

Mathematik – Das Werkzeug der Erkenntnis

2. Wahrscheinlichkeit & Statistik

Ulrich Eckhardt
Universität Hamburg
Department Mathematik
— Optimierung und Approximation —
Bundesstraße 55
20 146 Hamburg
E-Mail: Eckhardt@math.uni-hamburg.de
<http://www.math.uni-hamburg.de/home/eckhardt/>

19. Januar 2010

Inhaltsverzeichnis

1	Einleitung	3
2	Worüber reden wir?	5
2.1	Ein Blick in die Geschichte	7
3	Was also ist Wahrscheinlichkeit?	8
4	Das Rechnen mit Wahrscheinlichkeiten	12
4.1	Verknüpfung von Wahrscheinlichkeiten	13
4.2	Die bedingte Wahrscheinlichkeit	15
5	Verteilungen	16
5.1	Die Normalverteilung	16
5.2	Grenzwertsätze	18
	Anhang: Beispiele und Anwendungen	21
A	Das Beispiel von Pacioli	21

Vorlesung am 13. 1. 2010 in der Bucerius Law School Hamburg.

B Noch einmal: Das Rotwild	21
C Eine aktuelle Gerichtsentscheidung	22
D Der Fall der zwei Mädchen	22
E Bedingte Wahrscheinlichkeit vor Gericht	24
F Umwelt-Östrogene und Brustkrebs	27
G Wie kann man an der Börse gewinnen?	27
H Und wie ist es mit dem Zahlenlotto?	28
I Bringen wirklich Störche die Kinder?	29
Literatur	30

Unsere Tage sind gezählt – durch die Statistiker.

Stanislaw Jerzy Lec (1909 - 1966)

1 Einleitung

Gerade die Wahrscheinlichkeitsrechnung und die Statistik haben eine besondere Stellung innerhalb der Mathematik und auch unter den Nichtmathematikern. Bekannt ist die Winston Churchill¹⁾ (fälschlich) zugeschriebene Aussage *Ich glaube nur an Statistiken, die ich selbst gefälscht habe*, oder auch das Benjamin Disraeli zugeschriebene Scherzwort: *Es gibt drei Arten von Lügen: Lügen, verdammte Lügen und Statistiken*. (There are three kinds of lies: lies, damned lies, and statistics)²⁾. Diese häufig behauptete Nähe von Statistik zur Lüge, die bewußt falsche und törichte Interpretation von Statistik scheint verbreitet zu sein. So heißt es, Franklin D. Roosevelt habe gesagt: *Ich stehe Statistiken etwas skeptisch gegenüber. Denn laut Statistik haben ein Millionär und ein armer Kerl jeder eine halbe Million*. Ähnliches habe ich selbst einmal von Franz Josef Strauß gehört. Man könnte hier einwenden, daß es beispielsweise für das Finanzamt völlig unerheblich ist, ob zwei „Halbreiche“ Steuern zahlen oder aber ein ganz Reicher und ein Armer – sieht man einmal von der Steuerprogression und Ähnlichem ab. Ich habe die Beobachtung gemacht, daß sich professionelle Lügner – hierunter verstehe ich nicht ausschließlich Politiker – sich des subtilen und immerhin einige Fachkenntnis voraussetzenden Instruments der Statistik sehr selten bedienen. Ungleich viel wirkungsvoller ist doch immer noch die einfache plumpe und unverschämte Lüge.

Was ist es, das die Statistik in den Augen der Öffentlichkeit so verdächtig macht. Woher kommt dieses Unbehagen? Man beobachtet beispielsweise, daß niemand die statischen Berechnungen eines Ingenieurs anzweifeln würde. Stürzt dann das Gebäude doch ein, vermutet man Pusch am Bau oder Fahrlässigkeit, aber niemals einen Fehler in der verwendeten Mathematik – warum eigentlich nicht? Dieses Phänomen kann man leicht verfolgen, indem man das Presseecho auf den Einsturz der Eislaufhalle in Rosenheim am 2. Januar 2006 nachvollzieht, als 15 Menschen den Tod fanden.

Das Unbehagen an der Statistik ist auch bei Wissenschaftlern zu finden. Man kennt die Aussage von Albert Einstein, der in einem Brief vom 4. 12. 1926 an Max Born schrieb [17, S. 129 f.]:

Die Quantenmechanik ist sehr achtung-gebietend. Aber eine innere Stimme sagt mir, daß das doch nicht der wahre Jakob ist. Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns kaum näher. Jedenfalls bin ich überzeugt, daß *der* nicht würfelt,

Weniger bekannt ist die – man möchte fast sagen – verzweifelte Ablehnung der statistischen Interpretation der Quantenmechanik in einem früheren Brief vom 29. 4. 1924, ebenfalls an Max Born [17, S. 118]:

Bohrs Meinung über die Strahlung interessiert mich sehr. Aber zu einem Verzicht auf die strenge Kausalität möchte ich mich nicht treiben lassen, bevor man sich nicht noch

¹⁾Vgl. hierzu den Aufsatz von Werner Barke: „Ich glaube nur der Statistik, die ich selbst gefälscht habe“ *Statistisches Monatsheft Baden-Württemberg* 11/2004:50–53.

²⁾Diese (allem Anschein nach ebenfalls falsche) Zuschreibung geht auf Mark Twain zurück (siehe etwa [32, S. 153] oder [31, S. 471].

ganz anders dagegen gewehrt hat als bisher. Der Gedanke, daß ein einem freien Strahl ausgesetztes Elektron *aus freiem Entschluß* den Augenblick und die Richtung wählt, in der es fortspringen will, ist mir unerträglich. Wenn schon, dann möchte ich lieber Schuster oder gar Angestellter in einer Spielbank sein als Physiker. Meine Versuche, den Quanten greifbare Gestalt zu geben, sind allerdings immer wieder gescheitert, aber die Hoffnung gebe ich noch lange nicht auf. Und wenns nicht gehen will, dann bleibt noch der Trost, daß der Mißerfolg nur an mir liegt.

Auf der anderen Seite werden wir mit statistischen Erhebungen und deren Resultaten überflutet. Das Statistische Bundesamt, die Statistischen Landesämter, die Europäische Statistische Behörde EUROSTAT, das UNESCO Institute for Statistics, die WHO und die zahlreichen statistischen Einrichtungen aller Nationen sammeln Daten über alle Lebensbereiche, Wirtschaftsunternehmen finanzieren aufwendige Befragungen, es gibt zahlreiche Institute für Marktforschung, das Institut für Demoskopie Allensbach sammelt Daten über die öffentliche Meinung, es gibt Wahlstatistiken und Wahlprognosen, es gibt Verkehrszählungen und Volkszählungen, es gibt ein Krebsregister und eine Verkehrssünderkartei. Supermärkte untersuchen vermittels statistischer Methoden das Käuferverhalten, Versicherungen erstellen Sterbetafeln, um die Versicherungsprämien festzusetzen Zu dieser letzteren Anwendung schreibt Karl Heinrich Marx (1818 – 1883) [22, S. 218]:

Jeder Mensch stirbt täglich um 24 Stunden ab. Man sieht aber keinem Menschen genau an, wieviele Tage er bereits verstorben ist. Dies verhindert Lebensversicherungsgesellschaften jedoch nicht, aus dem Durchschnittsleben der Menschen sehr sichere, und was noch viel mehr ist, sehr profitliche Schlüsse zu ziehn.

Gerade dies letztere ist wichtig: Für mich als Versicherungsnehmer sagt die Sterbetafel überhaupt nichts. Für die Versicherung jedoch ist die in diesen Tafeln enthaltene statistische Information sehr wichtig, eben „profitlich“.

Es gibt eine Reihe von sehr populären und zum Teil auch gut lesbaren Buchveröffentlichungen über Wahrscheinlichkeitsrechnung und Statistik. Hans-Peter Beck-Bornholdt und Hans-Hermann Dubben haben eine Anzahl von Büchern geschrieben, in denen Probleme und Irrtümer mit Statistik dargestellt werden, etwa das Buch [4]. Von Gero von Randow stammt ein hübsches und unterhaltsames Buch über ein wahrscheinlichkeitstheoretisches Problem, welches die Gemüter in den USA und hier in Europa erhitzt hat [28]. Ein Autor, der von Beruf Mathematiker ist und der auch zu dem Thema *Statistik vor Gericht* schreibt, ist John Allen Paulos [27].

Wieso also hat Statistik einen solch schlechten Ruf? Hierfür kann man eine Reihe von Gründen nennen:

- Die Wahrscheinlichkeitsrechnung und noch mehr die Statistik werden im allgemeinen zur Angewandten Mathematik gezählt, das heißt, sie leben in dem eingangs erwähnten Niemandsland zwischen Welt und Mathematik. Wenn eine statistische Aufgabenstellung vorgestellt wird, dann häufig in Form einer kleinen Geschichte, in der Würfel, Urnen mit Kugeln, Menschen und Sachen eine Rolle spielen. Alle diese Dinge sind nicht ganz kontextfrei, und es besteht die Gefahr, daß man über den Kontext auf falsche Fährten geführt wird. Hierfür werden wir noch einige hübsche Beispiele kennenlernen.
- Damit zusammenhängend: Statistische Anwendungen haben häufig eine wichtige Bedeutung im realen Leben, ob es nun Konjunkturschätzungen, Krebsregister oder die erwähnten Sterbetafeln sind. Es besteht eine gewisse Versuchung, diese Daten zu manipulieren, zu

schönen oder sogar zu verfälschen. Es soll Menschen geben, die einer solchen Versuchung erliegen – zumindest besteht bei einigen Menschen ein gewisser Anfangsverdacht, daß derlei möglich sei.

- Statistik – im angewandten Sinne – beginnt mit der Erfassung von Daten und endet mit ihrer Interpretation und Präsentation. Überwiegend sind die mit der Erfassung und Bearbeitung dieser Daten befaßten Personen keine „Berufsstatistiker“, sondern bestenfalls interessierte Laien. Beispiele für den manchmal sehr leichtfertigen Umgang mit statistischen Daten findet man beispielsweise in den erwähnten Büchern von Beck-Bornholdt und Dubben. Wahrscheinlichkeitstheorie und Statistik sind aber anspruchsvolle Gebiete der Mathematik. Es stimmt schon nachdenklich, daß man kein Fahrzeug fahren darf, ohne eine Prüfung abgelegt zu haben, daß aber medizinische Daten, Wirtschaftsdaten, Interpretationen von Arzneimitteltests und viele andere Dinge von Ungelernten und Ungelehrten ausgeführt werden, die bestenfalls einen Schnellkurs in Statistik absolviert haben.

2 Worüber reden wir?

Wir haben die Begriffe noch nicht geklärt. Was ist Wahrscheinlichkeitstheorie beziehungsweise Wahrscheinlichkeitsrechnung und was ist Statistik? In seinem Buch über das „Ziegenproblem“ schildert Gero von Randow – der kein Mathematiker ist – seine Nöte mit der Wahrscheinlichkeit [28, S. 14]:

Wer ...hoffnungsfroh im Brockhaus unter «Wahrscheinlichkeit» nachschlägt, wird ...mit philosophischen Streitfragen und der Drohung konfrontiert, daß eine mathematische Definition des Begriffes «sehr komplizierter mengentheoretischer Untersuchungen» bedürfe. Ingenieure wiederum, die traditionsgemäß in die <Hütte> gucken, das enzyklopädische Handbuch der Ingenieurwissenschaften, müssen dort erfahren: «Es gibt keine gleichzeitig anschauliche wie umfassende und exakte Definition der Wahrscheinlichkeit.»

Es sieht also aus, als würde es schwierig werden!

Die Begriffe „Wahrscheinlichkeit“ und „Zufall“ haben in der Mathematik eine andere Bedeutung als im täglichen Leben oder in Philosophie oder Theologie. Etymologisch leitet sich das Wort *Zufall* von dem lateinischen *accidēns* her, und dieses Wort ist aus der Zusammensetzung von *ad* ‚zu‘ und *cadere* ‚fallen‘ entstanden [20]. Das Wort *wahrscheinlich* leitet sich aus dem lateinischen *verisimilis* her, also etwas, das „der Wahrheit ähnlich ist“, was „den Schein der Wahrheit hat“.

Gelegentlich wird das Wort „Wahrscheinlichkeit“ verwendet, um einer Aussage den Anschein eines gewichtigen wissenschaftlichen Resultats zu verleihen. Dies scheint der Fall zu sein, wenn im Wetterbericht die Aussage „Regenwahrscheinlichkeit siebzig Prozent“ zu hören ist. Zwar hat diese Aussage eine wohlbestimmte Bedeutung, jedoch wird sie mangels Interpretation von fast allen Zuhörern falsch verstanden [29, S. 27]. Bereits im Jahre 1958 schrieb der Philosoph und Wissenschaftstheoretiker Wolfgang Stegmüller (1923 – 1991) [15, S. 345]:

Die Aussage >die Wahrscheinlichkeit, daß es morgen regnen wird, ist $\frac{2}{3}$ < ist ... eine unvollständige Aussage; es müssen die relevanten Erfahrungsdaten angeführt werden (z.B. Kenntnis der heutigen Wettersituation, meteorologisches Wissen usw.), in bezug auf welche die fragliche Hypothese beurteilt wird.

In der Alltagssprache bezeichnen wir häufig etwas als „wahrscheinlich“, wenn wir sein Eintreffen für möglich halten, aber nicht für unbedingt zwingend. So steht in einem Buch über Wissenschaftstheorie [24, S. 33]:

Man wird sich also mit der Feststellung zufriedengeben müssen, daß aller **Wahrscheinlichkeit** nach Thales als erster die Möglichkeit gesehen hat, theoretische Sätze (und das heißt für den Fall der Geometrie: Sätze über *ideale Gegenstände*) zu formulieren.

Wir haben hier also den Begriff der Wahrscheinlichkeit als den Gegensatz zu dem Begriff der Notwendigkeit, also des Wissens auf der Grundlage allgemeiner Gesetze.

Häufig wird durch das Wort „Zufall“ ein Ereignis beschrieben, welches unerwartet kommt, welches man aber bei besserer Information hätte voraussehen können. Betrachten wir als Beispiel die Geschichte eines Reisenden, der eine fremde Stadt besucht hat:

Es war höchste Zeit, wenn ich meinen Zug noch erreichen wollte. Weit und breit war kein Taxi zu sehen, aber **zufällig** kam ein Bus.

Natürlich kam der Bus nicht zufällig, sein Erscheinen zu der bestimmten Zeit am bestimmten Ort war schon seit Monaten festgelegt, der Reisende kannte nur den Fahrplan nicht. Hier steht das Wort „Zufall“ für „Nichtwissen“ aus Mangel an Information.

Wahrscheinlichkeit und Zufall können aber auch verwendet werden, um komplizierte Vorgänge für die praktische Verwendung handhabbar zu machen. Obwohl meine Lebensversicherungsgesellschaft mehr Daten über mich (und andere) akquiriert hat, als ich mir träumen lasse, wird sie sich bei der Festlegung der Versicherungsprämie doch der oben genannten Sterbetafeln bedienen, weil dies den Vorgang erheblich vereinfacht. Bei dieser Verwendung ist Wahrscheinlichkeit einfach eine Datenkompression.

In seinem Buch *Zufall und Notwendigkeit* bringt Jacques Monod zur Illustration des „wirklichen“ Zufalls eine Geschichte [25, S. 107]:

Doch in anderen Situationen nimmt der Zufallsbegriff eine nicht mehr bloß operationale, sondern eine wesensmäßige Bedeutung an. Das ist zum Beispiel der Fall, wenn man von »wesensmäßiger Koinzidenz« sprechen kann; ein solches unabhängiges Zusammentreffen resultiert aus der Überschneidung zweier voneinander völlig unabhängiger Kausalketten. Nehmen wir zum Beispiel an, Dr. Müller sei zu einem dringenden Besuch bei einem Neuerkrankten gerufen worden, während der Klempner Krause mit der dringenden Reparatur am Dach eines Nachbargebäudes beschäftigt ist. Während Dr. Müller unten am Hause vorbeigeht, läßt der Klempner durch Unachtsamkeit seinen Hammer fallen; die (deterministisch bestimmte) Bahn des Hammers kreuzt die des Arztes, der mit zertrümmerten Schädel stirbt.

