Gesamtproduktion ist $0.5 \cdot 0.03 + 0.4 \cdot 0.06 + 0.1 \cdot 0.11 = 0.05$.

1.6 Der Satz von Bayes

Thomas Bayes behandelte das Problem, welcher Zusammenhang zwischen $P(A|B)$ und der inversen Wahrscheinlichkeit $P(B|A)$ besteht. Zur Herleitung seines Satzes gehen wir auf die Definition der bedingten Wahrscheinlichkeit gemäß (1.7) zurück; nach ihr gilt:

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(A \cap B)}{P(A)} \cdot \frac{P(A)}{P(B)} = P(B|A) \cdot \frac{P(A)}{P(B)}, \quad \text{d.h.}$$

$$(1.12) \qquad \boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

Dies ist bereits der Kern der Bayesschen Regel. Ihre eigentliche Bedeutung erhält sie, wenn man (1.12) auf den Fall einer Zerlegung $A_1, \ldots, A_n$ von Ω anwendet. Setzt man für A ein A_j ein und ersetzt man $P(B)$ gemäß (1.11), so erhält man als Ergebnis den **Satz von Bayes** (**Bayessche Regel**):

(1.13)

Bilden die Ereignisse $A_1, \ldots, A_n$ eine Zerlegung von Ω und ist B irgendein weiteres Ereignis, so gilt für jedes A_j

$$P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Beispiel 1.13

Wir betrachten zu Beispiel 1.12 (mit den dort eingeführten Ereignissen A_1, A_2, A_3 und B) das folgende Problem. Ein zufällig aus der Gesamtproduktion gewähltes Stück wird überprüft und zum Ausschuß gerechnet; wie groß ist die Wahrscheinlichkeit dafür, daß ein derartiges Stück von Maschine M_j hergestellt wurde? In dieser Situation ist das Ereignis B eingetreten, gesucht ist die bedingte Wahrscheinlichkeit von A_j unter der Bedingung B, also $P(A_j|B)$. Wir wenden (1.13) an und erhalten (zu den einzelnen Zahlen vergleiche man Beispiel 1.12):

$$P(A_1|B) = \frac{P(B|A_1)\,P(A_1)}{\sum_{i=1}^{3} P(B|A_i)\,P(A_i)} = \frac{P(B|A_1)\,P(A_1)}{P(B)} = \frac{0.03 \cdot 0.5}{0.05} = 0.30$$

$$P(A_2|B) = \frac{0.06 \cdot 0.4}{0.05} = 0.48 \quad \text{und} \quad P(A_3|B) = 0.22$$

Seine charakteristische Bedeutung erfährt der Satz von Bayes durch die folgende Anwendungsweise:

1) Die Versuchsanordnung eines Zufallsexperimentes ist oft nur unvollständig bekannt; man sagt dann, das Experiment hängt noch von gewissen Umweltzuständen ab. $A_1, \ldots, A_n$ mögen die das Experiment beeinflussenden Zustände sein; dabei wird vorausgesetzt, daß die $A_1, \ldots, A_n$ sich gegenseitig ausschließen (paarweise disjunkt sind) und alle möglichen Zustände beschreiben (vollständig sind), also eine Zerlegung bilden. Man weiß zwar nicht, welcher der n Zustände vorliegt, besitzt aber Anhaltspunkte für die Chancen, mit denen dies der Fall ist; dieses *Vorwissen* möge sich in Form der Wahrscheinlichkeiten $P(A_i)$, $i = 1, \ldots, n$, angeben lassen. Man nennt die $P(A_i)$ die **a priori Wahrscheinlichkeiten**. Oft handelt es sich dabei um *subjektive* Wahrscheinlichkeiten. Ist über $A_1, \ldots, A_n$ kein weiteres Wissen vorhanden, so wendet man häufig das *Prinzip vom unzureichenden Grund* an und betrachtet jedes A_i als gleichwahrscheinlich.

2) Die Wahrscheinlichkeit dafür, daß bei Durchführung des Experimentes ein Ereignis B eintritt, läßt sich nun nicht mehr unmittelbar angeben, sondern nur als bedingte Wahrscheinlichkeit $P(B|A_i)$ unter der Bedingung, daß der Zustand A_i vorliegt, $i = 1, \ldots, n$. Man nennt die $P(B|A_i)$ auch die **Experiment-** bzw. **Modellwahrscheinlichkeiten**. Sie müssen für jedes i bestimmbar sein.

