

Neue Ideen zur Stochastik

von
Dr. Wolfgang Riemer
Köln



Bibliographisches Institut Mannheim/Wien/Zürich
B. I.-Wissenschaftsverlag

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Riemer, Wolfgang:

Neue Ideen zur Stochastik / von Wolfgang Riemer.

- Mannheim; Wien; Zürich: Bibliographisches

Institut, 1985.

(Lehrbücher und Monographien zur Didaktik der
Mathematik; 3)

ISBN 3-411-03119-0

NE: GT

Alle Rechte, auch die der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieses Werkes darf ohne schriftliche Einwilligung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder ein anderes Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

© Bibliographisches Institut, Zürich 1985

Druck und Bindearbeit: Hain-Druck GmbH, Meisenheim/Glan

Printed in Germany

ISBN 3-411-03119-0

VORWORT

Dieses Buch behandelt die zentralen Themen der Stochastik in der Sekundarstufe:

- Wie kommt es zu den immer wieder beklagten Verständnisschwierigkeiten mit der Statistik? Sie scheinen nicht primär durch die "schwierige Materie", sondern durch einen unglücklichen Curriculaufbau verursacht, der die im Alltag gewachsenen intuitiven Strukturen vernachlässigt. In diesem Buch wird - unterstützt durch zahlreiche Beispiele - ein Rahmen für die Stochastik entwickelt, der diesen Verständnisschwierigkeiten langfristig entgegenwirkt.

Im Mittelpunkt stehen der subjektive Wahrscheinlichkeitsbegriff, die Bayessche Regel und das frühzeitige Abwägen zwischen verschiedenen Alternativhypothesen, nicht aber Mengenalgebra und Laplace-Wahrscheinlichkeiten. Die Erfahrung hat gezeigt, daß sich auf diese Weise die Schwierigkeiten beträchtlich reduzieren lassen.

- Im zweiten Teil wird der Leser mit einem Abakus bekanntgemacht, der gestattet, Wahrscheinlichkeitsverteilungen zu beliebigen Summenexperimenten spielerisch zu bestimmen. Durch eine intrinsisch motivierte Umkehrung gelingt es, Erwartungswert und Varianz problemorientiert einzuführen und deren Additivitätseigenschaften auf nicht formalem Darstellungsniveau herzuleiten. Es entfaltet sich die Möglichkeit, die gesamte Wahrscheinlichkeitsrechnung bis zum Gesetz der großen Zahlen enaktiv zu entwickeln. Gleichsam von selbst wächst ein intuitives Verständnis für den zentralen Grenzwertsatz.

- Überraschenderweise gelingt es, die genannten spielerischen Ansätze formal zu einem heuristischen Beweis des zentralen Grenzwertsatzes auszubauen. Es dürfte sich hier um die erste Elementarisierung handeln, die unterhalb des Hoch-

schulniveaus praktikabel ist, ohne sich auf den Spezialfall von De Moivre-Laplace zu beschränken.

- In weiteren Teilen werden physikalische Experimente zur Entwicklung und Veranschaulichung stochastischer Vorstellungen genutzt.

- Darüberhinaus soll das Vorgehen in diesem Buch unter lernpsychologischen Aspekten reflektiert werden. Dabei spielt der Begriff des generativen Modells eine wichtige Rolle.

Abschließend möchte ich Herrn Professor Dr. H. Scheid für die Anregung zu diesem Buch und dem Verlag Bibliographisches Institut für die Aufnahme in diese Reihe danken. Besonderer Dank gilt auch vielen meiner Schüler, die durch kreative und kritische Beiträge wesentlich zur Genese der vorliegenden Ideen beitrugen.

Köln, im Dezember 1984

Wolfgang Riemer

INHALT

0	Einleitung	
0.1	Inhaltsübersicht	11
0.2	Der umgekehrte Wahrscheinlichkeitsbaum	13
0.3	Zur "Grundkursproblematik"	14
1	Die Regel von Bayes und der subjektive Wahrscheinlichkeitsbegriff	
1.1	Problemstellung	15
1.2	Revision subjektiver Wahrscheinlichkeiten	20
1.3	Heuristische Begründung der Bayesschen Regel	28
1.4	Iterierte Anwendung der Bayesschen Regel	32
1.5	Die subjektiven Wahrscheinlichkeiten werden objektiv	39
1.6	Testen von Hypothesen	45
2	Primäre Intuitionen, sekundäre Intuitionen und heuristische Modelle aus der Sicht Fischbeins	
2.1	Primäre Intuitionen	51
2.2	Sekundäre Intuitionen	53
2.3	Modelle	55
2.4	Der umgekehrte Wahrscheinlichkeitsbaum	57
3	Der Varianzabakus	
3.1	Vorbemerkung	63
3.2	Der Faltungsabakus	66
3.3	Der Varianzabakus	70
3.4	Herleitung der Formel für die Varianz aus den Spielregeln des Abakus	80
3.5	Beweise zum Varianzabakus	81

3.6	Beweis der Additivität von Erwartungswert und Varianz über den Faltungsabakus und den Varianzabakus ..	84
3.7	Die Ungleichung von Tschebyscheff	88
4	Der elastische Stab	
4.1	Das Modell	91
4.2	Herleitung der Formel für die Varianz	94
4.3	Physikalischer Hintergrund	97
4.4	Die Ungleichung von Tschebyscheff	100
4.5	Curriculare Aspekte	101
5	Brownsche Molekularbewegung im elastischen Zentralkraftfeld, der Grenzwertsatz von de Moivre-Laplace	
5.1	Vorbemerkung	103
5.2	Simulation der Brownschen Bewegung im elastischen Zentralkraftfeld	105
5.3	Diskretes Modell der Brownschen Bewegung, die Binomialverteilung als stationäre Verteilung	108
5.4	Verfeinerung der Diskretisierung, Normalverteilung	111
5.5	Grenzwertsatz von de Moivre-Laplace, allgemeiner Fall	112
5.6	Der umgekehrte Wahrscheinlichkeitsbaum und Einführung der Binomialverteilung ohne Kombinatorik	113
6	Brownsche Molekularbewegung im konstanten Zentralkraftfeld	
6.1	Simulation der Brownschen Bewegung im konstanten Zentralkraftfeld	119
6.2	Diskretes Modell der Brownschen Bewegung	122
6.3	Verfeinerung der Diskretisierung, Übergang zur Exponentialverteilung	123
6.4	Anmerkung zur einseitigen Exponentialverteilung ...	124

7	Heuristische Begründung des zentralen Grenzwertsatzes	
7.1	Vorbemerkung	126
7.2	Der zentrale Grenzwertsatz	128
7.3	Herleitung der Differentialgleichung für die Grenz- verteilung im Spezialfall gleicher Summanden	131
7.4	Herleitung der Differentialgleichung im allgemeinen Fall nicht identischer Summanden	138
7.5	Die Lindebergbedingung	141
	Paradoxien zum Bayesschen Prinzip	143
	Literatur	148
	Index	155

0 Einleitung

Dieses Buch möchte anregen, über Stochastik und ihre Vermittlung im Unterricht nachzudenken. Es möchte versuchen, mit neuen Ideen auch altbekannte Sachverhalte in neuem Licht erscheinen zu lassen. Es möchte bewußt machen, wie man zentrale Inhalte der Stochastik bei geeigneter Darstellungsweise gemäß den Vorstellungen Bruners schon früher vorbereiten und oft treffender erarbeiten kann als allgemein üblich.

0.1 Inhaltsübersicht

Kapitel 1 ("Der subjektive Wahrscheinlichkeitsbegriff und die Bayessche Regel"): Die Schwierigkeiten beim Verständnis der Testtheorie im Anschluß an die Wahrscheinlichkeitsrechnung sind hinreichend bekannt. Sie lassen sich auch durch noch so überzeugende Erläuterungen im Unterricht nur schwer ausräumen. Die Ursache liegt darin begründet, daß wir die Wahrscheinlichkeitsrechnung und Statistik begrifflich nicht konsistent behandeln:

Der auf den **unteren Curriculumstufen** und in der Alltagserfahrung entwickelte subjektive Wahrscheinlichkeitsbegriff wird auf den **mittleren Curriculumstufen** ersetzt durch den Laplaceschen, der hier zusammen mit Kombinatorik und Mengen algebra das Bild der Wahrscheinlichkeitsrechnung prägt.