In der streng deterministischen Welt von Leibniz und Laplace wäre das Zusammentreffen des Hammers mit dem Schädel des Arztes schon am ersten Schöpfungstage vorherberechenbar gewesen. Wir werden uns noch fragen müssen, ob es wirklich einen „echten“ Zufall geben kann. Noch im Jahre 1908 schrieb der Mathematiker Henri Poincaré (1854 – 1912) [23, Text 114]:

Eine sehr kleine Ursache, die für uns unbemerkt bleibt, bewirkt einen beachtlichen Effekt, den wir unbedingt bemerken müssen, und dann sagen wir, daß dieser

Effekt vom Zufall abhängen. Würden wir die Gesetze der Natur und den Zustand des Universums für einen gewissen Zeitpunkt genau kennen, so könnten wir den Zustand dieses Universums für irgendeinen späteren Zeitpunkt genau voraussagen. Aber selbst wenn die Naturgesetze für uns kein Geheimnis mehr enthielten, können wir doch den Anfangszustand immer nur *näherungsweise* kennen. Wenn wir dadurch in den Stand gesetzt werden, den späteren Zustand mit demselben *Näherungsgrade* vorauszusagen, so ist das alles, was man verlangen kann; wir sagen dann: die Erscheinung wurde vorausgesagt, sie wird durch Gesetze bestimmt. Aber so ist es nicht immer; es kann der Fall eintreten, daß kleine Unterschiede in den Anfangsbedingungen große Unterschiede in den späteren Erscheinungen bedingen; Ein kleiner Irrtum in den ersteren kann einen außerordentlich großen Irrtum für die letzteren nach sich ziehen. Die Vorhersage wird unmöglich und wir haben eine »zufällige Erscheinung«.

Diese Aussage von Poincaré ist in vieler Hinsicht sehr interessant. Insbesondere wird hier deutlich unterschieden zwischen theoretischer Vorhersagbarkeit und der praktischen Unmöglichkeit, Ausgangsdaten beliebiger Genauigkeit zu beschaffen. Dieser Gesichtspunkt wird uns noch beschäftigen. Poincaré deutet hier an, daß für ihn Zufall auf der praktischen Unmöglichkeit der Vorhersage beruht, das heißt, für ihn ist Zufall ein anderer Ausdruck für Nichtwissen oder doch nicht genaues Wissen.

Werfen wir erst einmal einen Blick in die Geschichte!

2.1 Ein Blick in die Geschichte

Aristoteles (384 v. Chr. – 322 v. Chr.) unterscheidet zwischen notwendigen Vorgängen und kontingenten oder zufälligen Vorgängen. Im *Organon* gibt er ein Beispiel [1, Kategorien, 9. Kapitel, S. 65]:

Ich meine, dass z.B. es notwendig ist, dass morgen eine Seeschlacht entweder geschehen oder nicht-geschehen wird; aber deshalb ist es nicht notwendig, dass morgen eine Seeschlacht erfolgen wird; und es ist auch nicht notwendig, dass sie nicht-erfolgen wird; nur dass sie *entweder* erfolgt *oder* nicht-erfolgt ist notwendig.

Es ist interessant, daß Aristoteles als Beispiel für einen notwendigen Vorgang eine logische Tautologie anführt, die in jedem denkbaren Universum gültig ist. Wir würden als notwendig jeden naturgesetzlich bestimmten Vorgang bezeichnen, zum Beispiel die Tatsache, daß ein Gegenstand, den man losläßt, zu Boden fällt. Die nicht notwendigen Vorgänge behandelt Aristoteles ebenso wie in dem Zitat auf Seite 6: Es handelt sich um Vorgänge, die eintreten können, die aber auch nicht eintreten können. Solche Ereignisse nennt er auch zufällig.

Nach Aristoteles sind zufällige Ereignisse nicht durch Gesetze determiniert, können also nicht Gegenstand von Wissenschaft sein. In seiner *Metaphysik* schreibt er demgemäß kurz und knapp [2, III. Teil Einteilung und Objekt der Wissenschaft, S. 90]:

Damit mag die Frage nach dem Zufall und dem Grunde seines Vorkommens erledigt und erwiesen sein, daß es vom Zufall keine wissenschaftliche Erkenntnis gibt.

Diese Einstellung blieb bis in die Neuzeit, und erst im vierzehnten Jahrhundert begann man, sich mit zufälligen Ereignissen mathematisch zu befassen.

Der Mathematiker Luca Pacioli (etwa 1445 – nach 1514) veröffentlichte im Jahre 1494 ein Lehrbuch *Summa de Arithmetica Geometria Proportioni et Proportionalita* [7, 57. Kapitel]. In diesem Buch gab Pacioli auch eine Anleitung zur doppelten Buchführung. Es ist recht interessant, Pacioli mit Kopernikus zu vergleichen, der etwa zur gleichen Zeit lebte. Wenn Pacioli heute wieder zum Leben erweckt werden könnte, dann wäre er nach einer kurzen Einweisung in der Lage, alle Details einer modernen Buchhaltung zu verstehen und nachzuvollziehen – mit Ausnahme der benutzten Computer. Kopernikus hingegen hätte nicht die mindeste Chance, ein modernes Buch über Himmelsmechanik zu verstehen. Dies Beispiel mag die ungeheure Entwicklung, die die Himmelsmechanik und auch die Mathematik seit Kopernikus’ Zeiten durchlaufen hat, verdeutlichen.

In seinem Buch geht Pacioli auch auf eine Frage der Wahrscheinlichkeitsrechnung ein. Auch hier ist der praktische Anstoß, der zu einer neuen Mathematik führte, eher banal. Pacioli untersuchte die folgende Frage [7, S. 327]:

Ein Spiel, welches auf 6 gewonnene Punkte gespielt wird, muss in einem Augenblicke abgebrochen werden, in welchem der eine Spieler auf 5, der andere auf 2 steht; wie ist der Einsatz zwischen ihnen zu theilen?

Die Lösung, die Pacioli vorschlug, nämlich den Einsatz im Verhältnis 5 zu 2 zu teilen, ist offenbar nicht richtig. Man kann dies leicht sehen, wenn man einen extremen Fall betrachtet. Wenn der eine Spieler nur einen Punkt erreicht hätte und der andere noch keinen, dann müßte nach Pacioli’s Vorschlag der erste Spieler den ganzen Einsatz erhalten.

Im Jahre 1539 erschien das Buch *Practica Arithmetica generalis* von Geronimo Cardano (1501 – 1576) [7, 65. Kapitel]. In diesem Buch wurde ein anderer Vorschlag, der ebenfalls nicht befriedigend war, zur Lösung der Aufgabe vorgestellt. Cardano kritisierte an dem Pacioli’schen Ansatz, daß die Zahl 6 bei der Berechnung nicht verwendet wird. Problemata, die nicht in die Berechnung eingehen, machen diese allemal verdächtig.

Auch der bedeutende Mathematiker, Theologe und Philosoph Blaise Pascal (1623 – 1662) beschäftigte sich etwa im Jahre 1654 auf Anregung des Spielers De Mére mit der Aufgabe von Pacioli. Pascal gab die „moderne“ Lösung an, die Wahrscheinlichkeit für den Gewinn des ersten Spielers ist $\frac{15}{16}$ oder 93.75%, und demgemäß die für den Gewinn des zweiten Spielers $\frac{1}{16}$ oder 6.25%. In diesem Verhältnis wäre demgemäß der Einsatz aufzuteilen (eine detaillierte Begründung findet man auf Seite 9 und im in Anhang A).

Den Beginn der mathematischen Wahrscheinlichkeitstheorie kann man auf das Jahr 1713 datieren, in dem das Buch *Ars Conjectandi* des Baseler Mathematikers Jakob I. Bernoulli (1655 – 1705), also acht Jahre nach seinem Tode, erschien. Bernoulli faßte die Ergebnisse früherer Autoren zusammen und entwickelte diese weiter. Wichtig ist, daß Bernoulli zeigte, wie man mit Wahrscheinlichkeiten rechnen kann. Damit wurde die moderne mathematische Wahrscheinlichkeitstheorie begründet.

3 Was also ist Wahrscheinlichkeit?

Wir haben alle in der Schule gelernt, was die Wahrscheinlichkeit eines Ereignisses ist. Zum Beispiel gibt es beim Würfeln sechs mögliche Resultate, nämlich die Augenzahlen von 1 bis 6. Die Wahrscheinlichkeit, eine bestimmte Augenzahl, etwa die 6, zu würfeln, berechnet sich aus dem Quotienten der Anzahl der „günstigen“ Fälle – dies ist hier ein Fall, nämlich der Wurf einer 6 – und der Anzahl der möglichen Fälle, hier also sechs Fälle.

Demnach kann man im Falle des Würfelspiels von Pacioui (siehe Seite 8) etwa wie folgt argumentieren:

- Die einzige Gewinnmöglichkeit für den zweiten Spieler ist, viermal in Folge eine höhere Augenzahl zu würfeln.
- In vier aufeinanderfolgenden Partien gibt es 16 Möglichkeiten. In der folgenden Tabelle bedeutet 1, daß der erste Spieler eine höhere Augenzahl würfelt, und 2 bedeutet, der zweite Spieler würfelt eine höhere Augenzahl (der Fall, daß beide die gleiche Augenzahl würfeln, wird hier nicht berücksichtigt). Dann ist die Tabelle der möglichen Wurffolgen:

1. Partie	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
2. Partie	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
3. Partie	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
4. Partie	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2

- Der zweite Spieler hat unter 16 möglichen Ausgängen des Spiels nur einen für ihn günstigen Ausgang, die Gewinnwahrscheinlichkeit für den zweiten Spieler ist demnach $\frac{1}{16}$.

Beim näheren Überlegen stößt man alsbald auf Schwierigkeiten mit dieser Erklärung. Wir müssen voraussetzen, daß die benutzten Würfel „fair“ sind. Es gibt in Zauberläden Trickwürfel, die bevorzugt eine bestimmte Augenzahl anzeigen. Solch ein Würfel wäre nicht geeignet zur Definition der Wahrscheinlichkeit. Man könnte jetzt physikalische Bedingungen angeben, unter denen ein Würfel „fair“ ist (etwa homogene Massenverteilung), und man könnte nach Art des *Urmeters*, welches in Paris aufbewahrt wird, einen *Urwürfel* in einem Museum aufbewahren. Einen solchen Weg über die „Materie“ möchte man in der Mathematik nicht einschlagen, wir brauchen eine abstrakte „Platonische“ Definition. Diese könnte etwa lauten, daß ein „fairer“ Würfel einer ist, bei dem alle Augenzahlen gleich wahrscheinlich auftreten. Dies wäre aber eine zirkuläre Definition: Zur Begriffsbestimmung der Wahrscheinlichkeit setzen wir voraus, daß der Würfel „fair“ ist, und dieser Begriff wird unter Verwendung des Wortes „gleichwahrscheinlich“ eingeführt.

Man könnte natürlich einfach sagen, daß ein Würfel „fair“ ist, wenn bei einer „großen“ Anzahl von Würfeln die geworfenen Augenzahlen etwa mit einer Häufigkeit von $\frac{1}{6}$ auftreten.

In Abbildung 1 ist das Resultat eines Würfelxperiments gezeigt. Es wurden auf einem Rechner 2000 Würfe mit einem fairen Würfel simuliert. Nach jedem Wurf wurde berechnet, wie oft unter den bis zu diesem Wurf erschienenen Augenzahlen die 6 vorkam und dieser Wert wurde durch die Gesamtzahl der bis dahin getätigten Würfe dividiert. Bei einem fairen Würfel müßte sich der Quotient dem Wert $\frac{1}{6}$ annähern. Wie die Abbildung zeigt, liegt der Quotient bei 2000 Würfeln etwas zu hoch. Ist also der verwendete Würfel nicht fair? Eine genaue Analyse der Würfe ergibt, daß bei 2000 Würfeln 354-mal die 6 auftrat, das heißt, unter 2000 Würfeln trat 20-mal zu oft die 6 auf. Dies ist durchaus normal, es hat eben „zufällig“ eine „Glückssträhne“ von 6-er-Würfeln gegeben. Wenn man das Experiment neu startet, dann ergibt sich unter Umständen, also „normalerweise“, ein völlig anderes Bild, es kann ebensogut vorkommen, daß unter 2000 weiteren Würfeln die 6 seltener auftritt, als dem erwarteten Mittelwert entspricht. Wie könnte man nun testen, ob der Würfel wirklich „fair“ ist? Man könnte natürlich wesentlich öfter würfeln, das wird allerdings irgendwann langweilig, und es kann durchaus „zufällig“ geschehen, daß bei einem solchen Experiment die Häufigkeit der 6 nach unten oder oben abweicht. Die Mathematiker und Philosophen Richard von Mises (1883 – 1953), Hans Reichenbach (1891 – 1953) und Rudolf Carnap (1891 – 1970) begründeten einen Wahrscheinlichkeitsbegriff, der auf einer Grenzwertüberlegung beruht

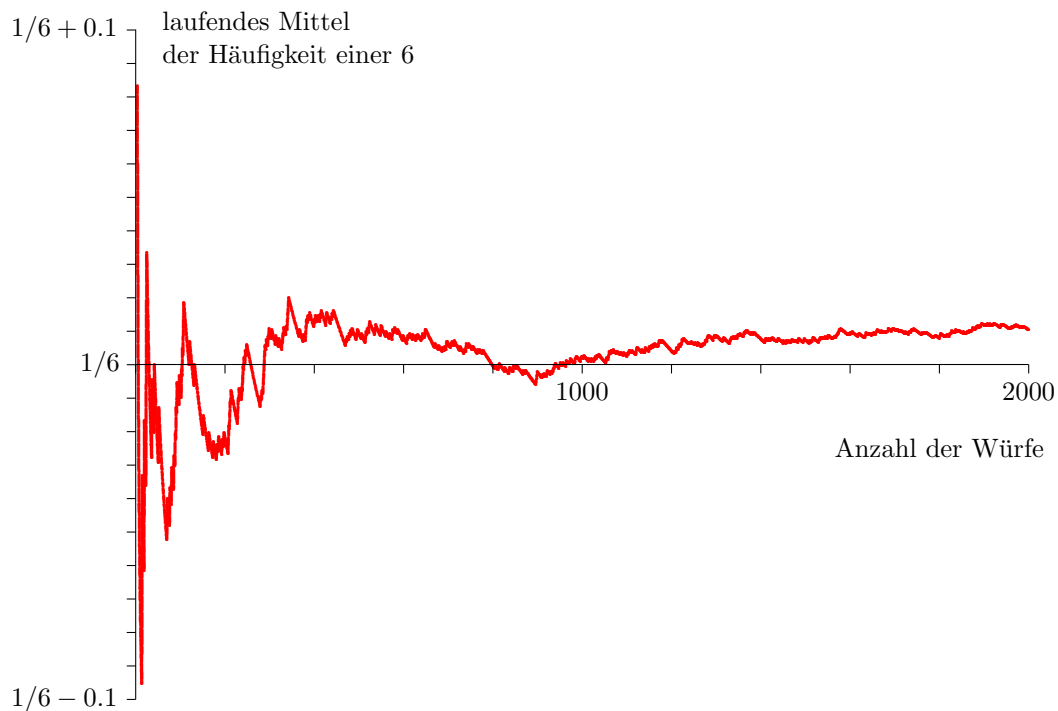


Abbildung 1: Die mittlere Anzahl der Augenzahl 6 bei einem Würfelexperiment.