3) Nun wird der Versuch durchgeführt und das Ereignis B beobachtet. Mit Hilfe der Regel von Bayes lassen sich dann die bedingten Wahrscheinlichkeiten $P(A_i|B)$ berechnen, sie heißen die **a posteriori Wahrscheinlichkeiten** und gelten als *Verbesserung* der a priori Wahrscheinlichkeiten mittels der Beobachtung B. Der Wissensstand über die Umweltzustände wird vor der Beobachtung durch die a priori Wahrscheinlichkeiten, nach der Beobachtung durch die a posteriori Wahrscheinlichkeiten wiedergegeben: *Man lernt durch die Beobachtung gemäß der Bayesschen Regel hinzu.*

Beispiel 1.14

In einer Urne befinden sich N Kugeln, darunter M schwarze und N−M weiße. *M ist unbekannt.* (N soll bekannt sein.) Die möglichen Werte von M sind 0, 1, ..., N. Sie werden als die möglichen Zustände $A_0, A_1, \ldots, A_N$ der Urne aufgefaßt, wobei wir mit A_i den Zustand „M = i" bezeichnen. Über die Urne möge kein weiteres Wissen vorhanden sein, so daß man nach dem Prinzip vom unzureichenden Grund annehmen kann, daß jedes A_i gleichwahrscheinlich ist. Die a priori Wahrscheinlichkeiten lauten also $P(A_i) = \frac{1}{N+1}$ für jedes i.

Das Zufallsexperiment besteht in der zufälligen Wahl einer Kugel. Mit B bezeichnen wir das Ereignis, daß die gewählte Kugel schwarz ist. Dann gilt: $P(B|A_i) = \frac{i}{N}$ für jedes i.

Nun wird das Experiment durchgeführt und der Eintritt von B beobachtet. Dann erhalten wir nach (1.13) für jedes j = 0, 1, ..., N:

$$P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\sum\limits_{i=0}^{N} P(B|A_i)\,P(A_i)} = \frac{\frac{j}{N}\,\frac{1}{N+1}}{\sum\limits_{i=0}^{N} \frac{i}{N}\,\frac{1}{N+1}}$$

Im Nenner dieses Ausdrucks steht P(B). Er läßt sich berechnen als

$$P(B) = \sum_{i=0}^{N} \frac{i}{N}\,\frac{1}{N+1} = \frac{1}{N(N+1)} \sum_{i=1}^{N} i = \frac{1}{N(N+1)}\,\frac{N(N+1)}{2} = \frac{1}{2}.$$

Wir erhalten daher als a posteriori Wahrscheinlichkeiten:

$$P(A_j|B) = \frac{2j}{N(N+1)} \quad \text{für } j = 0, 1, \ldots, N$$

Nach der Beobachtung einer schwarzen Kugel ist die a posteriori Wahrscheinlichkeit für alle $j > \frac{N}{2}$ größer als die dazugehörigen a priori Wahrscheinlichkeiten, für $j < \frac{N}{2}$ sind sie kleiner – ein Ergebnis, das wir auch intuitiv erwartet hätten.

An einer Gegenüberstellung der a priori und a posteriori Wahrscheinlichkeiten in Abbildung 4 erkennt man den durch die Beobachtung B erzielten Wissenszuwachs.

Der Vollständigkeit halber sind dort auch noch die a posteriori Wahrscheinlichkeiten nach der Beobachtung $\bar{B}$ („weiße Kugel") abgebildet. Sie lauten: $P(A_j|\bar{B}) = \frac{2(N-j)}{N(N+1)}$.