Für ein Verständnis der Testtheorie auf den **oberen Stufen** des Curriculums braucht man aber den Begriff der Hypothese und eine intuitive Vertrautheit mit dem Gedanken, daß Wahrscheinlichkeiten nur Annahmen darstellen, die nicht selten auf Grund von Indizien revidiert werden müssen.

Der Laplacesche Wahrscheinlichkeitsbegriff der mittleren Stufen läuft aber gerade wegen seiner griffigen Definition der Entwicklung dieser Intuition entgegen.

In Kapitel 1 soll gezeigt werden, wie man den auf unteren Curriculumstufen gepflegten subjektiven Wahrscheinlichkeitsbegriff durch geeigneten Einsatz der Bayesschen Regel

weiterentwickeln und zu dem später in der Statistik benötigten hypothetischen ausbauen kann, um so die genannten Diskontinuitäten zu überwinden, die den Prinzipien von Fortsetzbarkeit und spiralischem Curriculumaufbau widersprechen.

In Kapitel 2 soll das Vorgehen aus Kapitel 1, aber auch der folgenden Kapitel, aus dem Blickwinkel lernpsychologischer und didaktischer Prinzipien reflektiert werden.

Kapitel 3 ("Der Varianzabakus"): Warum benutzt man als Maß für die Streuung einer Verteilung die Varianz? Im allgemeinen bleibt man eine ehrliche Antwort auf diese Frage schuldig. Wir stellen Wahrscheinlichkeitsverteilungen durch Chips dar und ermitteln "spielerisch" Verteilungen zu Summen beliebiger Zufallsgrößen (Faltungen). Durch Umkehren des Spiels entsteht ein Abakus, der zu beliebigen Verteilungen ein Streuungsmaß (die Varianz und simultan den Erwartungswert) liefert. Die Formel für die Varianz leitet sich aus dem Abakus her. Auch das Gesetz der großen Zahlen läßt sich entnehmen. Man erhält eine enaktive (handlungsbezogene), also nicht formale Darstellung der Wahrscheinlichkeitsrechnung.

Kapitel 4 ("Der elastische Stab") beschreibt einen heuristischen Zugang zur Varianz über ein physikalisches Modell: Befestigt man Gewichte an einem elastischen Waagebalken, so hängt dessen Durchbiegungswinkel proportional von der Belastung und quadratisch vom Hebelarm ab. Diese Tatsache läßt sich zu einer beziehungshaltigen Einführung quadratischer Parabeln in der Sekundarstufe I nutzen. Für die Stochastik wird das Modell dadurch interessant, daß sich bei Belastung mit mehreren Gewichten die bekannte Analogie Gewichtsverteilung - Wahrscheinlichkeitsverteilung und Schwerpunkt - Erwartungswert zur Analogie Durchbiegungswinkel - Varianz erweitern läßt. Letztere ermöglicht ein enaktives Erarbeiten und experimentelles Überprüfen der Tschebyscheff-Ungleichung.

Die Kapitel 5 bis 7 weisen neue Wege, auf denen man das "schwierige" und didaktisch bisher wenig reflektierte Gebiet der stochastischen Grenzwertsätze (Zentraler Grenzwertsatz, Grenzwertsatz von de Moivre-Laplace) auf elementarem Niveau heuristisch angehen kann. Dabei werden einerseits Methoden früherer Curriculumstufen problemorientiert weiterentwickelt, andererseits wichtige Ideen der Hochschulmathematik vorbereitet.

0.2 Der umgekehrte Wahrscheinlichkeitsbaum

Nach diesem Inhaltsüberblick sei auf einen Zusammenhang zwischen den einzelnen Teilen dieses Buches hingewiesen. Er wird hergestellt durch das generative Modell des umgekehrten Wahrscheinlichkeitsbaumes, der trotz der weitgespannten Thematik in fast allen Kapiteln auftaucht und bei konsequentem Einsatz ein tragfähiges und nützliches Hilfsmittel darstellt:

Erstens ermöglicht er durch geeignete Darstellung der Bayeschen Regel einen "subjektivistischen Aufbau" der Stochastik, in dem weder Mengenalgebra noch deskriptive Statistik vorherrschen. Stattdessen werden die Lernenden auf der Grundlage eigener Primärerfahrungen von Anfang an mit der fundamentalen Idee der Statistik, nämlich dem Prinzip des induktiven Schließens vertraut.

Er ermöglicht **zweitens** die rekursive Bestimmung von Wahrscheinlichkeitsverteilungen zu Summenexperimenten ohne den Formalismus von Produkträumen und Zufallsvariablen, dadurch

- einen frühen problemorientierten Zugang zum Themenkreis um die Varianz (Varianzabakus) bis hin zum schwachen Gesetz der großen Zahlen, außerdem
- eine anschauliche Vorbereitung des zentralen Grenzwertsatzes und schließlich den Ansatz zu seiner heuristischen Begründung.

Drittens veranschaulicht er die rekursive Berechnung von Markoffketten ohne Matrizenalgebra, so daß schon auf mittleren Curriculumstufen die strukturellen Gemeinsamkeiten

etwa mit der Binomialverteilung (allgemeiner Faltungen) prägnant sichtbar werden.

0.3 Zur "Grundkursproblematik"

Wie wir wissen, ist das Unterrichten in Grundkursen der gymnasialen Oberstufe zu einem ernsthaften Problem geworden. Im Hinblick darauf sei betont, daß sich der curriculare Rahmen für die Stochastik aus Kapitel 1 in der Praxis sehr gut bewährt hat. Durch Einsatz eines geeigneten Darstellungsmodus wird es nämlich auch unter ungünstigen Umständen möglich, einen problemorientierten Unterricht durchzuführen, in dem Schüler vor dem Aufbau eines umfangreichen, oft nicht ganz einfach durchschaubaren Formalismus zentrale Ideen operativ erfassen und den Bezug zu ihren vorbewußten Primärerfahrungen erkennen können.

1 Die Regel von Bayes und der subjektive Wahrscheinlichkeitsbegriff

(Ein Einstieg in die Stochastik über den subjektiven Wahrscheinlichkeitsbegriff und die Bayessche Regel als Bindeglied zur Statistik)

1.1 Problemstellung

In der Test- und Schätztheorie treten bei Schülern häufig schwer behebbarere Verständnisschwierigkeiten auf. Das folgende oft gehörte Zitat: "Wir haben die **Hypothese H_0 ($p=0,8$)** auf dem 5% Signifikanzniveau verworfen, also gilt die Alternative H_1 mit 95%iger Wahrscheinlichkeit" belegt eindrucksvoll, wie gründlich die Aussagen der Testtheorie mißverstanden werden. Nehmen wir eine weitere Aussage hinzu, die sich zur Schätztheorie in vielen Büchern findet: "In einer Stichprobe vom Umfang 20 hat ein Ausfall die relative Häufigkeit $h=0,6$. Dann liegt die **Wahrscheinlichkeit p** für diesen Ausfall mit der Wahrscheinlichkeit 80% im Konfidenzintervall $[0,45;0,72]$ ". Während das erste Zitat offensichtlich falsch ist, wird die zweite Aussage häufig falsch interpretiert (vgl. [Schmetterer, S.309]).

Gemeinsam ist den Zitaten, daß einerseits von Wahrscheinlichkeitshypothesen (fett gedruckt), andererseits von subjektiven Wahrscheinlichkeiten (unterstrichen) gesprochen wird, mit denen diese Hypothesen gelten sollen.

Die Schwierigkeiten beim Verständnis statistischer Aussagen rühren nun daher, daß wir

- während der ganzen unteren und mittleren Curriculumstufen nie von Hypothesen sprechen, sondern Wahrscheinlichkeiten stets genau berechnen,
- die so wichtigen subjektiven Wahrscheinlichkeiten nicht thematisieren, mit denen Hypothesen gelten - obwohl uns gerade diese Wahrscheinlichkeiten im täglichen Leben unablässig begleiten.

Hier schließen wir nämlich induktiv, indem wir unsere subjektiven Wahrscheinlichkeiten durch Beobachtungen ständig verändern. Wir erläutern das etwas genauer: Um Beobachtungen zu erklären, stehen uns oft mehrere Hypothesen zur Auswahl. Einige halten wir für wahrscheinlicher, andere für weniger wahrscheinlich. Zur Klärung besorgen wir uns Informationen (Indizien), die meist in verschieden starker Weise für oder gegen die einzelnen Hypothesen sprechen. Möglicherweise werden dadurch zunächst unwahrscheinlich erscheinende Hypothesen wahrscheinlicher oder wahrscheinliche nahezu sicher. Jedenfalls werden wir auf Grund der Informationen unserer subjektiven Wahrscheinlichkeiten für die einzelnen Hypothesen überdenken und möglicherweise revidieren müssen.