(Man vergleiche hierzu die ausgezeichnete Darstellung von Carnap [8, Teil I]). Hier tritt uns das Unendliche als ein Hilfsmittel entgegen, um begrifflichen Schwierigkeiten aus dem Wege zu gehen. Eine solche Einführung der Wahrscheinlichkeit ist zwar logisch einwandfrei, jedoch praktisch nicht ausführbar.

Es gibt ein weiteres Problem bei der Definition von Wahrscheinlichkeiten. Wenn man die kombinatorische Definition, also die genannte – „Anzahl der günstigen Fälle / Anzahl der möglichen Fälle“ – verwendet, dann wird es schwierig bei Vorgängen, die unendlich viele Möglichkeiten zulassen. Bei der oben erwähnten „Regenwahrscheinlichkeit“ ist es – zumindest auf den ersten Blick – nicht möglich, „günstige“ und „mögliche“ Fälle aufzuzeigen. Eben solche Probleme treten auf, wenn man solche Begriffe wie Erkrankungswahrscheinlichkeiten, Unfallwahrscheinlichkeiten oder auch Wahrscheinlichkeiten von Gewinnen oder Verlusten an der Börse untersucht.

Bei der Berechnung einer Wahrscheinlichkeit hat man ähnliche Probleme, wie sie bei der Festlegung von Flächeninhalten oder Rauminhalten auftreten. Tatsächlich kann man beispielsweise einen Flächeninhalt durch ein Zufallsexperiment ermitteln. In Abbildung 2 ist grün schraffiert eine parabolisch berandete Fläche gegeben. Wenn man in dem umschreibenden Rechteck Punkte mit zufälligen Koordinaten ausstreut und aus zählt, welcher Anteil davon in die zu messende Fläche fällt, so erhält man aus dem Quotienten der Anzahl der Punkte in der Fläche zur Gesamtzahl eine Näherung für das Verhältnis des Flächeninhalts der Figur zu dem Flächeninhalt des umschreibenden Rechtecks. In Abbildung 3 sind die sich ergebenden Mittelwerte für bis zu 100 zufällig gewählte Punkte dargestellt. Wie man sieht, konvergieren die Näherungen (zufällig) sehr gut gegen den „richtigen“ Wert $\frac{4}{3}$ (das umschreibende Rechteck soll dabei eine Länge von 2 und eine Höhe von 1 haben). Diese Methode der Inhaltsbestimmung durch Zufallszahlen ist in der Physik sehr beliebt und unter dem Namen *Monte-Carlo-Methode* bekannt.

Eine andere Möglichkeit, die Wahrscheinlichkeit präzise zu definieren, besteht darin, daß man sich zunächst nicht darum bekümmert, *wie* die Wahrscheinlichkeit eines Ereignisses bestimmt werden sollte, sondern man überlegt, welche Eigenschaften die Wahrscheinlichkeit haben soll und wie man mit Wahrscheinlichkeiten rechnet. Dies hatte ja auch schon Bernoulli getan. Diese moderne *axiomatische* Methode wird heute gelehrt. Sie hat viel zu tun mit der Bestimmung von Flächen- oder Rauminhalten. Im 20. Jahrhundert entdeckte man, daß die Begriffe des Inhalts von Flächen, des Volumens von Körpern oder der Länge von Kurven längst nicht so einfach sind, wie man vermuten möchte. Es treten Paradoxa auf, die ganz typisch für den Umgang mit dem Unendlichen sind. Aus den Bemühungen, diese Begriffe – und natürlich auch die der Wahrscheinlichkeitstheorie sauber festzulegen, entstand die *Maßtheorie*. Die Entwicklung dieser Theorie ist verknüpft mit den Namen Émile Félix Édouard Justin Borel (1871 – 1956), Constantin Carathéodory (1873 – 1950), Henri Léon Lebesgue (1875 – 1941), Stefan Banach (1892 – 1945), Stanislaw Marcin Ulam (1909 – 1984) und anderen. Bei diesem Ansatz befindet man sich vollständig im Reich der Mathematik. Der Mathematiker Heinz Bauer (1928 – 2002), der auch in Hamburg lehrte, hat ein Buch über *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* geschrieben, welches ich selbst sehr schätze [3]. Das Buch hat 342 Seiten. Das erste „anschauliche“ Beispiel, ein Urnenexperiment, erscheint auf Seite 113. Weitere „konkrete“ Bezüge sind in diesem Buch nicht zu finden. Bei einer solchen Vorgehensweise hat man natürlich keinerlei Schwierigkeiten mit Problemen der korrekten Interpretation und Übersetzung vorgelegter praktischer Probleme, alles spielt sich in der Welt der reinen Mathematik ab.

Eine sehr interessante moderne Grundlegung der Wahrscheinlichkeitstheorie hat mit der Informationstheorie und Informatik zu tun. Sie wurde von dem bedeutenden und außerordentlich vielseitigen russischen Mathematiker Andrej Nikolajewitsch Kolmogorow (1903 – 1987) eingeführt. Einer der bedeutendsten noch lebenden Forscher auf diesem Gebiet ist Gregory Chaitin [9, 10, 11]. Die zugrundeliegende Idee ist eigentlich recht einfach. Wenn man eine Folge von Zahlen übermitteln möchte, dann kann man dies auf verschiedene Weise tun. Wenn ich jemandem die sämtlichen Nummern des Hamburger Telephonbuches senden möchte, dann besteht die einzige Methode darin, alle Nummern, so wie sie sind, zu übermitteln. Wenn ich aber jemanden sämtliche Primzahlen unter zehn Millionen übermitteln möchte, dann kann ich das recht einfach und kurz tun, indem ich ein kleines Programm schreibe, mit dessen Hilfe der Empfänger diese Primzahlen ausrechnen kann. In der algorithmischen Wahrscheinlichkeitstheorie wird nun eine Zahlenfolge zufällig genannt, wenn die sparsamste Möglichkeit, sie zu übermitteln, die Angabe der einzelnen Mitglieder der Folge ist. Sobald die Folge irgendwelche Regelmäßigkeiten aufweist, ist es zweckmäßiger, das „Bildungsgesetz“ zu übermitteln. Diese Definition ist hochinteressant, jedoch hat sie ebenfalls ihre Probleme. Die in obigen Beispielen benutzten „Zufallszahlen“ wurden mit einem sehr kurzen Standardprogramm berechnet, sie sind also nach der Kolmogorowschen Definition ganz sicher keine Zufallszahlen.

Zum Abschluß noch eine originelle Verwendung des Zufalls auf einem Gebiet, wo er eigentlich nicht hingehört. In der Zahlentheorie interessiert man sich besonders für *Primzahlen*, das sind natürliche Zahlen, die keine echten Teiler haben, zum Beispiel die Zahlen 2, 3, 5, 7, 11, 13, 17, 19, Man weiß seit Euklid, daß es unendlich viele solche Zahlen gibt (vgl. Teil 1 dieser Vorlesungsreihe), aber man hätte gerne so etwas wie „Bildungsgesetze“ dieser Zahlen. Man weiß aber nur relativ wenig über deren Vorkommen und Eigenschaften. Carl Friedrich Gauß (1777 – 1855) kam im Alter von etwa fünfzehn Jahren auf die Idee, die Primzahlen als „zufällig“ zu betrachten und nach ihrer Verteilung zu fragen. Nennt man $\pi(n)$ die Anzahl der Primzahlen bis zu der Zahl n (also etwa $\pi(20) = 8$), dann wäre der Quotient $\pi(n)/n$ die „Wahrscheinlichkeit“, unter den Zahlen bis n eine Primzahl zu finden. Gauß schrieb über diese Idee an den Astronomen Johann Franz Encke (1791 – 1865) am 24. Dezember 1849 [36, S. 212, Anmerkung 6]:

Die gütige Mitteilung über die Frequenz der Primzahlen ist mir in mehr als einer Beziehung interessant gewesen. Sie haben mir meine eigenen Beschäftigungen mit demselben Gegenstande in Erinnerung gebracht, deren erste Anfänge in eine sehr entfernte Zeit fallen, ins Jahr 1792 oder 1793, wo ich mir die Lambertschen Supplemente zu den Logarithmentafeln angeschafft hatte. Es war, noch ehe ich mit feineren Untersuchungen aus der höheren Arithmetik mich befaßt hatte, eines meiner ersten Geschäfte, meine Aufmerksamkeit auf die abnehmende Frequenz der Primzahlen zu richten, zu welchem Zweck ich dieselben in den einzelnen Chiliaden [Tausendern] abzählte und die Resultate auf einem der angehefteten weißen Blätter verzeichnete. Ich erkannte bald, daß unter allen Schwankungen diese Frequenz durchschnittlich nahe dem Logarithmus verkehrt proportional sei, so daß die Anzahl aller Primzahlen unter einer gegebenen Grenze n nahe durch das Integral $\int \frac{dn}{\log n}$ ausgedrückt werde ... [gemeint ist hier der Logarithmus zur Basis e]. In späterer Zeit, als mir die in Vegas' Tafeln (von 1796) abgedruckte Liste bis 400031 bekannt wurde, dehnte ich meine Abzählung weiter aus, was jenes Verhältnis bestätigte. Eine große Freude machte mir 1813 die Erscheinung von Chernacs cribrum [Sieb], und ich habe (da ich zu einer anhaltenden Abzählung der Reihe nach keine Geduld hatte) sehr oft einzelne unbeschäftigte Viertelstunden verwandt, um bald hier bald dort eine Chiliade abzuzählen; ich ließ jedoch zuletzt es ganz liegen, ohne mit der Million ganz fertig zu werden. Erst später benutzte ich Goldschmidts Arbeitsamkeit, teils die noch gebliebenen Lücken in der ersten Million auszufüllen, teils nach Burckhardts Tafeln die Abzählung weiter forzusetzen. So sind (nun schon seit vielen Jahren) die drei ersten Millionen abgezählt und mit dem Integralwert verglichen. Ich setze hier nur einen kleinen Extrakt her:

Unter	gibt es Primzahlen:	Integralwert:
500 000	41 556	41 606.4
1 000 000	78 501	79 627.5
1 500 000	114 112	114 263.1
2 000 000	148 883	149 054.8
2 500 000	183 016	183 245.0
3 000 000	216 745	216 970.6 ...

Wie groß das Interesse war, das noch der greise Gauß dem Verteilungsgesetz der Primzahlen entgegenbrachte, zeigt die Tatsache, daß auf seine Veranlassung hin der Hamburger Rechenkünstler Dase Primzahlentafeln von 6 000 000 bis 8 000 000 aufstellte.

4 Das Rechnen mit Wahrscheinlichkeiten

Ein wichtiger Vorteil der Wahrscheinlichkeitstheorie ist, daß man Wahrscheinlichkeiten numerisch, also als Zahlen zwischen 0 und 1, darstellen kann und daß man mit diesen Zahlen rechnen kann. Ein Ereignis, welches mit Wahrscheinlichkeit 1 auftritt, nennt man *sicheres Ereignis*, ein Ereignis mit Wahrscheinlichkeit 0 *unmögliches Ereignis*. Wir nehmen an, wir kennen die Wahrscheinlichkeit des Auftretens eines Ereignisses S . Wir bezeichnen die entsprechende Zahl mit $P(E)$ (von lateinisch probabilitas, Glaubhaftigkeit, Wahrscheinlichkeit). Ist E das Ereignis, mit einem (fairen) Spielwürfel eine 6 zu würfeln, dann ist also $P(E) = 1/6$. Mit $\neg E$ bezeichnen wir das Ereignis, daß E **nicht** eintritt. Bei dem Würfelbeispiel wäre $\neg E$ das Ereignis, daß **keine** 6 gewürfelt wird. Offenbar ist

$$P(\neg E) = 1 - P(E) \quad \text{für jedes Ereignis } E.$$

4.1 Verknüpfung von Wahrscheinlichkeiten

Wir betrachten als Beispiel das Würfeln mit einem (fairen) Spielwürfel. Die Wahrscheinlichkeit, eine 6 zu würfeln, ist $\frac{1}{6}$. Wir würfeln nun zweimal hintereinander und fragen uns, wie groß die Wahrscheinlichkeit ist, mit beiden Würfeln eine 6 zu erreichen. In formalisierter Sprechweise formulieren wir dies so:

Gesucht ist die Wahrscheinlichkeit für das Ereignis
 der erste Wurf ergibt eine 6 **und**
 der zweite Wurf ergibt eine 6.

Noch formaler: Wir bezeichnen mit A_1 das Ereignis „der erste Wurf ergibt eine 6“ und mit A_2 das Ereignis „der zweite Wurf ergibt eine 6“. Die Copula „und“ symbolisieren wir mit dem Zeichen $\wedge$. Damit ist die gesuchte Wahrscheinlichkeit ganz formal $P(A_1 \wedge A_2)$. Die beiden Wahrscheinlichkeiten $P(A_1)$ und $P(A_2)$ sind bekannt, wie wir gesehen haben, betragen sie je $\frac{1}{6}$.

Man kann die gesuchte Wahrscheinlichkeit leicht durch Auszählen bestimmen: Es gibt 36 mögliche Fälle für zwei Würfe mit einem Würfel, wie man aus Abbildung 4 erkennt. Von diesen 36 möglichen Fällen ist einer „günstig“, also ist die gesuchte Wahrscheinlichkeit gerade $\frac{1}{36}$.

Wir wollen noch einen Moment bei dieser Wahrscheinlichkeit verbleiben. Wenn wir das Experiment mit zwei Würfeln sehr oft wiederholen, dann wird etwa in $\frac{1}{6}$ aller Fälle der erste Würfel eine 6 zeigen. Wenn die Anzahl der Wiederholungen so groß ist, daß dieser Fall, daß der erste Würfel eine 6 zeigt, sehr häufig ist, dann wird unter diesen Fällen bei einem Anteil von $\frac{1}{6}$ der zweite Würfel ebenfalls eine 6 aufweisen. Wichtig ist hierbei, daß das Resultat des zweiten Würfelwurfs in keiner Weise von dem des ersten Wurfs abhängt. Wir können dann sagen, daß die gesuchte Wahrscheinlichkeit gerade das Produkt der Einzelwahrscheinlichkeiten ist:

$$P(A \wedge B) = P(A) \cdot P(B) \quad \text{bei voneinander unabhängigen Ereignissen } A \text{ und } B.$$

Im Beispielfall eines Wurfes von zwei Sechsen ist also in der Tat die gesuchte Wahrscheinlichkeit $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

Die Forderung der Unabhängigkeit ist hier sehr wichtig. Wenn ich in einem Wald wandere und ein Stück Rotwild kreuzt meinen Weg, dann ist dies ein Ereignis, welches eine recht geringe Wahrscheinlichkeit hat. Das Ereignis, daß unmittelbar darauf ein zweites Stück Rotwild meinen Weg kreuzt ist nicht das Quadrat der Wahrscheinlichkeit, daß ein Wild meinen Weg kreuzt, sondern bedeutend höher, denn bekanntlich tritt Rotwild in Rudeln auf, die beiden Ereignisse sind nicht unabhängig voneinander.

Eine weitere wichtige Verknüpfung von Wahrscheinlichkeiten ist die Verknüpfung durch „oder“. Wir betrachten als Beispiel den Wurf mit einem Würfel und fragen nach der Wahrscheinlichkeit, daß dabei eine 6 oder eine 4 erscheint. Diese beiden Ereignisse schließen einander offenbar aus,

denn man kann ja nicht gleichzeitig eine 6 und eine 4 würfeln. Die Berechnung der Wahrscheinlichkeit, daß entweder das eine oder das andere Ereignis eintritt, ist nun leicht: Die Anzahl der „günstigen“ Fälle ist einfach die Summe der Anzahlen der einzelnen Fälle des Wurfs einer 6 und des Wurfs einer 4. Wenn wir jetzt mit A das Ereignis bezeichnen, daß eine 6 geworfen wird und mit B das Ereignis, daß eine 4 geworfen wird, und wenn wir die Copula „oder“ durch $\vee$ symbolisieren, dann haben wir die Formel

$$P(A \vee B) = P(A) + P(B) \quad \text{bei einander ausschließenden Ereignissen } A \text{ und } B.$$

Im Beispielfall eines Wurfs einer 6 oder einer 4 ist die gesuchte Wahrscheinlichkeit also $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. Dasselbe Resultat erhält man natürlich auch durch Auszählen der „günstigen“ Fälle.