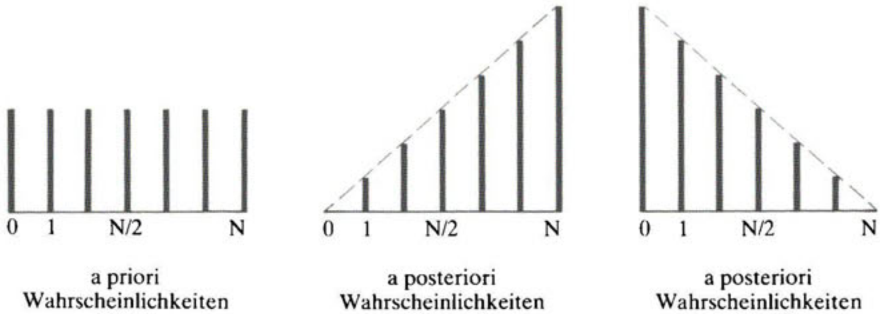

Abb. 4 A priori und a posteriori Wahrscheinlichkeiten zu Beispiel 1.14

1.7 Unabhängige Ereignisse

Wir betrachten den Fall, daß für zwei Ereignisse A und B die bedingte Wahrscheinlichkeit von B unter der Bedingung, daß A eintritt, genauso groß ist wie die Wahrscheinlichkeit von B selbst, d.h.: $P(B|A) = P(B)$. Dann beeinflußt das Eintreten von A die Chance für das Eintreten von B nicht, man nennt B **unabhängig von** A.

Umgekehrt erhalten wir, falls B unabhängig von A ist, für die bedingte Wahrscheinlichkeit $P(A|B)$ mit Hilfe von (1.12):

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

d.h.: A ist auch unabhängig von B. Daher ist der Unabhängigkeitsbegriff *symmetrisch in* A *und* B, man spricht von der **Unabhängigkeit von A und B**. Um diese Symmetrie auch formal zum Ausdruck zu bringen, beachten wir, daß nach Definition von $P(B|A)$ in (1.7) die Beziehung $P(B) = P(B|A)$ äquivalent ist mit $P(A \cap B) = P(A)\,P(B)$, so daß wir erhalten:

(1.14) A und B sind genau dann voneinander unabhängig, wenn gilt: $P(A \cap B) = P(A)\,P(B)$

Davon ausgehend läßt sich der Unabhängigkeitsbegriff auf mehr als zwei Ereignisse verallgemeinern: Die Ereignisse $A_1, \ldots A_n$ heißen **unabhängig**, wenn für jede Auswahl $A_{i_1}, A_{i_2}, \ldots A_{i_m}$ aus ihnen (mit $1 \leqq m \leqq n$ und $A_{i_j} \neq A_{i_k}$ für alle $i_j \neq \neq i_k$) gilt:

$$(1.15) \qquad \boxed{P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1})\,P(A_{i_2}) \cdots P(A_{i_m})}$$

Man nennt (1.14) und auch (1.15) die **Multiplikationsregel für unabhängige Ereignisse** und benutzt sie folgendermaßen: Wenn man aufgrund der Kenntnis der Versuchsanordnung davon ausgehen kann, daß sich zwei Ereignisse A und B aus sachlichen Gründen nicht beeinflussen können, so darf man $P(A \cap B)$ gemäß (1.14) berechnen. Entsprechendes gilt für (1.15).

Beispiel 1.15

Ein Würfel wird n mal nacheinander geworfen. Man kann davon ausgehen, daß die einzelnen Würfe sich gegenseitig nicht beeinflussen. Daher sind Ereignisse, die verschiedene Würfe betreffen, stets voneinander unabhängig. Wir betrachten das Ereignis A_i: „Beim i-ten Wurf wird eine Sechs erzielt", $i = 1, \ldots, n$.

Für jedes i ist $P(A_i) = \frac{1}{6}$. Da $A_1, A_2, \ldots, A_n$ unabhängig voneinander sind, gilt

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\,P(A_2) \cdots P(A_n) = \left(\frac{1}{6}\right)^n$$

d.h.: die Wahrscheinlichkeit, n mal hintereinander eine Sechs zu werfen, ist $\left(\frac{1}{6}\right)^n$; sie wird mit größer werdendem n verschwindend klein.

Wir wollen auch noch die folgende Frage beantworten: Wie groß muß n mindestens sein, damit die Wahrscheinlichkeit, daß sich unter den ersten n Würfen mindestens eine Sechs befindet, größer als 0.9 ist?