Die Bayessche Regel ermöglicht es, diesen Revisionsprozeß zu verstehen und quantitativ zu beschreiben.

Leiten wir die Bayessche Regel nach den Vorschlägen des derzeit praktizierten Curriculums deduktiv aus Axiomen her, so geht diese weitreichende Bedeutung und der Alltagsbezug leider meist verloren. Dafür ist aber auch die in der Einleitung erwähnte Inkonsistenz verantwortlich, die durch die starke Überbetonung von Kombinatorik und Laplace-Wahrscheinlichkeiten verursacht wird.

Sie verhindert, daß sich der in Alltagserfahrung und Primarstufe gepflegte subjektive Wahrscheinlichkeitsbegriff (vgl. [Heitele 2, S. 219]) weiterentwickelt zu einem hypothetischen, auf dessen Grundlage allein ein tieferes Verständnis der induktiven Bestandteile der Stochastik gedeihen kann. Dinges schreibt:

"Der objektivistische Wahrscheinlichkeitsbegriff, der sich an Urne und Glücksrad orientiert, scheint gut geeignet, am Anfang des Stochastikunterrichts zu stehen. Er hat aber Grenzen. Die Bayessche Formel paßt nicht dazu; sie braucht den subjektiven Wahrscheinlichkeitsbegriff. Die höhere Wahrscheinlichkeitstheorie braucht den Begriff der Hypothese." [Dinges, S. 121].

Daß auch der frequentistische Wahrscheinlichkeitsbegriff, der oft zusammen mit dem Laplaceschen verwendet wird, keinen Ausweg aus diesem Dilemma weist, kommt in der Arbeit von Schrage über "Schwierigkeiten mit stochastischer Modellbildung" zum Ausdruck:

"Die meisten Aussagen über Wahrscheinlichkeiten im täglichen Leben und Konsequenzen, die aus Wahrscheinlichkeiten für irgendwelche Ereignisse gezogen werden, beziehen sich auf subjektive Beurteilungen. Ebenso stützt sich das kindliche Vorverständnis, wie z.B. die Untersuchungen von [Cohen und Hansel] zeigen, auf einen subjektiven Wahrscheinlichkeitsbegriff. Die meisten Unterrichtskurse favorisieren demgegenüber einseitig einen frequentistischen Wahrscheinlichkeitsbegriff, stellen ihn eventuell als einzig möglichen hin." [Schrage, S. 99].

Shaughnessey charakterisiert die Situation: "Prior to any formal course work in probability students have had experience in and have dealt exclusively with 'subjective probability'. Suddenly students are confronted in a formal course with a completely mathematized model of what Carnap calls 'statistical probability'." [Shaughnessey, S. 298].

Mit Blick auf die Bayessche Regel, deren Fehlanwendungen er in seiner Arbeit analysiert, fährt Schrage fort: "Daß es dann zu Schwierigkeiten kommt, wenn Situationen zu mathematisieren sind, in denen es durchaus vernünftig ist, von Wahrscheinlichkeiten zu sprechen, in denen aber eine frequentistische Wahrscheinlichkeitsinterpretation keinen Sinn ergibt, und daß in solchen Fällen die Korrespondenz zwischen mathematischem Modell und der konkreten Problemstellung nicht mehr richtig gesehen wird, ist nicht zu verwundern." [Schrage, S. 100].

Anschließend verweist Schrage auf Dingess, der die Einleitung zu seiner Arbeit "Schwierigkeiten mit der Bayesschen Regel" durch die Worte beschließt:

"Ich hoffe, daß irgendwann Darstellungen der Bayesschen Regel gefunden werden, die den recht verstandenen Ansprüchen der Schule und der Lehrerausbildung genügen. Die vorliegende Arbeit will Anregungen liefern." [Dinges, S. 115].

Wir wollen die Anregungen aufgreifen und als Lösung der Probleme einen Einstieg in die Stochastik über den subjektiven Wahrscheinlichkeitsbegriff skizzieren, der seine Fortsetzung in einer problemorientierten Behandlung der Bayesschen Regel findet und dabei die Testtheorie ausgezeichnet vorbereitet. Der Gedankengang läßt sich wie folgt zusammenfassen:

Zu 1.2: Wie in dem Zitat von Schrage zum Ausdruck gebracht, tragen Schüler - wie alle Menschen - sehr viele subjektive Wahrscheinlichkeitsverteilungen in sich. Sie sind spontan bereit, diese zu äußern, zu diskutieren und möglicherweise zu revidieren. Dieses wertvolle Potential von Vorerfahrungen soll nicht durch eine Beschränkung auf objektive Wahrscheinlichkeiten zerstört, sondern durch aktives Schätzen und intuitives Revidieren subjektiver Wahrscheinlichkeitsverteilungen bewußt gemacht werden. Durch diesen Ansatz kommt von Anfang an der zentrale Gedanke der Statistik zum Tragen, der auch die folgenden Ausführungen wie ein "roter Faden" durchzieht:

Die genauen Wahrscheinlichkeiten werden stets verborgen bleiben, man kann sie höchstens so ansetzen, daß sie den experimentellen Ergebnissen nicht widersprechen.

Zu 1.3 und 1.4: Die Frage, wie der Mensch seine vielen subjektiven Wahrscheinlichkeitsverteilungen revidiert, wenn die Situationen nicht oder nur selten reproduzierbar sind, also relative Häufigkeiten nicht sinnvoll gebildet werden können, führt zur Bayesschen Regel. Aus lernpsychologischen Gründen soll sie durch eine heuristische Begründung, die bei jeder Anwendung neu vollzogen werden kann, den Charakter einer Sekundärintuition erhalten. Wie schon erwähnt, ist dazu mengentheoretischer Formalismus eher hinderlich. Es

reichen Pfad- und Additionsregel, die wir im generativen Modell des umgekehrten Wahrscheinlichkeitsbaumes zusammenfassen.

Zu 1.5: Wir haben zwei Methoden zur Revision subjektiver Wahrscheinlichkeiten kennengelernt, nämlich die Angleichung der subjektiven Wahrscheinlichkeiten an die relativen Häufigkeiten bei oft reproduzierbaren Situationen (1.2) und die Bayessche Methode (1.3, 1.4) bei selten reproduzierbaren Situationen. Wie diese Methoden zusammenhängen, soll eine Computersimulation zeigen, die auch vom begrifflichen Hintergrund überaus lohnend ist.

1.6 zeigt, mit welcher strukturellen Kontinuität sich das Testen von Hypothesen anschließt. Von zentraler Bedeutung ist, daß jetzt die Beziehungen zwischen Signifikanzniveau und Irrtumswahrscheinlichkeiten deutlich werden.

Vielleicht erkennt man schon an diesem Überblick, wie unser Rahmen für einen genetischen Aufbau der Stochastik die Brunerschen Prinzipien von Präfiguration, Fortsetzbarkeit und Strukturorientierung realisiert. Daß darüberhinaus auch wichtige lernpsychologische Prinzipien verwirklicht werden, soll zusammen mit der Theorie von Fischbein in Kapitel 2 erläutert werden.

1.2 Revision subjektiver Wahrscheinlichkeiten

Wie groß ist die Wahrscheinlichkeit, daß beim Werfen einer Münze Kopf ($K \hat{=} 1$) oder Zahl ($Z \hat{=} 0$) fällt? Jeder Mensch besitzt die gleiche subjektive Wahrscheinlichkeitsverteilung

ω	0	1
$p(\omega)$	0,5	0,5

und man erntet im allgemeinen nur mitleidiges Lächeln, wenn man diese Selbstverständlichkeit durch 1000 Münzwürfe bestätigen läßt. (Eine interessantere Fragestellung ist die Untersuchung von Schwankungen der relativen Häufigkeiten bei wachsender Versuchszahl, vgl. [Freudenthal 1] .)