Die Forderung, daß beide Fälle einander ausschließen, ist hier wichtig. Um unsere Erkenntnisse zu vertiefen, betrachten wir wieder der Fall zweier Würfe mit einem Würfel und fragen nun nach der Wahrscheinlichkeit, mit mindestens einem der Würfe eine 6 zu erhalten. Wir nennen dieses Ereignis A und suchen demgemäß dessen Wahrscheinlichkeit $P(A)$. Dazu betrachten wir zunächst die Gesamtheit der Möglichkeiten (Abbildung 4). Es gibt $6 \times 6 = 36$ gleichwahrscheinliche Möglichkeiten. Wir könnten die gesuchte Wahrscheinlichkeit natürlich durch einfaches Auszählen ermitteln, aber wir wollen jetzt rechnen!

Man könnte etwas voreilig so argumentieren: Die Frage ist doch, wie groß die Wahrscheinlichkeit ist, mit dem ersten **oder** dem zweiten Würfel eine 6 zu würfeln. Wir haben soeben gesehen, daß bei der „oder“-Verknüpfung die Wahrscheinlichkeiten einfach addiert werden, also wäre die gesuchte Wahrscheinlichkeit $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. So kann man offenbar nicht rechnen, denn mit dem gleichen Argument würde sich ergeben, daß die Wahrscheinlichkeit, mit sechs Würfeln eine 6 zu würfeln, gleich 1 wäre, und jeder Mensch, der einmal *Mensch ärgere dich nicht* gespielt hat, weiß, daß eine Folge von sechs Würfeln ohne eine 6 nicht selten ist. Die beiden Ereignisse „Würfeln einer 6 beim ersten Wurf“ und „Würfeln einer 6 beim zweiten Wurf“ schließen nämlich einander nicht aus, so daß unsere Formel nicht anwendbar ist.

Wir führen die folgenden Abkürzungen ein:

A₁ der erste Würfel zeigt eine 6 (Wahrscheinlichkeit $P(A_1) = \frac{1}{6}$),

A₂ der zweite Würfel zeigt eine 6 (Wahrscheinlichkeit $P(A_2) = \frac{1}{6}$),

Zur genaueren Untersuchung unterteilen wir die Gesamtheit aller Möglichkeiten in die folgenden Teilmengen von Ereignissen:

A₁ ∧ ¬A₂ der erste Würfel zeigt eine 6, der zweite nicht,

A₂ ∧ ¬A₁ der zweite Würfel zeigt eine 6, der erste nicht,

A₁ ∧ A₂ bei beiden Würfeln erhält man eine 6,

¬A₁ ∧ ¬A₂ keiner der beiden Würfe zeigt eine 6.

Diese Unterteilung hat die Eigenschaften:

- Die einzelnen durch $\wedge$ miteinander verknüpften Ereignisse sind durchweg voneinander unabhängig (sie betreffen zwei aufeinanderfolgende Würfe, und das Ergebnis des zweiten hängt nicht vom Ergebnis des ersten ab).
- Keiner der möglichen Fälle taucht in mehr als einer der Teilmengen auf. Das bedeutet aber, daß die Ereignisse in zwei verschiedenen Teilmengen einander ausschließen.
- Jede der Möglichkeiten des zweimaligen Wurfes liegt in eine dieser Teilmengen.

Wir zerlegen nun die eingangs gestellte Aufgabe nach Maßgabe unserer Teilmengen:

Mindestens eine 6 erhält man durch

eine 6 beim ersten Wurf und keine 6 beim zweiten Wurf	oder
keine 6 beim ersten Wurf und eine 6 beim zweiten Wurf	oder
eine 6 bei beiden Würfeln.	

Bei voneinander unabhängigen Ereignissen erhält man die Wahrscheinlichkeit des mit der Copula „oder“ verknüpften Ereignisses nach der obigen Formel durch einfache Addition der Einzelwahrscheinlichkeiten. Wir können also die gesuchte Wahrscheinlichkeit in „gelehrter“ Notation wie folgt berechnen:

$$P(A) = P((A_1 \wedge \neg A_2) \vee (A_2 \wedge \neg A_1) \vee (A_1 \wedge A_2))$$

Nun ist alles ganz einfach:

$$\begin{aligned} P(A_1 \wedge \neg A_2) &= \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}, \\ P(A_2 \wedge \neg A_1) &= \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}, \\ P(A_1 \wedge A_2) &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \end{aligned}$$

Damit ist die gesuchte Wahrscheinlichkeit, daß bei zwei Würfeln mindestens eine 6 auftritt = $2 \cdot \frac{5}{36} + \frac{1}{36} = \frac{11}{36}$, also etwas geringer als die „naiv“ erhaltene Zahl von $\frac{1}{3}$.

Man erhält natürlich das gleiche Resultat durch Auszählen der möglichen und der günstigen Fälle in Abbildung 4.

4.2 Die bedingte Wahrscheinlichkeit

Ein Begriff, der zuweilen auch Experten in Verwirrung bringt, ist die *bedingte Wahrscheinlichkeit*. Dieser Begriff geht auf Thomas Bayes (1702 – 1761) zurück, der in seiner zwei Jahre nach seinem Tode von Richard Price publizierten Schrift *An Essay towards solving a Problem in the Doctrine of Chances* (1763) dieses Problem der bedingten Wahrscheinlichkeit ansprach.

Betrachten wir ein Beispiel. Ein Gastwirt (die Geschichte möge in Bayern spielen) beobachtet, daß 40% seiner Gäste Schweinebraten und 20% der Gäste Rinderbraten bestellen. Der Schluß, daß 60% der Gäste Schweine- oder Rinderbraten bestellen, ist korrekt, da beide Ereignisse – bei Gästen mit normalem Appetit – einander ausschließen. Der Wirt ermittelt weiter, daß 85% der Gäste Bier trinken. Der Schluß, daß 125% aller Gäste Schweinebraten oder Bier genießen,

ist offenbar falsch, da die beiden Ereignisse „ein Gast bestellt Schweinebraten“ und „ein Gast bestellt Bier“ einander sicher nicht ausschließen.

Der Schluß, daß in dem obigen Beispiel $0.85 \times 0.4 = 0.34$, also 34% aller Gäste Schweinebraten *und* Bier bestellen, ist auch nicht zulässig, da diese beiden Ereignisse nicht voneinander unabhängig sind.

Bei einer Durchsicht der auf dem Computer des Gastwirts gespeicherten Belege stellt dieser fest, daß 40 % aller Gäste Schweinebraten **und** Bier bestellt haben. Der Wirt fragt sich nun: Gesetzt, ein Gast bestellt Schweinebraten. Wie hoch ist die Wahrscheinlichkeit, daß er dazu Bier bestellt?

Man spricht hier von der *bedingten Wahrscheinlichkeit* daß ein Gast Bier bestellt, wenn er Schweinebraten zu verzehren gedenkt. Man bezeichnet die Wahrscheinlichkeit eines Ereignisses A unter der Voraussetzung daß das Ereignis B eingetroffen ist, durch das Symbol $P(A|B)$.

Ereignis: Ein Gast bestellt	Bezeichnung	Wahrscheinlichkeit
Rinderbraten	R	0.20
Schweinebraten	S	0.60
Bier	B	0.85
Bier und Schweinebraten	$B \wedge S$	0.40

Nun zu unserer Wahrscheinlichkeitsaufgabe. Wir dürfen die bekannte Wahrscheinlichkeit des Ereignisses $B \wedge S$ nicht mit der bedingten Wahrscheinlichkeit verwechseln. $P(B)$ ist die Wahrscheinlichkeit, daß irgendein Gast Bier trinkt, $P(A \wedge B)$ die Wahrscheinlichkeit, daß ein Gast sowohl Schweinebraten als auch Bier bestellt. Die Wahrscheinlichkeit $P(B|S)$ bedeutet: Wieviele von der Gesamtheit der Schweinebratenesser trinken Bier? Das heißt, wir müssen die Wahrscheinlichkeit $P(S \wedge B)$ bezüglich der Gesamtheit aller Personen, die das Ereignis B realisieren, berechnen.

Man kann etwa so argumentieren, daß von einer sehr großen Anzahl n von Gästen $60 \cdot n$ Schweinebraten essen und $40 \cdot n$ essen Schweinebraten **und** trinken Bier. Der Prozentsatz der Schweinebratenesser, der auch noch Bier bestellt, ist demnach $40 \cdot n / 60 \cdot n = 2/3$. Abstrakt formuliert, erhalten wir die Formel für die bedingte Wahrscheinlichkeit, die von manchen Autoren eben auch die *Bayessche Formel* genannt wird:

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

Einige Beispiele im Anhang sollen den Begriff der bedingten Wahrscheinlichkeit illustrieren (Anhänge ?? und B) und auch verdeutlichen, wie schwer sich selbst Fachleute mit der Bayesschen Formel tun (Anänge D und E). Weitere Beispiele findet man in den zitierten Büchern von Beck-Bornholdt und Dubben sowie von Gero von Randow.

5 Verteilungen

5.1 Die Normalverteilung

Die mittlere tägliche Regenmenge in Hamburg, die Börsenkurse einer Aktie, das tägliche Fahrzeugaufkommen auf einer gewissen Autobahnstrecke, alles dies sind zufällige Ereignisse, die von

der Zeit abhängen. Man spricht hier von *Zeitreihen* oder *stochastischen Prozessen*³⁾. Wir beschränken uns auf solche Prozesse, deren Ergebnis ein Zahlenwert ist. Ziel des Studiums solcher Prozesse ist dabei die Analyse der Vorgänge, etwa die Suche nach Zusammenhängen und natürlich die Prognose.

Man charakterisiert solche Prozesse durch ihre *Verteilungsfunktion*. Diese gibt an, mit welcher Wahrscheinlichkeit die möglichen Ausgänge des Prozesses auftreten. Häufig begnügt man sich mit einfachen charakteristischen Zahlenwerten, etwa dem *Mittelwert*, also das arithmetische Mittel, das ist die Summe einer (großen) Anzahl von aufeinanderfolgenden Zahlenwerten des Prozesses, dividiert durch die Anzahl. Wir hatten diesen Begriff schon bei dem Würfelexperiment kennengelernt.

Eine weitere charakteristische Größe eines solchen Prozesses ist die *Varianz*. Sie wird gebildet, indem man für eine (große) Anzahl von aufeinanderfolgenden Ereignissen die Abweichung vom Mittelwert quadriert und den Mittelwert dieser Quadrate bildet.

Als ein häufig vorkommendes Beispiel einer Wahrscheinlichkeitsverteilung betrachten wir die sogenannte *Normalverteilung* oder *Gauß-Verteilung* (Abbildung 5).

Wir betrachten folgendes Beispiel: Eine Maschine produziert einen Faden der Dicke 0.1 mm (Dies Beispiel ist nicht realistisch). Wegen unterschiedlicher Qualität des Rohmaterials und wegen statistischer Schwankungen der Maschinen- und Umweltdaten hat dieser Faden nicht genau die Normdicke von 0.1 mm sondern weicht davon ab. Diese Abweichungen kann man regelmäßig zur Kontrolle messen und auf einer Graphik die relative Häufigkeit der einzelnen Abweichungen auftragen. Wir nehmen hier einmal an, daß die Bezifferung an der horizontalen Achse die Abweichung von der mittleren Dicke in $\frac{1}{100}$ mm darstelle. Dann könnte sich eine solche Verteilung wie in Abbildung 5 ergeben. Die meisten Meßwerte liegen dicht am Mittelwert, das heißt die Abweichung Null. Größere Abweichungen werden seltener und ganz große Abweichungen treten so gut wie nie auf. Bei der blauen Kurve ($\sigma = 0.4$ beziehungsweise $\sigma^2 = 0.16$) hat man eine starke Häufung der Abweichungen um den Mittelwert, bei der roten Kurve ($\sigma = 0.8$ beziehungsweise $\sigma^2 = 0.64$) hingegen sind die Abweichungen „stärker gestreut“, das heißt man hat hier weniger Werte mit geringen Abweichungen vom Sollwert und mehr Werte mit größeren Abweichungen. Man könnte also sagen, daß eine Maschine, deren Fehlerverhalten durch die Kurve mit $\sigma = 0.4$ beschrieben wird, besser eingestellt ist, als eine Maschine mit $\sigma = 0.8$.

In der Physik benutzt man die Standardabweichung, um die Zuverlässigkeit von gemessenen Resultaten zu charakterisieren. In einem Lehrbuch der Experimentalphysik [19] findet man etwa die folgenden Angaben:

$$\text{Lichtgeschwindigkeit im Vakuum} \quad c \quad (2.997925 \pm 1) \cdot 10^8 \text{ m s}^{-1}$$

Das bedeutet, daß die Lichtgeschwindigkeit $299\,792\,500 \frac{\text{m}}{\text{s}}$ beträgt oder $299\,792.5 \frac{\text{km}}{\text{s}}$. Die Streuung dieses Meßresultats beträgt $100 \frac{\text{m}}{\text{s}}$ ($0.000\,001 \cdot 10^8$). Man kann ausrechnen, daß der wahre Wert vom Mittelwert mit einer Wahrscheinlichkeit von 63 % nicht mehr als σ abweicht [19,

³⁾**Stochastik** [griechisch], eine an der Wahrscheinlichkeit orientierte Betrachtungsweise in den Naturwissenschaften und in der Statistik; besagt, dass bei Massenerscheinungen Aussagen nicht ganz exakt gemacht werden können, sondern nur unter Berücksichtigung gewisser zufälliger Abweichungen vom empirisch ermittelten Mittelwert. (dtv-Lexikon).

Mathematische Stochastik beschäftigt sich mit der Beschreibung und Untersuchung von Zufallsexperimenten wie zum Beispiel dem Werfen von Reißzwecken, Würfeln oder Münzwurf sowie vom Zufall beeinflussten zeitlichen Entwicklungen und räumliche Strukturen. (Wikipedia)

Frage: Warum wirft man Reißzwecken?

1.1.7 Meßfehler], und das heißt hier, daß die Lichtgeschwindigkeit mit dieser Wahrscheinlichkeit zwischen $299\,792\,490 \frac{m}{s}$ und $299\,792\,510 \frac{m}{s}$ liegt.

Die Gaußsche Normalverteilung wurde von Carl Friedrich Gauß im Jahre 1795 gefunden [36, S. 23]. Auch die mittlere quadratische Abweichung wurde von Gauß eingeführt. Der kleine Planet Ceres wurde 1801 entdeckt [36, S. 41 ff.], ging aber nach kurzer Beobachtungszeit in der Dämmerung verloren. Es kam nun darauf an, möglichst schnell die Bahn des Planeten zu berechnen, um ihn wiederfinden zu können. Je länger die Berechnung dauerte, desto größer würde die Abweichung des Planeten von seinem berechneten Ort wegen der unvermeidlichen Beobachtungsfehler werden. Man kann ohne Übertreibung sagen, daß ganz Europa fieberhaft zu rechnen begann [36, S. 44]. Gauß löste diese Aufgabe auf typisch mathematische Weise. Er überlegte erst einmal, wie das Problem genau zu formulieren sei, dann entwickelte er drei neue mathematische Methoden und schließlich rechnete er – und war als Erster fertig.