Sei B_n das Ereignis, daß sich unter den ersten n Würfen mindestens eine Sechs befindet. Dann gilt: $\bar{B}_n = \bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_n$ und wegen der Unabhängigkeit der $\bar{A}_1, \ldots, \bar{A}_n$ folgt:

$P(\bar{B}_n) = P(\bar{A}_1)\,P(\bar{A}_2) \cdots P(\bar{A}_n) = (\frac{5}{6})^n$

$P(B_n) = 1 - P(\bar{B}_n) = 1 - (\frac{5}{6})^n$ ist größer als 0.9, wenn $(\frac{5}{6})^n < 0.1$ ist; wegen $(\frac{5}{6})^{12} = 0.112$ und $(\frac{5}{6})^{13} = 0.093$ lautet die Antwort: n muß mindestens gleich 13 sein.

Fragen zur Wiederholung

1. Erklären oder erläutern Sie die folgenden Begriffe: Zufallsexperiment, Ereignis, sicheres Ereignis, unmögliches Ereignis, Elementarereignis.
2. Was versteht man unter einem Laplace-Experiment?
3. Wie lautet die Laplacesche Wahrscheinlichkeitsdefinition? Welchen Einwand (aus logischer Sicht) gibt es gegen sie?
4. Wie versuchte *von Mises,* die Wahrscheinlichkeit eines Ereignisses zu definieren? Welche Einwände gibt es dagegen?
5. Beschreiben Sie die axiomatische Definition der Wahrscheinlichkeit durch die drei Axiome der Wahrscheinlichkeitsrechnung.
6. Welche verschiedenen Auffassungen des Wahrscheinlichkeitsbegriffes gibt es?
7. Welche drei Möglichkeiten gibt es, die Wahrscheinlichkeit eines Ereignisses zu bestimmen?
8. Was versteht man unter der Additivität der Wahrscheinlichkeit? Wie lautet der Additionssatz für zwei (nicht notwendig disjunkte) Ereignisse?
9. Wie ist die bedingte Wahrscheinlichkeit erklärt? Wie lautet der Multiplikationssatz für zwei (nicht notwendig unabhängige) Ereignisse?
10. Wann nennt man zwei Ereignisse unabhängig? Wie lautet die Multiplikationsregel für zwei unabhängige Ereignisse?
11. Welchen Zusammenhang beschreibt der Satz von Bayes?
12. Was versteht man unter a priori bzw. a posteriori Wahrscheinlichkeiten? Welche Rolle spielt dabei der Satz von Bayes?

Aufgaben

1. A, B und C seien drei Ereignisse eines Zufallsexperimentes mit der Ergebnismenge Ω. Man gebe in mengentheoretischer Schreibweise die Ereignisse an, daß
   a) A und B eintreten, aber nicht C
   b) keines der drei Ereignisse eintritt
   c) genau eines der drei Ereignisse eintritt
   d) höchstens zwei der drei Ereignisse eintreten.
2. Eine Münze wird dreimal geworfen. Bestimmen Sie die Wahrscheinlichkeiten der Ereignisse
   A: Es tritt mindestens einmal „Wappen“ auf

B: Es tritt genau einmal „Wappen“ auf
C: Es tritt höchstens einmal „Wappen“ auf.