1.2.1 Beispiel 1

Bei der entsprechenden Frage für die Anzahl der Köpfe in einem Vierfachmünzwurf gehen die Meinungen jedoch erheblich auseinander. Man notiert die verschiedenen subjektiven Wahrscheinlichkeitsverteilungen:¹⁾

ω	0	1	2	3	4
$P_{\text{Frank}}(\omega)$	0,2	0,2	0,2	0,2	0,2
$P_{\text{Jeanette}}(\omega)$	0,1	0,25	0,3	0,25	0,1
$P_{\text{Arndt}}(\omega)$	0,05	0,2	0,5	0,2	0,05

Wer hat recht? Jetzt dient das Experiment der Lösung eines echten Konfliktes. Das Zusammentragen der Versuchsergebnisse

1) Die Gleichverteilung wird meist als erste vorgeschlagen, doch werden nach kurzem Überlegen Zweifel daran wach. Daß die Wahrscheinlichkeiten zwischen 0 und 1 liegen und ihre Summe 1 ergibt, gehört zu dem in der Alltagserfahrung gewonnenen Fundament, das die Schüler ab einem Alter von 12 Jahren in den Unterricht mitbringen. Vgl. dazu die Untersuchung [Fischbein/Barbat/Minzat]. Die numerischen Ergebnisse dieses Kapitels stammen aus Unterrichtsversuchen.

mehrerer Schüler liefert die folgende relative Häufigkeitsverteilung (hier für 200 Versuche):

ω	0	1	2	3	4
$r_{200}(\omega)$	0,055	0,245	0,405	0,235	0,06

und das Experiment zwingt die Schüler zur Revision ihrer subjektiven Wahrscheinlichkeitsverteilungen.

Nach kurzer Diskussion einigt man sich auf:

ω	0	1	2	3	4
$P_{\text{rev}}(\omega)$	0,06	0,24	0,4	0,24	0,06

Offensichtlich werden uns aber weitere Experimente zu erneuten Revisionen zwingen; mit anderen Worten: **Die genauen Wahrscheinlichkeiten werden stets verborgen bleiben. Man kann sie höchstens so ansetzen (d.h. Hypothesen so aufstellen), daß sie den experimentellen Ergebnissen nicht widersprechen.**

Beachtenswert ist in diesem Zusammenhang, daß Schüler auch bei weniger symmetrischen experimentellen Ergebnissen ihre Annahmen $p(\omega)=p(4-\omega)$ stets beibehalten. Das belegt, daß nicht "allzustarke" Abweichungen relativer Häufigkeiten von den Hypothesen von vornherein akzeptiert werden.

Statistisches Denken ist angelegt, man muß es nur bewußt machen.

Die Tatsache, daß in diesem Beispiel aus der Laplace-Hypothese für eine Münze die Wahrscheinlichkeiten für den Vierfachmünzwurf (Binomialverteilung) direkt mit dem Wahrscheinlichkeitsbaum oder rekursiv mit dem umgekehrten Wahrscheinlichkeitsbaum (vgl. 2.4) bestimmt werden können, erscheint jetzt als "Clou":

ω	0	1	2	3	4
$B_{4;1/2}(\omega)$	0,0625	0,25	0,375	0,25	0,0625

In diesem Zusammenhang kann leicht bewußt gemacht werden, daß die zu Grunde liegende Laplace - Hypothese auch nur eine plausible subjektive Annahme darstellt, die man möglicherweise zu revidieren hat. Das belegt etwa folgende fruchtbare Variation des Experimentes:

1.2.2 Beispiel 2

Statt der Münzwürfe machen wir Vierfachzüge aus einer Urne mit roten ($\hat{=1}$) und weißen ($\hat{=}0$) Kugeln von unbekannter Zusammensetzung. Drei Schülergruppen erhielten folgende relative Häufigkeitsverteilungen:

ω	0	1	2	3	4
$r_{100}(\omega)$	0,2	0,35	0,30	0,15	0,0
$r_{100}(\omega)$	0,17	0,48	0,17	0,12	0,06
$r_{100}(\omega)$	0,21	0,38	0,34	0,07	0,0
$r_{300}(\omega)$	0,1933	0,4033	0,27	0,1133	0,02

Man entschließt sich, die Symmetriehypothese aufzugeben und einigt sich auf folgende subjektive Wahrscheinlichkeitsverteilung:

ω	0	1	2	3	4
$p(\omega)$	0,19	0,4	0,27	0,12	0,02

Doch sind davon nicht allzusehr abweichende Verteilungen ebenso akzeptabel (unter anderem auch die relative Häufigkeitsverteilung r_{300} selber).

Lediglich die Tatsache scheint festzustehen, daß $\omega=1$ am wahrscheinlichsten ist, gefolgt von $\omega=2$, $\omega=0$, $\omega=3$. Am unwahrscheinlichsten ist $\omega=4$.

Die Urne enthält also sicher mehr weiße als rote Kugeln. Frank tippt auf 1/3 rote, Manfred auf 1/4 (Hypothesen). Unter Benutzung seiner Hypothese konstruieren wir für Frank mittels Wahrscheinlichkeitsbaum oder Faltung eine neue subjektive Wahrscheinlichkeitsverteilung (Binomialverteilung):

Frank: ω	0	1	2	3	4
$B_{4,1/3}(\omega)$	0,1975	0,3951	0,2963	0,0988	0,0123

analog für Manfred:

Manfred: ω	0	1	2	3	4
$B_{4,1/4}(\omega)$	0,3164	0,4219	0,2109	0,0469	0,0039

Manfreds Verteilung widerspricht der eben getroffenen Feststellung über die Rangordnung der einzelnen Wahrscheinlichkeiten, sie wird von den Klassenkameraden verworfen. Die Verteilung von Frank ist dagegen mit den experimentellen Ergebnissen recht gut vereinbar. Wir akzeptieren sie.

Erfahrungsgemäß kommen einige Schüler auf die Idee, die 300 Vierfachzüge als 1200 Einfachzüge zu deuten, bei denen man insgesamt 409 rote Kugeln zog. Das entspricht einer relativen Häufigkeit von $r_{1200}=0,3408$ für rot und die Hypothese von Frank ($p=1/3$) wird mit noch besserem Gewissen akzeptiert. Wer könnte der Versuchung widerstehen, in die Urne hineinzuschauen?

Es folgen zwei lehrreiche Beispiele, bei denen sich keine verborgenen Laplace - Hypothesen aufspüren lassen; sie eignen sich für die ersten Einführungsstunden in die Wahrscheinlichkeitsrechnung und die Überleitung zur Bayesschen Regel gleichermaßen.

1.2.3 Beispiel 3 (Zinkenwürfel)

Man sägt aus Profilleisten "kurze und lange Zinkenwürfel" gemäß der folgenden Abbildung. Nur aus der Geometrie der Objekte leiten sich vor jeglichem Experiment Vermutungen über das Verhalten der Würfel her:

kurz							lang					
	1	2	3	4	5	6	1	2	3	4	5	6
Ord	d	c	b	a	c	d	c	d	b	a	d	c
P _{Ingo}	,04	,16	,25	,35	,16	,04	,10	,05	,20	,50	,05	,10
P _{Nele}	,03	,08	,35	,43	,08	,03	,06	,05	,34	,44	,05	,06
r ₂₅₀	,056	,14	,228	,336	,20	,04	,092	,044	,236	,42	,08	,128
P _{rev}	,045	,17	,23	,34	,17	,045	,11	,065	,24	,42	,065	,11

Ord ist eine Hypothese über die Reihenfolge der relativen Häufigkeiten, die von fast allen Schülern vertreten wird: Bei dem langen Würfel sind die Stirnseiten 2 und 5 unwahrscheinlicher als die Seiten 1 und 6. Bei dem kurzen Würfel ist es umgekehrt. Die größte und schwerste Seite 3 wird meist unten liegen, also ist 4 bei beiden Würfeln an wahrscheinlichsten. Die folgenden zwei Zeilen enthalten die subjektiven Wahrscheinlichkeiten von Ingo und Nele. Man beachte wieder die partielle Symmetrie in allen Hypothesen! Die weit weniger symmetrischen relativen Häufigkeiten r_{250} zwingen uns, Neles Hypothese zu verwerfen, Ingos können wir akzeptieren oder unter Verwendung unserer experimentellen Ergebnisse revidieren (p_{rev}).