Eine der neuen Methoden Gauß' war eben die Methode der kleinsten Quadrate. Er hatte eine relativ große Anzahl von Beobachtungen vorliegen, diese waren jedoch mit unvermeidlichen Meßfehlern behaftet. Er formulierte daher die Aufgabenstellung wie folgt:

Gesucht ist eine Bahn des Planeten, für die die mittlere quadratische Abweichung zu den Beobachtungen möglichst klein wird.

Diese Vorgehensweise hat zwei ganz wichtige Vorteile

- Sie führt zu einer besonders „schönen“ Theorie.
- Man kann mit der mittleren quadratischen Abweichung besonders gut rechnen. Dies war insbesondere zu Gauß' Zeiten von enormer Bedeutung, aber auch heute noch ist diese Methode aus genau diesem Grunde verbreitet und beliebt.

Ein Nachteil der Methode ist, daß extrem abweichende oder falsche einzelne Meßwerte das Resultat völlig verfälschen können.

Als ein ganz einfaches Beispiel für Mittelwert und Varianz betrachten wir wieder den Würfelwurf. Bei einer langen Serie von Würfeln wird jede Augenzahl etwa mit der Häufigkeit $1/6$ auftreten. Daher ist der Mittelwert der Augenzahl

$$\mu = \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5.$$

Damit erhält man für die Varianz

$$\begin{aligned} \sigma^2 &= \frac{1}{6} \cdot ((3.5 - 1)^2 + (3.5 - 2)^2 + (3.5 - 3)^2 + (3.5 - 4)^2 + (3.5 - 5)^2 + \\ &\quad + (3.5 - 6)^2) = \frac{1}{6} \cdot 2 \cdot (2.5^2 + 1.5^2 + 0.5^2) \approx 2.92 \end{aligned}$$

oder $\sigma \approx 1.71$.

5.2 Grenzwertsätze

Natürlich ist die Gaußsche Normalverteilung nur eine Annäherung an die Wirklichkeit. Man sieht leicht ein, daß reale Fehlerverteilungen **nicht** immer der Normalverteilung genügen können. Es

kann beispielsweise vorkommen, daß die wahre Fehlerkurve nicht so schön symmetrisch ist wie in Abbildung 5, es könnte durchaus sein, daß Abweichungen „nach unten“ aus irgendwelchen Gründen häufiger auftreten als Abweichungen „nach oben“. Nach der Gauß-Verteilung treten extreme Abweichungen selten auf, aber die Wahrscheinlichkeit dafür ist eben nicht gleich Null. Das würde im obigen Fadenbeispiel heißen, daß mit einer gewissen Wahrscheinlichkeit auch Abweichungen vorkommen müssen, die negatives Vorzeichen haben und einen Betrag, der größer ist als 0.1 mm. Dann aber hätte der Faden eine negative Dicke, was nicht nur unvorstellbar sondern sogar physikalisch unmöglich ist. Trotz dieser Vorbehalte ist die Normalverteilung meist eine gute Annäherung an die Wirklichkeit. Die Gründe dafür findet man in den sogenannten *Grenzwertsätzen* der Wahrscheinlichkeitstheorie beziehungsweise im *Gesetz der großen Zahlen*. Einer dieser Grenzwertsätze ist nach dem bedeutenden Mathematiker Jakob Bernoulli (1654 – 1705) benannt. Man könnte ihn so formulieren:

Die Wahrscheinlichkeit für das Eintreffen eines Ereignisses E sei p . Nach n unabhängigen Versuchen soll das Ereignis E n_1 -mal eingetreten sein. Dann gilt für eine beliebig gewählte positive reelle Zahl ε

$$P\left(\left|\frac{n_1}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4\varepsilon^2 n}.$$

Umgangssprachlich würde dies etwa heißen: Wenn man hinreichend lange beobachtet (das heißt, wenn n groß ist), dann weicht der empirisch gewonnene Näherungswert n_1/n mit einer vorhersagbaren Wahrscheinlichkeit beliebig wenig von dem wahren Wert p ab. Illustrativer ist das folgende Beispiel dreier Münzwurfexperimente, wobei $\varepsilon = 0.1$ gewählt wurde [18, S. 658]:

Autor	n	Kopf oben n_1	$\frac{n_1}{n}$	$1 - \frac{1}{4\varepsilon^2 n}$
Buffon	4 040	2 048	0.507	0.993 8
K. Pearson	12 000	6 019	0.504 6	0.997 9
K. Pearson	24 000	12 012	0.500 5	0.999 0

Eine sehr wichtige und auch überraschende Aussage der Wahrscheinlichkeitstheorie ist der *zentrale Grenzwertsatz*. Er soll hier nicht in allen Details ausformuliert werden. Wir betrachten auch hier einen – etwa zeitabhängigen – Prozeß, der eine gewisse Verteilungsfunktion besitzt, die uns hier nicht zu interessieren braucht, wichtig ist nur, daß sie einen Mittelwert μ hat und eine Varianz σ^2 .

Gesucht ist die Wahrscheinlichkeit, daß der *gleitende Mittelwert* $\hat{z}_n = \frac{1}{n} \cdot \sum_{i=1}^n z_i$ von n zufälligen

Versuchen von dem Mittelwert μ um nicht mehr als eine vorgegebene positive reelle Zahl δ abweicht. Diese läßt sich angeben, es ist nämlich

$$P(\mu - \delta \leq \hat{z}_n \leq \mu + \delta) = \sqrt{\frac{1}{2\pi}} \cdot \int_{-\frac{\delta}{\sigma}\sqrt{n}}^{\frac{\delta}{\sigma}\sqrt{n}} e^{-t^2/2} dt + O\left(\frac{1}{\sqrt{n}}\right).$$

Wir wollen uns nicht bemühen, diese Formel zu „verstehen“, ich weise nur auf einige Besonderheiten hin:

- Der letzte Term $O\left(\frac{1}{\sqrt{n}}\right)$ ist ein Fehlerterm, der besagt, daß der Fehler dieser Darstellung mit $1/\sqrt{n}$ klein wird. Die gute Nachricht ist, daß er mit wachsendem n klein wird, die schlechte Nachricht, daß dies quälend langsam geschieht, wie wir schon an unserem Würfelexperiment

gesehen hatten. Um ein „Gefühl“ für die Geschwindigkeit der Annäherung zu bekommen, betrachten wir für einige Werte von n die Größe $1/\sqrt{n}$:

n	$1/\sqrt{n}$
100	0.1
10 000	0.01
1000 000	0.001

Mit anderen Worten, wenn wir bei den genannten Experiment bis auf eine Größenordnung von $1/100$ an den Mittelwert gelangen wollen, dann sind dazu immerhin 10 000 Würfe nötig.

- Der Ausdruck, der unter dem Integralzeichen steht, die Funktion $e^{-t^2/2}$, ist nichts anderes als die Gaußsche Fehlerfunktion. Dieses recht überraschende Resultat zeigt, daß diese Funktion die Bezeichnung *Normalverteilung* zu Recht trägt und daß die Tatsache, daß die Normalverteilung mehr oder weniger bei beliebigen Zufallsexperimenten auftritt, tiefliegende Gründe hat.
- Eine ganz kuriose Sache ist, daß in der Formel die Zahl π , das Verhältnis von Kreisumfang zu Durchmesser, auftritt. Dies ist wirklich überraschend: Wo ist hier der Kreis? Dieses unvermutete Auftreten von π an einer unerwarteten Stelle zeigt aber auch sehr schön, daß in der Mathematik alles mit allem verknüpft ist.

Der Physik-Nobelpreisträger Eugene Paul Wigner (1902 – 1995) schrieb einen Artikel mit dem Titel *The unreasonable effectiveness of mathematics in the natural sciences*. Der Artikel beginnt mit einer kleinen Geschichte [34], die diesen letztgenannten mysteriösen Zusammenhang von π und der Wahrscheinlichkeitstheorie herausstellt:

There is a story about two friends, who were classmates in high school, talking about their jobs. One of them became a statistician and was working on population trends. He showed a reprint to his former classmate. The reprint started, as usual, with the Gaussian distribution and the statistician explained to his former classmate the meaning of the symbols for the actual population, for the average population, and so on. His classmate was a bit incredulous and was not quite sure whether the statistician was pulling his leg. “How can you know that?” was his query. “And what is this symbol here?” “Oh,” said the statistician, “this is pi.” “What is that?” “The ratio of the circumference of the circle to its diameter.” “Well, now you are pushing your joke too far,” said the classmate, “surely the population has nothing to do with the circumference of the circle.”

Anhang: Beispiele und Anwendungen

A Das Beispiel von Pacioui

Wir kommen auf das Spielerbeispiel von Pacioui (Seite 8) zurück. Die fragliche Wahrscheinlichkeit kann man – mit heutigen Verständnis – recht leicht finden, wenn man etwa wie folgt argumentiert:

Wenn die Spieler weitergespielt hätten, dann würde beim nächsten Zug der eine Spieler (S1) mit 50 % Wahrscheinlichkeit das Spiel gewonnen haben. Mit ebenfalls 50 % Wahrscheinlichkeit würde nach dem nächsten Zug der Spieler S1 5 Punkte behalten haben, aber Spieler S2 würde 3 Punkte haben (wir ignorieren des Fall des Gleichstandes, dann erhält jeder Spieler 0 Punkte).

Beim darauffolgenden Zug bestehen wieder zwei Möglichkeiten: Entweder S1 gewinnt mit ebenfalls 50 % Wahrscheinlichkeit oder aber mit 50 % Wahrscheinlichkeit hätte Spieler S2 vier Punkte.

In Abbildung 6 sind die möglichen Fälle graphisch dargestellt. Wir sehen, daß es nur einen Fall gibt, in dem Spieler S2 gewinnt, und dieser Fall tritt mit einer Wahrscheinlichkeit von $0.5 \times 0.5 \times 0.5 \times 0.5 = 0.0625$ ein. Die Wahrscheinlichkeit, daß S2 gewinnt wäre demnach 6.25 %. Wie hoch ist die Gewinnwahrscheinlichkeit von S1? Nach den Regeln der Wahrscheinlichkeit müssen wir die Wahrscheinlichkeiten der einzelnen möglichen Spielausgänge addieren (unabhängige Ereignisse): Der Ausgang (6,2) tritt mit 50 % Wahrscheinlichkeit ein, der Ausgang (6,3) mit 25 %, (6,4) mit 12.5 % und (6,5) mit 6.25 %, das heißt, die gesuchte Gewinnwahrscheinlichkeit ist $0.5 + 0.25 + 0.125 + 0.0625 = 0.9375$ oder 93.75%.

B Noch einmal: Das Rotwild

Wir kehren noch einmal zu dem Beispiel mit dem Rotwild zurück. Es sei A das Ereignis: „Während einer Wanderung im Walde von sechs Stunden trifft man auf mindestens ein Stück Rotwild“. B sei das Ereignis: „Während einer Wanderung im Walde von sechs Stunden trifft man auf ein weiteres Stück Rotwild“. Um eine Zahl anzugeben, nehmen wir $P(A)$ mit $1/100$ an, daß heißt, auf hundert Wanderungen trifft man einmal auf ein Stück Rotwild. Wir brauchen nun noch die Wahrscheinlichkeit von $A \wedge B$. Diese ermitteln wir empirisch: Wir haben beobachtet, daß unter einer „großen“ Anzahl von Wanderungen (sagen wir, 10 000 Wanderungen) 100mal überhaupt Rotwild gesichtet wurde, was der Wahrscheinlichkeit von $1/100$ entspräche. Weiterhin hat man unter 100 Begegnungen mit Rotwild 80 mal beobachtet, daß einem Stück Rotwild unmittelbar ein weiteres folgte. Das würde hier bedeuten, daß die (empirische) Wahrscheinlichkeit von $A \wedge B$ sich wie folgt berechnet

$$\begin{aligned} P(A \wedge B) &= \frac{\text{Beobachtungen eines Stücks Rotwild, gefolgt von einem anderen}}{\text{Gesamtzahl der Wanderungen}} = \\ &= \frac{80}{10\,000} = 1/125 = 0.008. \end{aligned}$$

Die Wahrscheinlichkeit, nach Sichtung eines Stücks Rotwild unmittelbar ein weiteres zu sehen, ist nach der Formel für die bedingte Wahrscheinlichkeit damit

$$P(B|A) = \frac{P(B \wedge A)}{P(A)} = \frac{1/125}{1/100} = \frac{100}{125} = \frac{4}{5} = 0.8,$$

das heißt, wenn man ein Stück Rotwild gesehen hat, dann muß man mit 80 % Wahrscheinlichkeit damit rechnen, daß ein weiteres Rotwild folgt.

Es muß wohl nicht erwähnt werden, daß die hier angegebenen Zahlen nicht sauber erhoben, sondern grob geschätzt wurden, sie müssen also nicht realistisch sein.

C Eine aktuelle Gerichtsentscheidung

Am 26. November 2008 wurde vom Landesarbeitsgericht Berlin-Brandenburg ein Urteil gefällt, welches in der Presse lebhaftes Echo fand. Dies aus zwei Gründen: Einmal handelte sich um ein Verfahren zum Thema Diskriminierung am Arbeitsplatz und zum anderen wurde wohl zum ersten Male in Deutschland ein wahrscheinlichkeitstheoretisches Argument als Indiz angeführt. Der Sachverhalt⁴⁾: Eine Angestellte der Firma GEMA fühlte sich als Frau diskriminiert, weil der Posten des Personaldirektors an einen Mann vergeben wurde, obwohl die Klägerin gleich qualifiziert war. In der Firma war der Anteil der Frauen unter den Beschäftigten 85 %, jedoch waren sämtliche 16 Direktoren männlich. Die Klägerin hatte ein „mathematisches Gutachten“ erstellen lassen, nach dem die Wahrscheinlichkeit, daß sämtliche Direktoren *zufällig* Männer waren, bei dem genannten hohen Frauenanteil weniger als ein Prozent beträgt.

Diese Entscheidung wurde auch deshalb als besonders wichtig angesehen, weil eine derartige Argumentation, die in den USA gängiges Argument in solcherlei Fällen ist, nunmehr auch in Deutschland mit Erfolg angewandt wurde. Die Problematik der Verwendung statistischer Gutachten vor Gericht wird in zwei Artikeln in der Zeitschrift *Spektrum der Wissenschaft* diskutiert [26, 21].

D Der Fall der zwei Mädchen

Eine bekanntes Beispiel, das sogar Fachleute in Verwirrung gebracht hat, läßt sich wie folgt beschreiben [14]⁵⁾:

Sie wissen, daß eine Familie zwei Kinder hat, und Sie wissen zusätzlich, daß eines davon ein Mädchen ist. Wie groß ist die Wahrscheinlichkeit, daß beide Kinder Mädchen sind?

Diese Aufgabe scheint leicht lösbar zu sein. Vergessen wir das Kind, dessen Geschlecht wir kennen und fragen wir uns, wie groß ist die Wahrscheinlichkeit, daß das verbleibende Kind ein Mädchen

⁴⁾Siehe etwa die Information der Anwaltskanzlei Heinemann, Magdeburg
<http://www.raheinemann.de/16.98.Diskriminierung+per+Mathematik.html>
oder den Beitrag in SPIEGEL ONLINE vom 26. November 2008,
<http://www.spiegel.de/wirtschaft/0,1518,592973,00.html>.

⁵⁾Siehe auch Gero von Randow, DIE ZEIT Nr. 27, 28. Juni 1996, Seite 29

ist. Dies scheint klar: Die Wahrscheinlichkeit, daß irgendein Kind ein Mädchen ist, beträgt (ziemlich genau) $1/2$. Also müßte dies auch hier der Fall sein. Wenn wir nach dem kombinatorischen Rezept vorgehen, Anzahl der günstigen Fälle durch Anzahl der möglichen Fälle, dann stellt sich die Situation wie folgt dar:

Mögliche Fälle für zwei Kinder:

Kind 1	Kind 2	
Mädchen	Mädchen	
Mädchen	Junge	
Junge	Mädchen	
Junge	Junge	Dieser Fall kann hier nicht vorliegen.