3. Man zeige: Sind die Ereignisse A und B unabhängig, so auch
a) A und $\bar{B}$, b) $\bar{A}$ und B, c) $\bar{A}$ und $\bar{B}$.
4. Ist es wahrscheinlicher, bei vier Würfen mit einem Würfel mindestens eine Sechs zu werfen oder bei 24 Würfen mit je zwei Würfeln mindestens eine Doppel-Sechs? (Eine Frage des *Chevalier de Méré* an *Pascal*)
5. Zwei Personen A und B gehen das folgende Spiel ein: Eine Münze wird wiederholt geworfen; wenn bei einem Wurf „Wappen“ erscheint, erhält A einen Punkt, sonst B. Wer zuerst fünf Punkte erzielt, hat gewonnen und erhält den Einsatz (den A und B je zur Hälfte eingesetzt haben). Nach sieben Würfen hat A vier Punkte und B drei. Das Spiel muß abgebrochen werden. Wie lautet die gerechte Aufteilung des Einsatzes, wenn man unter „gerecht“ eine Aufteilung im Verhältnis der Gewinnchancen versteht? (Bekannteste Frage *de Mérés* an *Pascal*, wobei *de Méré* nur an die beiden folgenden Möglichkeiten dachte: entweder eine Aufteilung im Verhältnis der gewonnenen Spiele, also 4:3, oder im umgekehrten Verhältnis zu den noch fehlenden Punkten, also 2:1.)
6. Wieviele Würfe mit je zwei Würfeln braucht man mindestens, um mit einer Wahrscheinlichkeit von mehr als 50% mindestens eine Doppel-Sechs zu erzielen? (Eine Frage, die ebenfalls dem Briefwechsel zwischen *de Méré* und *Pascal* zugeordnet wird.)
7. Einem Urlauber ist von seinem Ferienort bekannt, daß auf einen Tag ohne Regen mit Wahrscheinlichkeit 4/5 wieder ein niederschlagsfreier Tag und auf einen Tag mit Regen mit Wahrscheinlichkeit 3/5 wieder ein Tag mit Niederschlag folgt. Er ist an einem Tag ohne Regen angekommen und möchte drei Tage später eine Tour unternehmen. Wie groß ist die Wahrscheinlichkeit, daß er dazu einen Tag ohne Niederschlag erwischt?
8. Aus einer Urne mit N Kugeln, darunter M schwarze und $N-M$ weiße, werden nacheinander und ohne Zurücklegen zwei Kugeln zufällig entnommen. A_i sei das Ereignis „beim i-ten Zug erscheint eine schwarze Kugel“, $i = 1, 2$. Man zeige:
a) $P(A_1) = P(A_2)$
b) A_1 und A_2 sind abhängige Ereignisse.
9. Von drei Urnen U_1, U_2 und U_3 wird eine zufällig ausgewählt; jede Urne hat die gleiche Wahrscheinlichkeit, gewählt zu werden. Die Urnen enthalten nur schwarze und weiße Kugeln, U_1 7 schwarze und 3 weiße, U_2 5 schwarze und 5 weiße, U_3 2 schwarze und 8 weiße. Aus der gewählten Urne wird anschließend eine Kugel zufällig gezogen.
a) Wie groß ist die Wahrscheinlichkeit, dabei eine schwarze Kugel zu ziehen?
b) Es wurde eine schwarze Kugel gezogen: Wie groß ist die Wahrscheinlichkeit, daß sie aus Urne U_1 (bzw. U_2 bzw. U_3) stammt?
10. Ein Labortest zur Erkennung einer Krankheit K, an der 5% einer bestimmten Bevölkerung leiden, besitze die folgende Wirkungsweise: Hat eine Person die Krankheit K, so zeigt der Test diese mit Wahrscheinlichkeit 0.96 auch an; hat eine Person die Krankheit K nicht, so zeigt der Test K immerhin noch

mit Wahrscheinlichkeit 0.16 an. Man berechne die Wahrscheinlichkeit dafür, daß eine zufällig aus der Bevölkerung gewählte Person
a) an der Krankheit K leidet, obwohl der Test „nicht K" indizierte
b) an der Krankheit K nicht leidet, obwohl der Test K indizierte

11. Die Wahrscheinlichkeit dafür, daß in einem gewissen Werk ein Erzeugnis der Norm genügt, sei gleich 0.90. Ein Prüfverfahren ist so angelegt, daß es für ein der Norm genügendes Stück das Resultat „normgerecht" mit einer Wahrscheinlichkeit von 0.95 anzeigt. Für ein Stück, das der Norm nicht genügt, zeigt das Prüfverfahren das Resultat „normgerecht" immerhin noch mit einer Wahrscheinlichkeit von 0.10 an.
    a) Wie groß ist die Wahrscheinlichkeit dafür, daß ein unter diesem Prüfverfahren für normgerecht befundenes Stück auch tatsächlich die Norm erfüllt?
    b) Wie groß ist diese Wahrscheinlichkeit, wenn das Prüfverfahren für dasselbe Stück zweimal unabhängig voneinander das Ergebnis „normgerecht" angezeigt hat?

12. Bei einem Würfelspiel mit zwei Würfeln betrachten wir die Ereignisse
    A: erster Würfel zeigt eine gerade Zahl
    B: zweiter Würfel zeigt eine ungerade Zahl
    C: die Summe der beiden Augenzahlen ist gerade.
    Man zeige, daß je zwei der drei Ereignisse voneinander unabhängig, alle drei Ereignisse aber voneinander abhängig sind.