Das Erstellen und Revidieren subjektiver Wahrscheinlichkeitsverteilungen ist nicht Selbstzweck, wir nutzen sie von Anfang an zur Vorbereitung entscheidungstheoretischer Grundgedanken:

Schüler, deren experimentelle Ergebnisse in obiger Tabelle noch nicht verwendet wurden, teilen uns diese mit. Wir haben zu entscheiden, ob es sich um den langen oder den kurzen Zinkenwürfel gehandelt hat. Fehlentscheidungen werden uns mitgeteilt:

abs. Häufigkeit für						gewürfelt von	unsere Verm.	Sicherheit der Verm.	Verm. war
1	2	3	4	5	6				
3	9	9	16	9	4	Heidi	kurz	1450:1	richtig
2	3	15	24	3	3	Achim	lang	82:1	richtig
0	14	26	4	6	0	Veronika	kurz	31Mio:1	richtig
9	3	19	14	3	2	Jörg	lang	2516:1	falsch
2	7	10	15	14	2	Oliver	kurz	450000:1	richtig

Die Spalte "Sicherheit" wurde in Vorgriff auf 1.4 aufgenommen. Man berechnet sie mit der Bayesschen Regel. Obwohl man sie im Unterricht an dieser Stelle noch nicht erwähnen wird, kann man sich bei Jörg wohl berechtigt fragen, ob unsere Vermutung wirklich falsch war, oder ob Jörg geschummelt hat. 1 und 6 traten nämlich sehr viel häufiger auf als 2 und 5, was für den langen Würfel charakteristisch ist.

1.2.4 Beispiel 4 (Münzen mit Rand)

Man sägt kleine Rundhölzer, die z.T. durch Ankleben von Metallplättchen "gezinkt" werden. Folgende Tabelle zeigt in Analogie zu Beispiel 3 subjektive Wahrscheinlichkeitsverteilungen und ihre Revisionen durch das Experiment:

($Z \hat{=}$ Zahl, $B \hat{=}$ Baum, $R \hat{=}$ Rand)

	Rundholz, ungezinkt			Rundholz, gezinkt		
	Z	B	R	Z	B	R
P_{Mark}	0,25	0,25	0,50	0,50	0,10	0,40
P_{Elke}	0,40	0,40	0,20	0,50	0,00	0,50
P_{Dieter}	0,333	0,333	0,333	0,50	0,05	0,45
r_{1500}	0,219	0,237	0,454	0,517	0,073	0,41
P_{rev}	0,225	0,225	0,55	0,52	0,07	0,41

Die Hypothesen von Mark sagten die experimentellen Ergebnisse wesentlich besser voraus als die von Elke und Dieter, welche wir verwerfen müssen. Noch größeres Vertrauen haben wir jedoch in die revidierte Verteilung p_{rev} . Auch nach weiteren Experimenten sehen wir keine Veranlassung, p_{rev} zu verwerfen.

Wie in Beispiel 3 schließen sich Versuche an, bei denen man auf Grundlage der letztgenannten Verteilungen zu entscheiden hat, ob ein unbekanntes Objekt eine echte Münze, ein gezinktes oder ungezinktes Rundholz ist.

In 1.4 und 1.5 werden wir die Beispiele mit Hilfe der Bayes'schen Regel vertiefend behandeln, doch dürfte schon hier deutlich geworden sein, daß der subjektive Wahrscheinlichkeitsbegriff tragfähig ist, und wie seine Verwendung statistisches Denken vorbereitet. Man kann für die hier vorliegenden reproduzierbaren Situationen den Unterschied zum frequentistischen Wahrscheinlichkeitsbegriff wie folgt charakterisieren:

- Wahrscheinlichkeiten werden nicht mehr als objektiv feststehende Grenzwerte relativer Häufigkeiten aufgefaßt, sondern als subjektiv "beste Schätzung (Voraussage) für die zu erwartende relative Häufigkeit in einer langen Versuchsreihe". [Scheid 2, S. 32].
- Dadurch erhalten die Wahrscheinlichkeiten den Charakter von Hypothesen.
- Manche Hypothesen halten wir für wahrscheinlicher als andere, manche müssen wir ganz verwerfen.
- Wahrscheinlichkeiten sind von relativen Häufigkeiten prinzipiell verschieden: Man erkennt das daran, daß alle Hypothesen in den Beispielen 1, 3 und 4 Symmetrien aufweisen, für die relativen Häufigkeiten wird das nur in Ausnahmefällen gelten.

Für die Entwicklung eines subjektiv- hypothetischen Wahrscheinlichkeitsbegriffs ist die oft praktizierte Beschränkung auf klassische Zufallsobjekte wie Münze, Würfel, Glücksrad und Reißnagel fatal, denn

- entweder endet man wegen totaler Symmetrie sofort bei Laplace, es **verschwindet der Gedanke, daß es verschiedene gleichberechtigte Hypothesen gibt,**
- oder man endet wegen vollständiger Asymmetrie bei der Definition von Wahrscheinlichkeiten als Grenzwert relativer Häufigkeiten. **Es verschwindet der Gedanke, daß Wahrscheinlichkeiten Vorhersagen machen und von relativen Häufigkeiten prinzipiell verschieden sind.**

Der Rückgriff auf teilweise symmetrische Objekte in den Beispielen 1.2.3 und 1.2.4 ist also wohlüberlegt. Daß sich der Laplacesche Wahrscheinlichkeitsbegriff als Spezialfall des subjektiv - hypothetischen einordnen läßt, ist klar. Auch das Arbeiten mit Wahrscheinlichkeitsbäumen (Additionsregel und Produktregel) ist sinnvoll, aber nicht als Berechnung objektiver Wahrscheinlichkeiten, sondern als Methode, wie man aus einfachen subjektiven Wahrscheinlichkeitsannahmen (Hypothesen) (vgl. Frank $p=1/3$, Manfred $p=1/4$) neue, meist kompliziertere subjektive Verteilungen herleiten kann, an denen man interessiert ist. Ein großer Teil des im bisherigen Curriculum eingesetzten Materials ist also auch im Rahmen des subjektivistischen Ansatzes verwendbar.

Der "Blickpunktwechsel" ist jedoch im Bezug auf Bruners Prinzipien von Präfiguration und Fortsetzbarkeit von entscheidender Bedeutung.

1.3 Heuristische Begründung der Bayesschen Regel

Der in 1.2 geschilderte Ansatz hat gezeigt, daß subjektive Wahrscheinlichkeiten durch Erfahrung ständig revidiert werden müssen. In oft reproduzierbaren Situationen kann dies durch Angleichung an relative Häufigkeitsverteilungen erfolgen. Doch wie revidiert man im Falle selten, möglicherweise nur einmal reproduzierbarer Situationen? Die Antwort darauf liefert die Bayessche Regel. Wir wollen ihre heuristische Begründung an einem Beispiel erläutern, das im Unterricht zu recht engagierten Diskussionen führen kann.

1.3.1 Ein Kriminalbeispiel

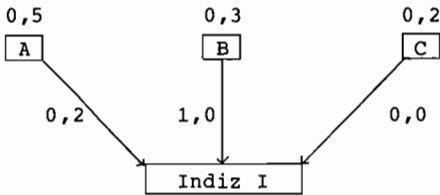
In einem Prozeß kommt auf Grund eindeutiger Indizien nur einer der Herren A, B, C als Täter in Frage. Der Kommissar ist sich noch unsicher. Nach den bisherigen Ermittlungen sagt ihm sein in langer Berufspraxis geschultes Gefühl: A war der Täter mit 50%iger Sicherheit, B mit 30%iger und C mit 20%iger Sicherheit. Nun tritt ein neues Indiz auf: Der Täter hat am Tatort eine Filterzigarette geraucht. Welche Informationen muß sich der Kommissar verschaffen? Er erfährt, daß C leidenschaftlicher Nichtraucher ist. (C hat sich vor kurzem scheiden lassen, weil er den Qualm seiner Frau nicht mehr ertragen konnte.) A und B sind starke Raucher, doch raucht A nach Aussagen der Arbeitskollegen mit ca. 20% nur selten Filterzigaretten, B raucht ausschließlich mit Filter. Auf wen tippt der Kommissar nun?