Wir haben also drei mögliche Fälle. In zwei Fällen ist das „andere“ Kind ein Junge, in einem ein Mädchen. Die fragliche Wahrscheinlichkeit ist also $1/3$.

Die auf den ersten Blick überraschende Lösung der Aufgabe ist korrekt, man kann sie etwa mit zwei Würfeln – oder noch besser mit zwei Münzen – „empirisch“ testen. Sie hat mit dem Begriff *bedingte Wahrscheinlichkeit* zu tun, einem Begriff, der – wie erwähnt – anschauliche Schwierigkeiten bereitet.

Wir nehmen an, wir hätten zwei Ereignisse, nennen wir sie A (hier: beide Kinder sind Mädchen) und B (mindestens ein Kind ist ein Mädchen). Die sogenannte *a-priori-Wahrscheinlichkeit* für das Ereignis A , wir nennen sie $P(A)$, ist die Wahrscheinlichkeit, die dieses Ereignis hat, wenn wir keinerlei Vorwissen haben. Diese ist in unserem Beispiel unstrittig $1/4$, denn es gibt die genannten vier Fälle, und einer davon ist der „günstige“. Wie groß ist $P(B)$, die a-priori-Wahrscheinlichkeit von B ? Auszählen der Fälle liefert $3/4$.

Gefragt ist nun die Wahrscheinlichkeit des Eintretens von A , wenn bekannt ist, daß das Ereignis B eingetreten ist. Das heißt aber doch, daß nunmehr die „möglichen Fälle“ in Sinne der kombinatorischen Wahrscheinlichkeitsdefinition die Fälle sind, bei denen B eingetreten ist, und das sind drei Fälle. Es gibt unter diesen nur einen „günstigen“ Fall, in dem A eintritt, also müssen wir wie oben den Quotienten $1/3$ bilden, und dieser ist – natürlich – gerade der Quotient $P(A)/P(B)$.

Interessanter ist die folgende Überlegung, die Dewdney anstellt: Wir kommen auf die Wahrscheinlichkeit von $1/3$, weil wir die Kinder nicht unterschieden haben. Würden wir wissen, daß das ältere Kind ein Mädchen ist, dann wäre die Wahrscheinlichkeit, daß das jüngere Kind auch ein Mädchen ist, $1/2$, denn das Geschlecht des jüngeren Kindes hängt natürlich nicht von dem des älteren ab, beide Ereignisse sind voneinander *unabhängig*. Wir kamen auf die Wahrscheinlichkeit von $1/3$, indem wir die beiden Fälle „das ältere Kind ist ein Mädchen“ und „das jüngere Kind ist ein Mädchen“ als verschieden betrachteten.

Diese letztgeannte Überlegung wäre geeignet, Verwirrung zu stiften. Also war die erste Überlegung, die auf die Wahrscheinlichkeit $1/2$ geführt hatte, doch richtig? Was diese kleine Aufgabe so interessant macht, ist, daß sich hier sehr schön zeigt, wie subtil Zusatzinformationen („Das ältere Kind ist ein Mädchen“) sich auswirken können. Wir wissen mehr, und dieses Zusatzwissen äußert sich in einer größeren Wahrscheinlichkeit.

Übrigens steigert Dewdney in seinem Artikel die Verwirrung noch weiter. Wenn unsere Information über die beiden Kinder in einem Bild besteht, auf dem das eine Kind deutlich als Mädchen identifizierbar ist, das andere jedoch mit einem großen Hund spielt und nicht vollständig sichtbar

ist, so daß sein Geschlecht nicht erkennbar wird, dann kann man die Kinder „individualisieren“, indem man argumentiert: Das Kind, welches nicht mit dem Hund spielt, ist ein Mädchen. Wie wahrscheinlich ist dann, daß das Kind, welches mit dem Hund spielt, ebenfalls ein Mädchen ist? Die Zusatzinformation „das mit dem Hund spielt“ reicht aus, um die Wahrscheinlichkeit von $1/2$ zu ergeben.

Fazit: Derartige Aufgabenstellungen, für die es eine „Geschichte“ gibt, sind in unserer Welt oder doch zumindest im Niemandsland an der Grenze unserer Welt angesiedelt. Sie hängen daher empfindlich vom Kontext ab und wir können als Fazit formulieren: Vorsicht!

E Bedingte Wahrscheinlichkeit vor Gericht

Dewdney bringt noch eine Anzahl von sehr schönen Beispielen für die Fallstricke des Wahrscheinlichkeitsbegriffs. Wer an einem sehr interessanten und vergnüglichen Beispiel interessiert ist, sollte in dem Buch von Gero von Randow über das „Ziegenproblem“ nachlesen [28].

Wir wollen hier ein etwas seröseres – wenngleich konstruiertes – Beispiel untersuchen. Wir nehmen an, ein Mensch werde angeklagt, eine Bank überfallen zu haben. Es gibt keine brauchbare Identifikation des Bankräubers, aber Zeugen haben angegeben, daß der Angeklagte zwei Tage vor dem Überfall „auf verdächtige Weise vor der Bank herumgelungert“ habe. Der Anwalt des Angeklagten findet heraus, daß die Bank aus irgendwelchen Gründen über einen langen Zeitraum die Aufzeichnungen der Überwachungskameras ausgewertet hat. Die Bänder der Kameras wurden natürlich gelöscht, die Aufzeichnungen wurden automatisch gewonnen und sind anonymisiert. Aus diesen Auswertungen geht hervor, daß in diesem Zeitraum eine sehr große Anzahl von Personen sich ohne ersichtlichen Grund vor der Bank aufgehalten habe, also genau wie der Beklagte. Der Anwalt argumentiert: Es haben im Beobachtungszeitraum tausende von Personen dieses Verhalten gezeigt. In diesem Zeitraum wurde die Bank einige wenige Male überfallen, das heißt aber doch, daß von dieser „repräsentativen Stichprobe“ von „verdächtigen“ Personen **höchstens** ein Bruchteil eines Promilles einen Banküberfall begangen haben kann. Und das bedeutet, daß die Tatsache des „Herumlungerns“ in keiner Weise ein Indiz für eine Überfallsabsicht sein kann.

Um konkrete Zahlen zu haben, nehmen wir an, in einem Beobachtungszeitraum von 1 000 Tagen habe die Auswertung 5 000 Fälle von „verdächtigem Verhalten“ ergeben. Im gleichen Zeitraum sei die Bank zweimal überfallen worden. Da unter den 5 000 Verdachtsfällen sich nur zweimal ein Täter befunden haben kann (wir vernachlässigen hier, daß es sich bei 5 000 beobachteten nicht unbedingt um verschiedene Personen handeln muß und daß der Täter die Bank vielleicht mehr als einmal beobachtet hat), dann ist also die Wahrscheinlichkeit, daß ein zufällig herausgegriffener verdächtiger Passant der Täter ist, in der Tat höchstens $2/5\,000 = 0.0004$ oder 0.04 %.

Hiergegen könnte der Anklagevertreter einwenden: Wir reden hier nicht von a-priori-Wahrscheinlichkeiten, sondern von bedingten Wahrscheinlichkeiten. Eine Person, die verdächtiges Verhalten zeigt, kann dies aus zahllosen harmlosen Gründen tun. Wirklich relevant für uns sind diejenigen Fälle, bei denen sich eine Person verdächtig machte **und** unmittelbar darauf ein Überfall stattfand. Wir definieren hier:

Ereignis A Eine sich verdächtig verhaltende Person begeht einen Überfall.

Die (a-priori-) Wahrscheinlichkeit dafür daß eine verdächtige Person einen Überfall begeht, ist – wie vom Anwalt des Beklagten richtig ausgeführt – (natürlich unter gewissen Annahmen, um die

wir uns hier nicht kümmern) gegeben durch

$$P(A) = \frac{\text{Anzahl der Überfälle}}{\text{Anzahl der verdächtigen Personen}} = \frac{2}{5\,000} = 0.0004.$$

Nun haben die meisten dieser Personen mit dem Überfall nichts zu tun. Es ist nicht anzunehmen, daß jemand, der sich ein halbes Jahr vor einem Überfall auffällig verhält, irgendwie im Zusammenhang mit dem Überfall steht. Wir definieren daher

Ereignis B Der Beobachtungstag des verdächtigen Verhaltens liegt höchstens drei Tage vor einem Überfall.

Es ist dann die empirische Wahrscheinlichkeit, die sich aus den vorliegenden Daten ergibt:

$$P(B) = \frac{\text{Anzahl der in Frage kommenden Tage}}{\text{Gesamtzahl der Beobachtungstage}} = \frac{2 \text{ Überfälle} \times 3 \text{ Tage}}{1\,000 \text{ Tage}} = 0.006.$$

Die Wahrscheinlichkeit für das Ereignis A unter der Zusatzbedingung, daß auch noch Ereignis B auftritt, wäre dann der Quotient aus der Anzahl der verdächtigen Personen in den sechs kritischen Tagen, die einen Überfall begehen und der Gesamtzahl der verdächtigen Personen in dieser Zeit. Unter den verdächtigen Personen begehen höchstens zwei einen Überfall, da ja nur zwei Überfälle stattgefunden haben. Wenn wir annehmen, daß in den insgesamt sechs Tagen, die vor den beiden Überfällen gelegen haben, dreißig Personen sich verdächtig machten, dann kann man bestenfalls schließen, daß die gesuchte Wahrscheinlichkeit höchstens $2/30 = .0667$ oder höchstens 6.67 % beträgt. Dies klingt schon etwas anders. Damit ist nicht erhärtet, daß der Angeklagte wirklich mit einer nennenswerten Wahrscheinlichkeit schuldig ist, es ist nur gezeigt, daß die Abschätzung des Anwalts des Beklagten nichts mit der Realität – genauer nichts mit korrekter Wahrscheinlichkeitsrechnung – zu tun hat.

Nun kommt die bedingte Wahrscheinlichkeit in ihrer vollen Schönheit zur Anwendung: Unsere „relevante“ Grundgesamtheit sind alle diejenigen Fälle, in denen sich eine Person verdächtig verhielt **und** daß dieses Ereignis höchstens drei Tage vor einem Überfall stattfand. Diese **und**-Verknüpfung symbolisiert man in der Wahrscheinlichkeitstheorie mit dem Symbol $\wedge$, das heißt, die für uns relevanten Ereignisse sind gegeben durch

$$A \wedge B = \begin{array}{l} \text{Eine sich verdächtig verhaltende Person begeht einen Überfall} \\ \textbf{und} \quad \text{Der Beobachtungstag des verdächtigen Verhaltens} \\ \text{liegt höchstens drei Tage vor einem Überfall.} \end{array}$$

Nun hatten wir angenommen, daß unter den verdächtigen Personen der Täter zu finden sei. Damit ist dann $P(A \wedge B)$ gleich 2 (nämlich die (maximale) Anzahl der Täter unter den verdächtigen Personen) dividiert durch die Anzahl (nämlich 30) der sich verdächtig verhaltenden Personen in den 2×3 relevanten Tagen.

Für diejenigen, die sich schon mit Wahrscheinlichkeitstheorie beschäftigt haben: Man kann hier die Wahrscheinlichkeit des Ereignisses $A \wedge B$ nicht aus den Wahrscheinlichkeiten von A und B unmittelbar ermitteln, da wir die Anzahl der Überfälle festhalten, und damit hängt die Wahrscheinlichkeit von A von der Anzahl der Bezugstage ab.

Wir hatten eben angenommen, daß in den sechs kritischen Tagen 30 Personen auffälliges Verhalten gezeigt haben. Dies entspricht dem Mittelwert des gesamten Beobachtungszeitraums, wir hatten

5 000 verdächtige Personen in 1 000 Tagen, also im Mittel fünf verdächtige Personen pro Tag. Bei zwei Überfällen und damit sechs relevanten Tagen beziehungsweise 30 in Frage kommenden Personen ist dann die Wahrscheinlichkeit des Ereignisses A unter der Bedingung, daß B vorliegt – diese Wahrscheinlichkeit nennt man eben die *bedingte Wahrscheinlichkeit* von A und bezeichnet sie mit $P(A|B)$ – gerade $2/30$. Nach der Formel für bedingte Wahrscheinlichkeiten, die gelegentlich auch die Bayessche Formel genannt wird (die eigentliche Bayessche Formel ist etwas allgemeiner), gilt

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

Setzen wir die Zahlen ein, dann wird

$$P(A|B) = \frac{2/30}{6/1000}$$

Im obigen Falle, dem Fall der beiden Mädchen, war $P(A \wedge B) = P(A)$, weil im Falle A (beide Kinder sind Mädchen) der Fall B (mindestens ein Kind ist Mädchen) ohnehin vorliegt.

Nun modifizieren wir wieder die Aufgabenstellung ein wenig, indem wir annehmen, daß ausgerechnet vor den beiden Überfällen die Anzahl der „Herumlungerer“ etwas geringer gewesen war. Nehmen wir an, daß in der unmittelbaren Nähe der Bank eine Bushaltestelle sei, und daß ein Teil der von der Überwachungskamera erfaßten Personen einfach nur auf den Bus wartete. Zufällig war etwa zeitgleich mit den Überfällen die Bushaltestelle wegen Straßenbauarbeiten verlegt worden. Dann würde unsere Rechnung so aussehen. In den fraglichen sechs Tagen gab es nur 10 Fälle von „verdächtigen Personen“, das heißt, es wir haben die empirische Wahrscheinlichkeit

$$P(A \wedge B) = \frac{10}{5\,000} = \frac{1}{500}.$$

Unnötig, anzumerken, daß diese ganze Geschichte nicht ganz realistisch ist. So haben wir sehr frei von der Tatsache Gebrauch gemacht, daß im Berichtszeitraum nur zwei Überfälle stattgefunden haben, und eine Gesamtheit von zwei Fällen ist für die Statistik nicht brauchbar. Wir könnten uns hier damit herausreden, daß wir ja ursprünglich nur eine Abschätzung der Höchstwahrscheinlichkeit im Sinne hatten.

Jedoch: Ich gebe noch zwei Beispiele dafür an, daß derartige Argumentationen nicht so ganz weit hergeholt sind. Das erste Beispiel ist eher anekdotisch: Anfang der sechziger Jahre hatte ich selbst Gelegenheit, eine Nacht in der Arrestzelle einer Polizeiwache zu verbringen. Der Grund: Ich hatte mich „verdächtig“ in der Nähe eines Automaten, der in der Vergangenheit mehrfach aufgebrochen worden war, aufgehalten. Der Fairness halber muß ich hinzufügen, daß der betreffende Beamte sich einige Tage später bei mir für seine Überreaktion entschuldigte und daß mein Verhalten bei der Festnahme nicht eben der Wahrheitsfindung dienlich war.

John Allen Paulos schreibt in einem seiner Bücher (und in anderen Publikationen) über verschiedene Anwendungen der Statistik in der Rechtsfindung [27]. Eine dieser Anwendungen betrifft den Fall von O. J. Simpson, der ein bekannter und populärer Footballstar in den USA war und den man 1994 des Mordes an seiner Ex-Frau angeklagt hatte. Das öffentliche Interesse an diesem Fall war gewaltig, und man kann vielleicht sagen, daß in einer solchen Situation in den USA eine Gerichtsverhandlung in Gefahr gerät, zu einer Show zu degenerieren. In einem umstrittenen Prozeß wurde Simpson freigesprochen. Eine interessante Nuance: In einem Zivilprozeß wurde Simpson schuldig gesprochen und zu einer hohen Schadensersatzleistung verurteilt, und im Jahre 2008

wurde er wegen Beteiligung an einem Raubüberfall zu 33 Jahren verurteilt. Inzwischen war seine Popularität allerdings deutlich zurückgegangen.

Paulos schreibt zu dem Mordprozeß [27, S. 85]:

Zu meiner Unzufriedenheit mit der Geschichte Simpsons trugen zahllose Beispiele für das bei, was man *Statistizid* nennen könnte. Lassen Sie mich mit einem Refrain beginnen, der von Anwalt Alan Dershowitz während der Verhandlung permanent wiederholt wurde. Er erklärte, die Misshandlungen in der Ehe Simpsons seien irrelevant für den Fall, da weniger als eine von 1 000 Frauen, die von ihren Ehemännern misshandelt wurden, später von ihnen getötet werden. Obwohl die Zahlen korrekt sind, ist Dershowitz' Behauptung ein verblüffendes non sequitur; sie ignoriert nämlich eine offensichtliche Tatsache: Nicole Simpson wurde getötet.

F Umwelt-Östrogene und Brustkrebs

Nach [5, S. 185 f.]:

In der – immerhin recht seriösen – Zeitschrift *Spektrum der Wissenschaft* erschien im Jahre 1995 ein Artikel mit dem Titel «Verursachen Umwelt-Östrogene Brustkrebs?» [12]. Dieser Beitrag beruht auf einer Arbeit, die in dem Fachjournal *Environmental Health Perspectives* publiziert worden war [13], will also immerhin als gewichtiger Beitrag zur Forschung verstanden werden. In diesem Artikel legten die Autoren dar, daß in der letzten Zeit in den USA das Brustkrebsrisiko gestiegen sei. In der Tat ist in der Zeit von 1973 bis 1991 die Anzahl der Brustkrebsfälle von etwa 80 auf etwa 110 Fälle pro 100 000 Frauen gestiegen. Bei Frauen über fünfzig Jahre stieg im gleichen Zeitraum die Anzahl der Fälle sogar von etwa 250 auf etwa 340 pro 100 000.

In ihrem interessanten und auch amüsanten Buch „Der Hund, der Eier legt“ weisen die beiden Autoren Hans-Peter Beck-Bornholdt und Hans-Hermann Dubben darauf hin, daß die in dem genannten Artikel gezogenen Schlüsse höchst fragwürdig sind und daß man aus den vorgelagten Daten keinesfalls auf eine Erhöhung der Brustkrebsrate schließen darf. Einflüsse wie verbesserte Diagnosemethoden oder bessere Früherkennung durch vermehrte Vorsorgemaßnahmen werden in dem Artikel zwar erwähnt, aber sie werden nicht quantitativ berücksichtigt. Insbesondere muß man bei den angegebenen Zahlen berücksichtigen, daß in dem betrachteten Zeitraum auch die Lebenserwartung erheblich zugenommen hat. Das bewirkt aber, daß früher eine ganze Anzahl von Frauen den Krebs nicht mehr „erlebt“ hat. Zudem müßte man untersuchen, ob mit zunehmenden Alter sich das Krebsrisiko erhöht. In dem Artikel werden auch die Fallzahlen für Frauen unter fünfzig Jahren angegeben, und diese Fallzahlen zeigen keine wachsende Tendenz, was die Hypothese stützen würde, daß in der Tat die Wahrscheinlichkeit von Brustkrebs mit zunehmendem Alter steigt.

G Wie kann man an der Börse gewinnen?

Aktienkurse stellen ein klassisches Beispiel eines stochastischen Prozesses dar, und alle Behauptungen, daß man den Kursverlauf mit irgendwelchen Methoden vorhersagen könne, entbehren der Grundlage. Das „Börsenspiel“ wird regiert von einem Prinzip der Wirtschaftswissenschaften, welches dem Energiesatz der Physik gleichzusetzen ist, dem Prinzip der *Arbitragefreiheit*. Unter Arbitrage versteht man das risikofreie oder doch risikoarme Erzielen von Gewinnen durch

Ausnutzen von Preisunterschieden ⁶⁾. Hätte man irgendeine Methode, die Börsenkurse etwa des nächsten Tages mit einer Wahrscheinlichkeit von mehr als 50 % vorherzusagen, dann hätte man die Möglichkeit, langfristig Gewinne zu erzielen. Da Gewinne eines Teilnehmers an der Börse Verluste der anderen Mitspieler bedeuten, werden diese irgendwann die Lust am Spiel verlieren und sich zurückziehen. Das heißt, die Tatsache, daß einer der Akteure an der Börse regelmäßig gewinnt, wird das Gesamtsystem „Börse“ langfristig destabilisieren. Aus diesem Grunde verhalten sich Börsenkurse auf eine Weise, die Vorhersagbarkeit ausschließen. Kurzfristige Vorhersagbarkeit würde aber impliziert, wenn die Kurven der Börsenkursentwicklungen etwa differenzierbar wären. Aus diesem Grunde weisen diese Kurven das „gezackte“ Verhalten auf, welches man in der Physik bei der Brownschen Bewegung kennt (Abbildung 7). Wir haben hier ein Beispiel einer nirgends differenzierbaren Funktion, ein Phänomen, welches noch Mitte des zwanzigsten Jahrhunderts als exotisch empfunden wurde. Diese Reaktion der Finanzmärkte, risikofreie Gewinnmöglichkeiten „automatisch“ auszugleichen, wird durch das Verhalten der Teilnehmer noch verstärkt [16, S. 97]:

In der Tat gibt es Hedgefonds, die systematisch mit Hilfe hochkarätiger Programme nach Arbitrage-Möglichkeiten in den Finanzmärkten suchen, sie ausnutzen und dadurch zum Verschwinden bringen.

In der letzten Zeit hat es eine Krise auf den Finanzmärkten gegeben. Über die Gründe gibt es zahlreiche mehr oder weniger begründete Vermutungen. Nach Eberlein ist eine wichtige Ursache darin zu suchen, daß völlig inadäquate und nicht aktuelle Mathematik eingesetzt wird [16].

H Und wie ist es mit dem Zahlenlotto?

Wir alle kennen das Zahlenlotto. Man muß in einem Formular sechs Zahlen aus 49 aussuchen (wir ignorieren einmal die Zusatzzahl), und wenn bei der Ziehung der Zahlen eine oder mehrere der angekreuzten mit den gezogenen Zahlen übereinstimmt, dann kann man einen mehr oder weniger großen Geldbetrag gewinnen.

Wir wollen versuchen, die Wahrscheinlichkeit, die „sechs richtigen“ Zahlen anzukreuzen, auszurechnen. Dies ist nicht schwierig. Für die erste Zahl gilt: Die Gesamtzahl der möglichen Ereignisse ist 47, nämlich die Anzahl der Zahlen auf dem Schein. Die Anzahl der günstigen Ereignisse ist 6, nämlich die Anzahl der gezogenen Zahlen. Also haben wir die Wahrscheinlichkeit, eine der Zahlen richtig anzukreuzen mit $\frac{6}{49}$ gefunden.

Wir groß ist die Wahrscheinlichkeit, noch eine weitere Zahl richtig zu treffen? Wir müssen hier nicht die Bayessche Formel bemühen, wir rechnen einfach so: Es sind 48 Zahlen verblieben, und wir möchten unter den verbleibenden 5 Zahlen die richtige finden. Die Wahrscheinlichkeit ist hier $\frac{5}{48}$. Beide Ereignisse – die eine Zahl zu finden und dann die zweite ebenfalls richtig anzukreuzen, sind (in unserer Modellierung) voneinander unabhängig, also ist die Wahrscheinlichkeit, zwei Zahlen richtig zu finden, $\frac{6 \cdot 5}{49 \cdot 48}$.

Die gesuchte Wahrscheinlichkeit für „sechs Richtige“ ist also

$$\frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{720}{10\,068\,347\,520} = 0.000\,000\,071\,511 \dots = 7.1511 \dots \cdot 10^{-8}$$

⁶⁾Nach [35]: **Arbitrage** Kombination eines Kauf- und Verkaufsaktes, der zeitliche oder lokale Preisunterschiede für $\rightarrow$ Güter, Edelmetalle, $\rightarrow$ Effekten, Guthaben von $\rightarrow$ Zentralbankgeld oder Zinssätzen ausnutzt. Der Arbitrageur kauft zum niedrigsten und verkauft zum höheren $\rightarrow$ Preis, wobei der Preisunterschied größer als die anfallenden Kosten sein muß. Da beim Arbitragegeschäft Marktkennnis gegeben ist, ist es ziemlich risikolos im Gegensatz zur $\rightarrow$ Spekulation. Durch Arbitrage wird ein $\rightarrow$ Gleichgewicht hergestellt.

oder in anderen Worten 1 : 13 983 816. Ein sehr illustrativer Vergleich für die Wahrscheinlichkeit, sechs Richtige im Lotto zu gewinnen, wird von Erhard Behrends angegeben [6]: Man stelle sich eine Strecke von 100 km Autobahn vor. Auf dieser Strecke ist ein Streifen von etwa 1 cm Breite quer zur Fahrtrichtung markiert. Wenn man diese Strecke abfährt mit einem Beifahrer, dessen Augen verbunden sind, und dieser Beifahrer dann irgendwo eine Münze aus dem Fenster wirft, dann ist die Wahrscheinlichkeit, den Streifen zu treffen etwa gleich groß wie die Wahrscheinlichkeit, sechs Richtige im Lotto zu gewinnen.

I Bringen wirklich Störche die Kinder?

Man kann zwei gegebene Funktionsverläufe durch die Korrelationsanalyse vergleichen. Hierbei wird die Wahrscheinlichkeit ermittelt, daß die beiden Verläufe „etwas miteinander zu tun haben“. Diese Methode wird häufig benutzt, man kann etwa untersuchen, inwieweit der Ölpreis vom Dollarkurs abhängt, welche Entwicklungen Einfluß auf Kursverläufe an der Börse haben oder ob gewisse hygienische Maßnahmen sich positiv auf das Erkrankungsrisiko auswirken.

Im Jahre 1988 veröffentlichte Helmut Sies in der hochwissenschaftlichen Zeitschrift *Nature* eine Notiz, die seitdem häufig zitiert wurde. In diesem Beitrag wies er darauf hin, daß in Westdeutschland die Geburtenrate auffällig gut korreliert ist mit der Anzahl der brütenden Storchpaare [30].

Literatur

- [1] Aristoteles. *Kategorien oder Lehre von den Grundbegriffen. Hermeneutica oder Lehre vom Urtheil*. Philosophische Bibliothek, Bd. 700. Erich Koschny, Leipzig, 1876. Übersetzt und erläutert von J. H. von Kirchmann.
- [2] Aristoteles. *Metaphysik*. Eugen Diederichs, Jena, 1907. Ins Deutsche übertragen von Adolf Lasson.
- [3] Heinz Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 1968.
- [4] Hans-Peter Beck-Bornholdt und Hans-Hermann Dubben. *Der Hund, der Eier legt. Erkennen von Fehlinformation durch Querdenken*. rororo Taschenbücher Nr. 61154. Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, Vollständig überarbeitete und erweiterte Neuauflage, Januar 2001.
- [5] Hans-Peter Beck-Bornholdt und Hans-Hermann Dubben. *Der Schein der Weisen. Irrtümer und Fehltritte im täglichen Denken*. rororo Taschenbücher Nr. 61450. Hoffmann & Campe, Hamburg, 2001.
- [6] Erhard Behrends. Ist Mathematik die Sprache der Natur? Vortrag auf der Herbsttagung 2009 der Mathematischen Gesellschaft Hamburg am 6. und 7. November 2009.
- [7] Moritz Cantor. *Vorlesungen über Geschichte der Mathematik. Zweiter Band: Von 1200 – 1668*. Verlag und Druck von B. G. Teubner, Leipzig, unveränderter Neudruck der zweiten Auflage, 1913.
- [8] Rudolf Carnap. *Einführung in die Philosophie der Naturwissenschaften*. Ullstein Materialien, Ullstein Buch Nr. 35243. Verlag Ullstein GmbH, Frankfurt/Main, Berlin, 1986.
- [9] Gregory J. Chaitin. Randomness and mathematical proof. *Scientific American*, 232(5):47–52, May 1975.
- [10] Gregory J. Chaitin. Algorithmic information theory. *IBM Journal of Research and Development*, 21:350–359, July 1977.
- [11] Gregory J. Chaitin. Grenzen der Berechenbarkeit. *Spektrum der Wissenschaft*, S. 86–93, Februar 2004.
- [12] Devra Lee Davis und H. Leon Bradlow. Verursachen Umwelt-Östrogene Brustkrebs? *Spektrum der Wissenschaft*, S. 38–44, Dezember 1995.
- [13] Devra Lee Davis, H. Leon Bradlow, Mary Wolff, Tracey Woodruff, David G. Hoel, and Hoda Anton-Culver. Medical hypothesis: Xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives*, 101(5):372–377, October 1993.
- [14] A. K. Dewdney. Der Trugschluss des Ermittlers. *Spektrum der Wissenschaft · Dossier 02/03: Mathematische Unterhaltungen II*, S. 6–8, 2003.
- [15] Alwin Diemer und Ivo Frenzel, Hrsg. *Das Fischer Lexikon Philosophie. Mit einer Einleitung von Professor Dr. Helmuth Plessner herausgegeben von Dr. Dr. Alwin Diemer und Ivo Frenzel*. Fischer Bücherei KG, Frankfurt am Main, 1958.

- [16] Ernst Eberlein. Mathematik und die Finanzkrise. *Spektrum der Wissenschaft*, S. 92–100, Dezember 2009.
- [17] Albert Einstein – Hedwig und Max Born. *Briefwechsel 1916–1955. Kommentiert von Max Born, Geleitwort von Bertrand Russell, Vorwort von Werner Heisenberg*. Ullstein-Buch; Nr. 34345: Ullstein-Sachbuch. Verlag Ullstein GmbH, Frankfurt am Main, Berlin, 1986.
- [18] W. Gellert, H. Küstner, M. Hellwich und H. Kästner, Hrsg. *Kleine Enzyklopädie Mathematik*. Verlag Enzyklopädie, Leipzig, 1967.
- [19] Christian Gerthsen, H. O. Kneser und Helmut Vogel. *Physik, ein Lehrbuch zum Gebrauch neben Vorlesungen*. Springer-Verlag, Berlin, Heidelberg, New York, Dreizehnte Auflage, neubearbeitet und erweitert von H. Vogel, 1977.
- [20] Kluge. *Etymologisches Wörterbuch der deutschen Sprache. Bearbeitet von Elmar Seebold*. Walter de Gruyter, Berlin · New York, 23., erweiterte Auflage, 1999.
- [21] Edwin Kube, Werner Deinet und Bernd Rieger. Zum Sachbeweis im deutschen Strafprozeß. *Spektrum der Wissenschaft*, S. 114–116, Juli 1990.
- [22] Karl Marx. *Das Kapital. Kritik der politischen Ökonomie. Erster Band (= Band 23 der Werke von Marx und Engels). Buch I: Der Produktionsprozeß des Kapitals. Nach der vierten, von Friedrich Engels durchgesehenen und herausgegebenen Auflage, Hamburg, Verlag von Otto Meissner, 1890*. Dietz Verlag, Berlin, 1962.
- [23] Karl von Meyenn, Hrsg. *Lust an der Erkenntnis. Triumph und Krise der Mechanik. Ein Lesebuch zur Geschichte der Physik*. Serie Piper, 1146. R. Piper Verlag, München, Zürich, Februar 1990.
- [24] Jürgen Mittelstraß. *Die Möglichkeit von Wissenschaft*. suhrkamp taschenbuch wissenschaft 62. Suhrkamp Verlag, Frankfurt am Main, 1974.
- [25] Jaques Monod. *Zufall und Notwendigkeit. Philosophische Fragen der modernen Biologie. Deutsch von Friedrich Giese, Vorrede zur deutschen Ausgabe von Manfred Eigen*. Deutscher Taschenbuch Verlag, München, 1975.
- [26] Peter J. Neufeld und Neville Colman. Wissenschaft im Zeugenstand. *Spektrum der Wissenschaft*, S. 106–114, Juli 1990.
- [27] John Allen Paulos. *Es war 1mal ... Die verborgene mathematische Logik des Alltäglichen, Aus dem Englischen übersetzt von Brigitte Post und Harald Höfer*. Spektrum Akademischer Verlag, Heidelberg · Berlin, 2000. Besprechung: *Spektrum der Wissenschaft*, September 2000:108.
- [28] Gero von Randow. *Das Ziegenproblem. Denken in Wahrscheinlichkeiten*. rororo 1090, Sachbuch 9337. Rowohlt Taschenbuch Verlag, Reinbek, 1992.
- [29] Gero von Randow. *Das Ziegenproblem. Denken in Wahrscheinlichkeiten*. rororo science, 61905. Rowohlt Taschenbuch Verlag, Reinbek, 2004.
- [30] Helmut Sies. A new parameter for sex education. *Nature*, 332:495, 7 April 1988.
- [31] Mark Twain. Chapters from my autobiography–XX. *North American Review*, CLXXXVI(618):465–474, July 5 1907.