Die Schüler versetzen sich in seine Lage und wägen ab: Auf Grund des Indizes I (Täter hat Rest einer Filterzigarette hinterlassen) wird C völlig entlastet, B wird stark belastet. Auch A wird belastet, jedoch lange nicht so stark wie B. Dafür sprechen die Indizien, die vor Entdeckung der Zigarettenreste gefunden wurden stärker gegen A (50%) als gegen B (30%). Folgende Tabelle gibt die durch das Indiz revidierten subjektiven Wahrscheinlichkeitsverteilungen dreier Schüler wider:

ω	A	B	C
$P_{\text{Manfred}}(\omega)$	0,20	0,80	0,00
$P_{\text{Guido}}(\omega)$	0,35	0,60	0,05
$P_{\text{Jeanette}}(\omega)$	0,30	0,70	0,00

Aus den Unterschieden, die trotz gleicher Information für alle entstehen, ergibt sich der Wunsch nach einer eindeutigen und rational begründbaren Revisionsmethode. Das wird die Bayessche Formel sein.

Zunächst veranschaulicht man alle verfügbare Information auf prägnante Weise durch den umgekehrten Wahrscheinlichkeitsbaum:



Die "zentrale Frage":

"Auf welchen Wegen kann das beobachtete Indiz entstanden sein? Vergleiche ihre Wahrscheinlichkeiten!"

hilft, die oben intuitiv und unstrukturiert geäußerten Gedanken zu präzisieren. Sind die Schüler mit der Pfadregel vertraut, dann finden sie schnell die

Antwort auf die "zentrale Frage":

1. Weg: A war der Täter und hinterläßt Filterreste
Wahrscheinlichkeit: $p_1 = 0,5 \cdot 0,2 = 0,1$
2. Weg: B war der Täter und hinterläßt Filterreste
Wahrscheinlichkeit: $p_2 = 0,3 \cdot 1,0 = 0,3$
3. Weg: C war der Täter und hinterläßt Filterreste
Wahrscheinlichkeit: $p_3 = 0,2 \cdot 0,0 = 0,0$

Die Wahrscheinlichkeit, daß B das Indiz hinterlassen hat, also der Täter war, ist somit dreimal so groß (75%) wie die

entsprechende Wahrscheinlichkeit (25%) für A.

Allgemeiner erkennt man, daß die Summe $p=p_1+p_2+p_3$ der Wahrscheinlichkeiten für die einzelnen Wege die Wahrscheinlichkeit dafür angibt, daß das Indiz überhaupt beobachtet werden würde (Summenregel). Der Anteil p_1/p gibt dann die Wahrscheinlichkeit an, mit der es von A herrührt. Und das ist der Inhalt der Bayesschen Regel. Die revidierte subjektive Wahrscheinlichkeitsverteilung sieht also wie folgt aus:

ω	A	B	C
Prev, Bayes (ω)	$P_1/P=$ 0,1/0,4=0,25	$P_2/P=$ 0,3/0,4=0,75	$P_3/P=$ 0/0,4=0

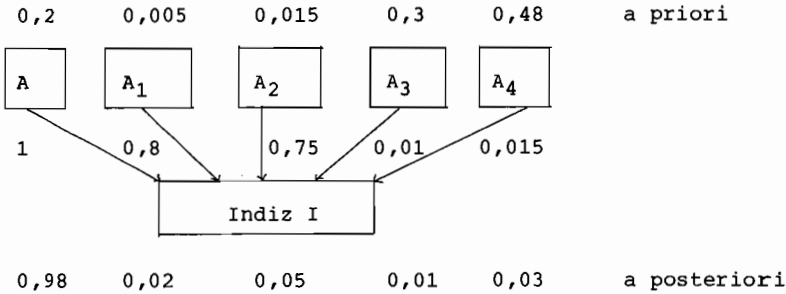
Sie stimmt mit den Ergebnissen unserer intuitiven Revision recht gut überein. Die Tatsache, daß Schüler die Bayessche Regel mitunter aus dem umgekehrten Wahrscheinlichkeitsbaum herauslesen, noch bevor man die zentrale Frage stellt, belegt die generative Kraft dieses Diagramms.

1.3.2 Bayessche Regel und induktive Logik

Aus einer wahren Implikation $A \Rightarrow I$ und der Wahrheit von I läßt uns ein psychologischer Mechanismus oft auf die Wahrheit von A zurückschließen. (In der Werbung macht man davon bekanntlich oft Gebrauch). Logisch ist das falsch, praktisch ist der Umkehrschluß vertretbar, wenn mögliche alternative Gründe für I (A_1, A_2) unwahrscheinlich sind oder (A_3, A_4) I nur selten zur Folge haben.

Im folgenden Beispiel hat A vor Beobachtung des Indizes die Wahrscheinlichkeit 0,2, nach der Beobachtung von I ist die Wahrscheinlichkeit auf 0,89 angewachsen. I macht also A sehr viel wahrscheinlicher, und meistens wird man mit den aus dem Umkehrschluß gezogenen Konsequenzen in der Realität Glück haben. Man spricht von induktiver Logik, sie hat sich während der menschlichen Entwicklung herausgebildet und gehört zum festen Repertoire an bewährten Verhaltensmustern (vgl. [Riedel]).

Numerisches Beispiel:



Zu schwerwiegenden Fehlschlüssen führt dieser Mechanismus, wenn man alternative Gründe für I übersehen hat (d.h. deren a priori Wahrscheinlichkeiten waren zu klein). Dann wird die nach Revision für A erhaltene Wahrscheinlichkeit wesentlich zu groß.

Beispiel aus dem Schulalltag: Aus $x=3 \Rightarrow x^2=9$ schließen Schüler immer wieder nach der induktiven Logik $x^2=9 \Rightarrow x=3$. Sie übersehen die negativen Zahlen (kleine subjektive a priori Wahrscheinlichkeit), weil negative Lösungen im Schulunterricht seltener vorkommen als positive. Ein Beispiel aus der Rechtssprechung findet sich bei [Schrage].

Ein intuitives Verständnis für den im Laufe der menschlichen Entwicklung erprobten Mechanismus der induktiven Logik (und seiner Grenzen) ist wegen der Universalität sicher ein Lernziel ersten Ranges. Es ist nur auf der Grundlage eines subjektiven Wahrscheinlichkeitsbegriffs zu erreichen.

1.4 Iterierte Anwendung der Bayesschen Regel

Interessant wird die Bayessche Regel bei wiederholter Anwendung, wenn die a posteriori Wahrscheinlichkeiten vorheriger Revisionen als a priori Wahrscheinlichkeiten späterer Revisionen dienen. Dann wird nämlich das Funktionieren der induktiven Logik und der prozeßhafte Charakter des Erkennens experimentell nachvollziehbar und in gewissem Sinne einer Überprüfung zugänglich. Auch die Grundgedanken der Testtheorie beginnen sich aufzudrängen, denn nach wiederholter Anwendung der Bayesschen Regel wird man so sicher, daß man bereit ist, Entscheidungen zu fällen bzw. das Risiko von Fehlentscheidungen in Kauf zu nehmen.

Wir wollen drei Beispiele geben, von denen sich das erste auch zur Einführung in die Bayessche Regel eignet.

1.4.1 Beispiel 1

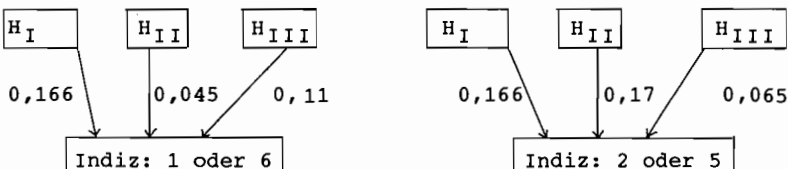
Man füllt drei Tüten: I (mit 100% weißen Kugeln), II (mit 75% weißen, 25% roten Kugeln) und III (mit 25% weißen und 75% roten Kugeln). Dann wird eine der Tüten ausgewählt. Man weiß aber nicht, welche. Infolgedessen besitzen wir für die Hypothesen H_I , H_{II} , H_{III} , die durch die Tüten symbolisiert werden, zunächst gleiche subjektive a priori Wahrscheinlichkeiten ($1/3$). Dann wird der besagten Tüte eine Kugel entnommen. Sie ist weiß. Das spricht am meisten für H_I , am wenigsten für H_{III} . Man einigt sich auf die subjektive Wahrscheinlichkeitsverteilung 55%, 30%, 15%... .