- [32] Mark Twain. *Autobiographie, zusammengestellt von Charles Neider*. Fischer Bücherei 1021. Fischer Bücherei GmbH, Frankfurt am Main und Hamburg, 1969.
- [33] Walter Vogel. *Wahrscheinlichkeitstheorie*. Studia Mathematica / Mathematische Lehrbücher, Band XXII. Vandenhoeck & Ruprecht, Göttingen, 1970.
- [34] E. P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, 13:1–14, 1960.
- [35] Artur Woll, Hrsg. *Wirtschaftslexikon von A –Z. Betriebswirtschaft · Volkswirtschaft · Statistik · Rechtswissenschaft*. Oldenbourg Verlag, München, 10., vollständig neubearbeitete Auflage, 2009.
- [36] Erich Worbs. *Carl Friedrich Gauß. Ein Lebensbild*. Koehler & Amelang, Leipzig, Zweite, verbesserte Auflage, 1955.

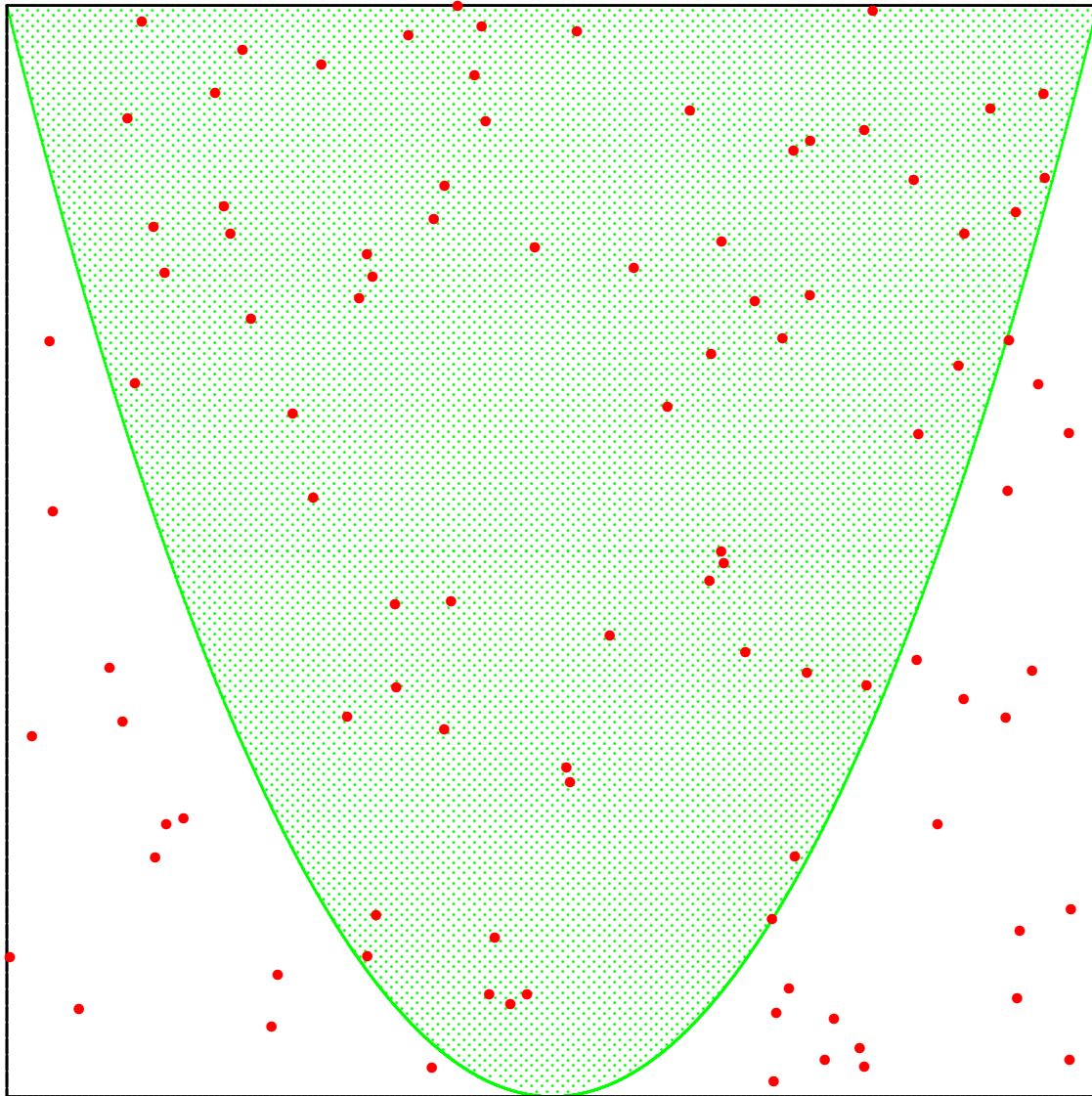


Abbildung 2: Berechnung einer Fläche (grün) vermittle zufällig gewählter Punkte (rot).

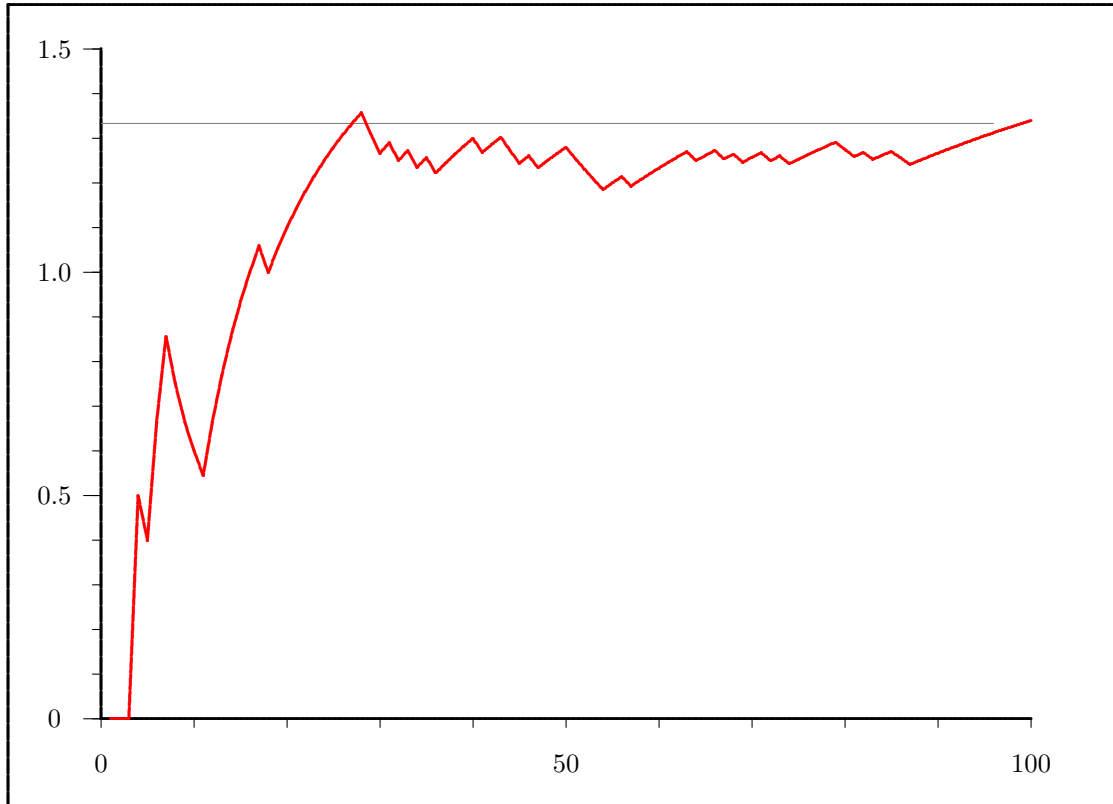


Abbildung 3: Berechnung eines Flächeninhalts durch ein Zufallsexperiment.

	1	2	3	4	5	6
1	•	•	•	•	•	•
2	•	•	•	•	•	•
3	•	•	•	•	•	•
4	•	•	•	•	•	•
5	•	•	•	•	•	•
6	•	•	•	•	•	•

Abbildung 4: Wurf zweier Würfel.

In der ersten Spalte ist die Augenzahl des ersten Würfels angegeben, in der Kopfzeile die Augenzahl des zweiten. • bezeichnet den Fall, daß bei zwei Würfeln keine 6 gewürfelt wurde, • den Fall, daß bei einer der Würfel eine 6 zeigt, der andere nicht, und • schließlich bezeichnet den Fall, daß beide Würfel 6 zeigen.

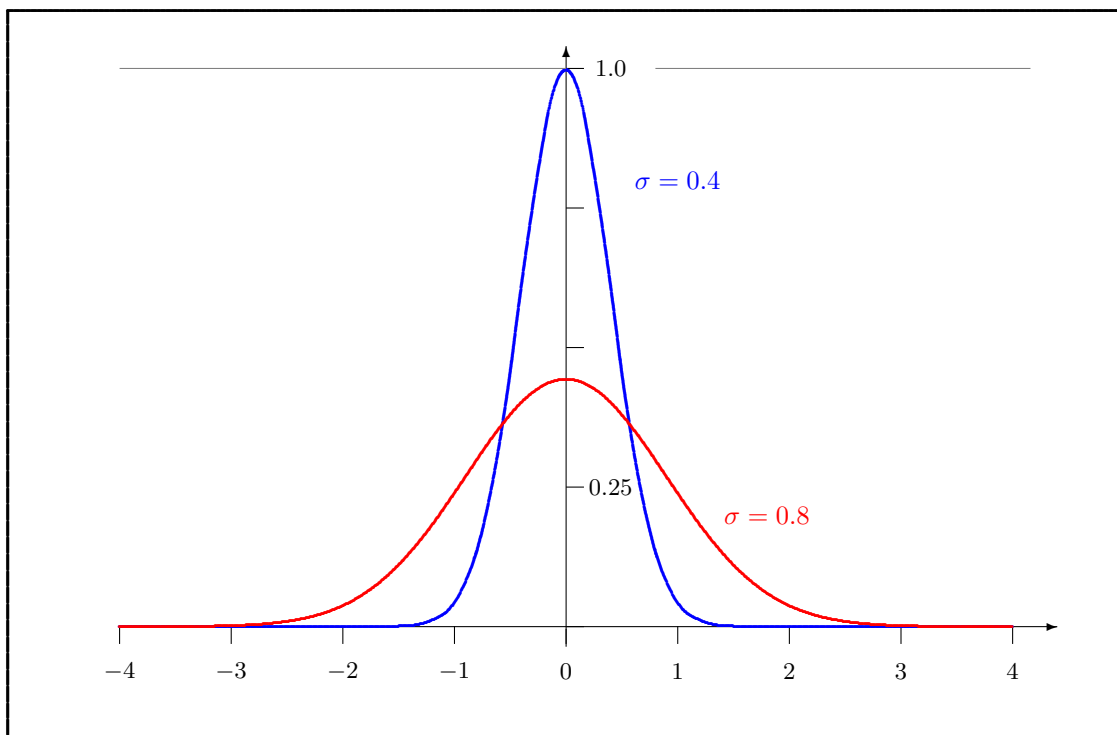


Abbildung 5: Die Gaußsche Fehlverteilungsfunktion für zwei verschiedene σ -Werte.

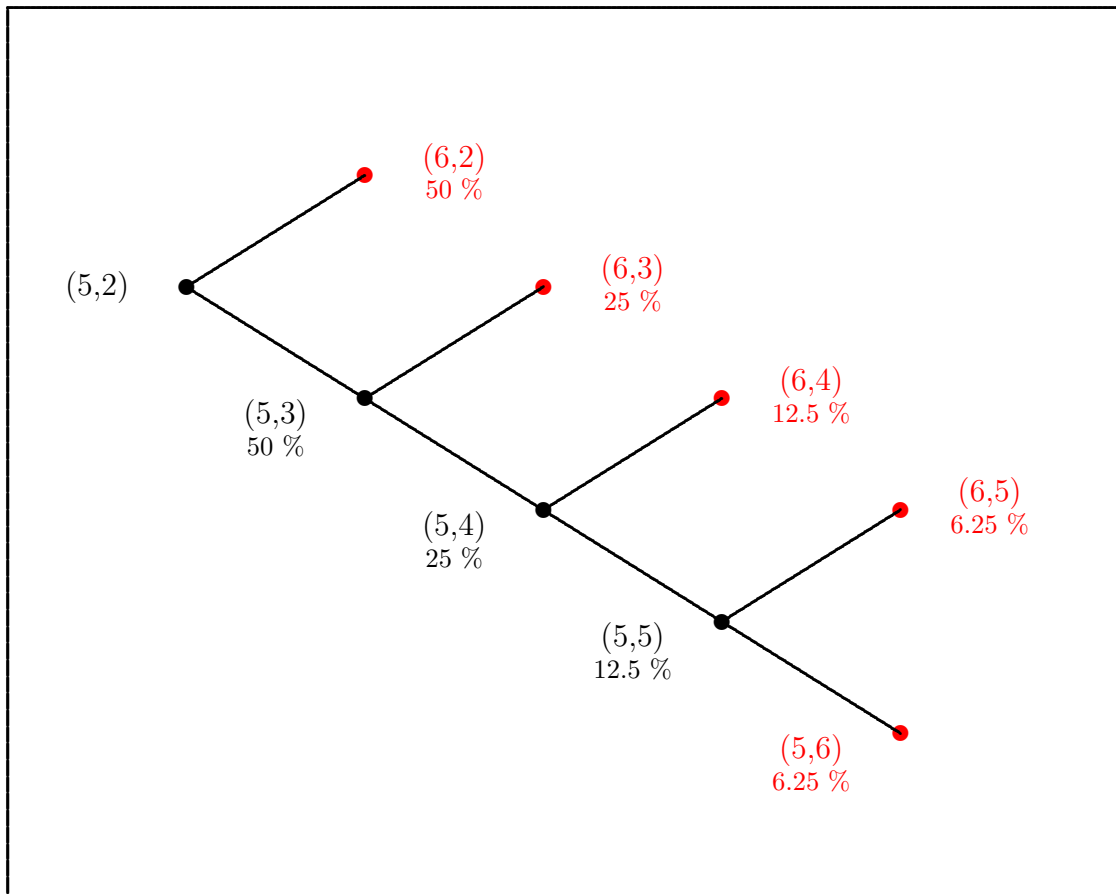


Abbildung 6: Die möglichen Spielausgänge.

An jeden Knoten ist der Punktestand in der Form (Spieler 1, Spieler 2) angegeben sowie die Wahrscheinlichkeit, mit der dieser Knoten erreicht wird. Gewinnknoten sind rot markiert.



Abbildung 7: Kursverlauf einer Aktie (ADIDAS SALOMON).

Dargestellt sind die Tageskurse vom 8. November 1996 bis zum 5. April 2002.
Die Markierungen auf der Abszisse entsprechen Intervallen von je 30 Tagen.