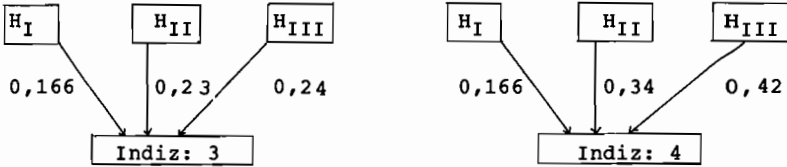
Nach der vierten weißen Kugel ist man sich ziemlich sicher, daß Tüte I vorliegt (80%, 18%, 2%), doch dann wurde (überraschenderweise) eine rote Kugel gezogen, wodurch sich die subjektiven Wahrscheinlichkeiten radikal ändern. Die folgende Tabelle zeigt links, wie Schüler einer Klasse 7 ihre subjektiven Wahrscheinlichkeiten revidierten. Daneben stehen zum Vergleich die Ergebnisse, die die Bayessche Regel liefert. Wie erwähnt dienen die revidierten (a posteriori) Verteilungen gleichzeitig als a priori Verteilungen für die nächste Revision.

Intuitive Revision:			Revision nach Bayes : (gerundet)			Rechnung:		
Hyp.I nur w	Hyp.II 3/4 w	Hyp.III 1/4 w	Hyp.I nur w	Hyp.II 3/4 w	Hyp.III 1/4 w			
33,3%	33,3%	33,3%	33,3%	33,3%	33,3%	1/3	1/3	1/3
\boxed{w}			\boxed{I} 1	\boxed{II} 3/4	\boxed{III} 1/4	1.Weg 2.Weg 3.Weg	4/12 3/12 1/12	8/12
55,0%	30,0%	15,0%	50,0%	37,5%	12,5%	4/8	3/8	1/8
\boxed{w}			\boxed{I} 1	\boxed{II} 3/4	\boxed{III} 1/4	1.Weg 2.Weg 3.Weg	16/32 9/32 1/32	26/32
60,0%	30,0%	10,0%	61,5%	34,6%	3,9%	16/26	9/26	1/26
\boxed{w}			\boxed{I} 1	\boxed{II} 3/4	\boxed{III} 1/4	1.Weg 2.Weg 3.Weg	64/104 27/104 1/104	92/104
75,0%	20,0%	5,0%	69,5%	29,4%	1,1%	64/92	27/92	1/92
\boxed{w}			\boxed{I} 1	\boxed{II} 3/4	\boxed{III} 1/4	1.Weg 2.Weg 3.Weg	256/368 81/368 1/368	338/368
80,0%	18,0%	2,0%	75,7%	23,9%	0,4%	256/338	81/338	1/338
\boxed{r}			\boxed{I} 0	\boxed{II} 1/4	\boxed{III} 3/4	1.Weg 2.Weg 3.Weg	0/1352 81/1352 3/1352	84/1352
0,0%	95,0%	5,0%	0,0%	96,4%	3,6%	0	81/84	3/84

1.4.2 Beispiel 2 (Vertiefung zum Zinkenwürfel aus 1.2.3)

Bei einem unbekanntem Objekt ist zu entscheiden, ob es sich um einen Laplace-Würfel (Hypothese H_I), einen kurzen (H_{II}) oder einen langen Zinkenwürfel (H_{III}) handelt. Dabei charakterisieren wir die Zinkenwürfel durch die Verteilungen, die sich in früheren Versuchen (1.2.3) bewährt haben. Entsprechend der gewürfelten Zahl revidieren wir unsere subjektiven a priori Wahrscheinlichkeiten nach einem der folgenden Diagramme:





Die Indizien 1 und 6 sprechen nach der induktiven Logik am meisten für den Laplace-Würfel, die Indizien 2 und 5 sprechen sehr für den kurzen und gegen den langen Zinkenwürfel, 4 spricht am meisten für den langen Zinkenwürfel. Den Prozeß des Erkenntnisgewinns verfolgt man anhand der untenstehenden Tabelle:

H_I :Laplace(%)						H_{II} : kurz (%)						H_{III} : lang (%)					
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
16,6	16,6	16,6	16,6	16,6	16,6	4,5	17	23	34	17	4,5	11	65	24	42	65	11
gewürfeltes subjektive Wahrscheinlichkeit für die Hypothesen: Indiz:																	
	33,33%					33,33%						33,33%					
3	25,00%					33,33%						42,00%					
4	12,30%					33,13%						54,75%					
4	4,49%					30,18%						64,33%					
3	3,58%					25,97%						70,45%					
3	2,29%					21,94%						75,76%					
4	0,93%					18,12%						80,95%					
4	0,37%					14,96%						84,94%					
4	0,14%					11,77%						<u>88,08%</u>					
5	0,37%					31,75%						67,88%					
4	0,15%					26,51%						73,34%					
5	0,30%					55,68%						44,02%					
3	0,21%					49,74%						50,05%					
2	0,31%					77,43%						22,26%					
4	0,14%					72,78%						27,08%					
2	0,17%					90,24%						9,59%					
6	0,56%					82,01%						17,43%					
3	0,41%					78,39%						21,21%					
5	0,46%					<u>92,40%</u>						7,14%					

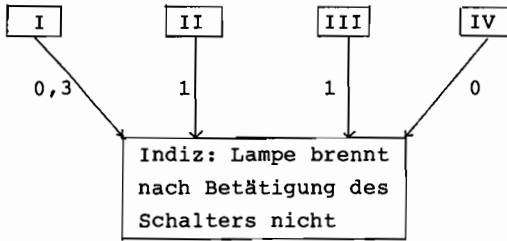
Nach 18 Versuchen wird die Beobachtung abgebrochen. Mit hoher Wahrscheinlichkeit liegt der kurze Zinkenwürfel vor. Hätte man sich nach 8 Versuchen für den langen Zinkenwürfel entschieden, wäre ein Irrtum unterlaufen. Man beachte, wie sich die Situation durch das wiederholte Auftreten der Indizien 2 und 5 verändert hat.

1.4.3 Beispiel 3 (Schreibtischlampe)

Eine nicht mehr ganz neue Schreibtischlampe soll angeknipst werden. Es ist bekannt, daß der Druckschalter in gewissen zeitlichen Abständen Wackelkontaktphasen hat, sie leuchtet dann beim Anknipsen mit 30%iger Wahrscheinlichkeit nicht auf. Auch brennt wegen des zu kleinen Schirmes die Birne öfter durch als normal. Aufgrund längerer Erfahrung (ich war immer zu faul, den Schalter auszuwechseln), besitze ich folgende subjektive Wahrscheinlichkeitsverteilung für den Lampenzustand:

ω	I: Schalter hat Wackelkontakt- phase, Birne ist ok	II: Birne ist kaputt, Schalter ist ok	III: Schalter hat Wackelkontakt- phase, Birne ist kaputt	IV: Alles ist in Ordnung
$p(\omega)$	0,1	0,01	0,001	0,889

Ich komme ins Zimmer, knipse die Lampe an und es bleibt dunkel (Indiz). Bevor ich lange nachdenke, werde ich den wackelkontaktverdächtigen Schalter erneut betätigen - und nochmals. Ist während - sagen wir - sechsmaliger Betätigung des Schalters das Licht immer noch nicht angegangen, werde ich es mit einer neuen Glühbirne versuchen. Warum? Die Antwort liefert die Regel von Bayes. Wir benutzen sie sechsmal hintereinander zur Revision unserer subjektiven Wahrscheinlichkeitsverteilungen für den Lampenzustand durch das Indiz. Dabei kommt stets der gleiche Wahrscheinlichkeitsbaum zur Anwendung, nur die a priori Verteilungen ändern sich:



Es ergeben sich folgende Zwischenergebnisse ($p_n(\omega)$ bezeichnet die subjektive Wahrscheinlichkeit für den Lampenzustand, nachdem der n 'te Druck auf den Schalter die Lampe immer noch nicht zum Leuchten brachte; H bezeichnet die Entropie der Wahrscheinlichkeitsverteilungen):

ω	I	II	III	IV	H
$P_0(\omega)$	0,100	0,010	0,001	0,889	0,547
$P_1(\omega)$	0,732	0,244	0,024	0,000	0,957
$P_2(\omega)$	0,450	0,500	0,050	0,000	1,234
$P_3(\omega)$	0,197	0,730	0,073	0,000	1,096
$P_4(\omega)$	0,069	0,847	0,085	0,000	0,770
$P_5(\omega)$	0,022	0,889	0,089	0,000	0,580
$P_6(\omega)$	0,007	0,903	0,090	0,000	0,494

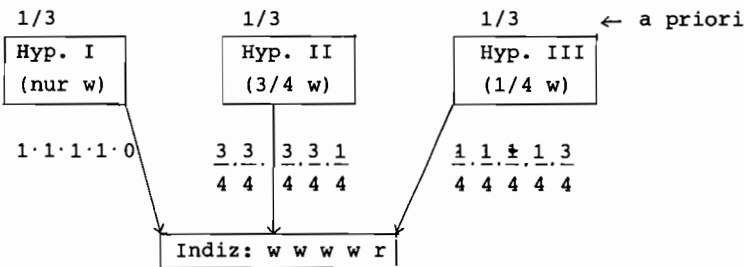
Nachdem bei der ersten Betätigung des Schalters kein Licht anging, schnellte die Wahrscheinlichkeit für einen Wackelkontakt von 10% auf 73% (weswegen man es sofort nochmals probiert). Nach sechs erfolglosen Versuchen ist diese Wahrscheinlichkeit jedoch auf 0,7% abgesunken, und der Gang zum Kasten mit den Ersatzbirnen ist sinnvoll. Funktioniert die Lampe auch nach Einschrauben einer neuen Birne nicht, werde ich zu weiteren Revisionen gezwungen sein und einer anderen Alternative (etwa Anschlußschnur kaputt oder Sicherung gelöst) eine vorher zu vernachlässigende Wahrscheinlichkeit einräumen müssen.

Nur am Rande sei angemerkt, wie man an Beispielen des obigen Typs Veränderungen der Entropie einer Situation durch ein-treffende Informationen studieren kann: Vor dem ersten Betä-

tigen des Schalters bin ich mir über den Zustand der Lampe ziemlich im klaren: "Sie wird höchstwahrscheinlich funktionieren." Die Entropie $H = -\sum p(\omega) \lg(1/p(\omega))$ der Situation ist gering, sie beträgt 0,547. Nachdem die Lampe nicht angegangen ist, bin ich über ihren Zustand unsicher geworden: "Mit recht hoher Wahrscheinlichkeit liegt der Fehler aber an einem Wackelkontakt." Die Entropie ist auf 0,975 angewachsen. Noch unsicherer macht mich ein erneutes Versagen der Lampe. "Liegt es womöglich doch an der Birne?" (Entropie 1,234). Die sechsmalige Betätigung des Schalters liefert wieder Klarheit. "Höchstwahrscheinlich liegt der Fehler an der Birne." Die Entropie der Situation ist auf 0,494 abgesunken.

Anmerkung:

Natürlich kann man in den obigen Beispielen das Endergebnis auch durch einmalige Anwendung der Bayesschen Regel erhalten, man hat nur ein "zusammengesetztes Indiz" zu betrachten, die bedingten Wahrscheinlichkeiten ergeben sich aus der Pfadregel. So erhält man für das Beispiel 1.4.1 der Tüten mit unbekanntem Inhalt folgenden Baum:



Rechnung:

1. Weg	(1/3)	1	1	1	1	0	= 0	}	28/1024
2. Weg	(1/3)	(3/4)	(3/4)	(3/4)	(3/4)	(1/4)	= 27/1024		
3. Weg	(1/3)	(1/4)	(1/4)	(1/4)	(1/4)	(3/4)	= 1/1024		

0/28

27/28

1/28

← a posteriori

Werden die auftretenden Zahlen zu klein, dann betrachtet man nur Verhältnisse von Wahrscheinlichkeiten. Auf diese Weise wurde die Spalte "Sicherheit" in der Tabelle aus 1.2.3 (Zinkenwürfel) berechnet. So erhält man für die a posteriori Wahrscheinlichkeiten dafür, daß die experimentellen Ergebnisse von einem kurzen bzw. langen Würfel stammen, das Verhältnis

$$\frac{p(\text{a post. kurz})}{p(\text{a post. lang})} = \frac{\left[\frac{0,045}{0,11} \right]^{n_1/6} \cdot \left[\frac{0,17}{0,065} \right]^{n_2/5} \cdot \left[\frac{0,23}{0,24} \right]^{n_3} \cdot \left[\frac{0,34}{0,42} \right]^{n_4} \cdot \frac{p(\text{a pr. kurz})}{p(\text{a pr. lang})}$$

wobei die Exponenten die absoluten Häufigkeiten der gewürfelten Indizien enthalten.

1.5 Die subjektiven Wahrscheinlichkeiten werden objektiv

Anfangs, als die betrachteten Situationen beliebig oft reproduzierbar waren (1.2), revidierten wir subjektive Wahrscheinlichkeiten dadurch, daß wir sie mit relativen Häufigkeiten abstimmten. Was würde eine wiederholte Revision unter Benutzung der Bayesschen Regel liefern? Diese Frage soll uns eine Computersimulation beantworten.

1.5.1 Beispiel

Greifen wir das Beispiel 1.2.1 des Vierfachmünzwurfs auf und betrachten wir die Wahrscheinlichkeit $w=w(2)$ für den Ausfall 2 (zwei der vier Münzen zeigen Kopf). Frank, Jeanette und Arndt hatten hierfür die subjektiven Wahrscheinlichkeiten $w=0,2$, $w=0,3$ und $w=0,5$ genannt. In der Sprache der Statistiker haben sie verschiedene Hypothesen über eine unbekannte Wahrscheinlichkeit aufgestellt:

Hypothese von Frank	$H_{Fr}: w=0,2$
Hypothese von Jeanette	$H_{Je}: w=0,3$
Hypothese von Arndt	$H_{Ar}: w=0,5$

Nehmen wir die Hypothese hinzu, die wir nach 200 Experimenten aufgestellt haben:

Hypothese eines Experimentators	$H_{Ex}: w=0,4$
---------------------------------	-----------------

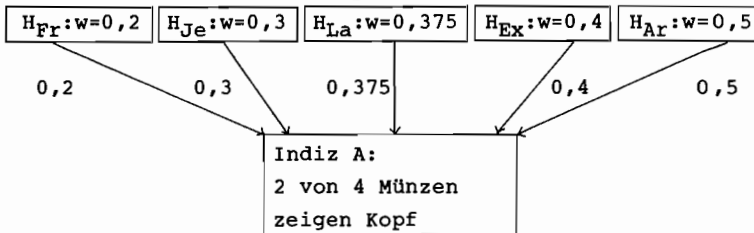
und geben wir auch Herrn Laplace mit seiner Hypothese eine Chance:

Hypothese von Laplace	$H_{La}: w=0,375$
-----------------------	-------------------

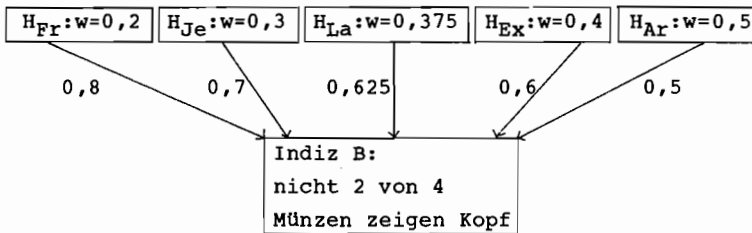
Frage: Wer hat recht? Oder welche Hypothese besitzt die größte Wahrscheinlichkeit? Vor allen Experimenten und theoretischen Überlegungen sind die drei Schüler unsicher, doch hält jeder seine eigene Hypothese für die wahrscheinlichste, sonst hätte er sie wohl kaum spontan vorgeschlagen. Sagen wir der Einfachheit halber: Jeder der drei glaubt mit

60% iger Wahrscheinlichkeit an seine eigene Hypothese, mit 10% iger Wahrscheinlichkeit an die vier anderen (subjektive a priori Verteilung auf den Hypothesen). Wir machen nun eine Reihe von Experimenten und lassen die drei Schüler je nach Ausgang ihre subjektiven Wahrscheinlichkeitsverteilungen gemäß der Bayesschen Regel revidieren. Dabei wird einer der folgenden Bäume benutzt:

Falls 2 von den 4 Münzen Kopf zeigen:



Falls nicht 2 von 4 Münzen Kopf zeigen:



Nach zwei tatsächlich durchgeführten Vierfachmünzwürfen und den zugehörigen Revisionen nimmt uns der Computer mit seinem Zufallsgenerator die Arbeit ab: Er druckt neben der Nummer N des Vierfachmünzwurfs (1. Spalte) das im N-ten Versuch eingetretene Indiz A oder B (zwei bzw. keine zwei Köpfe). Die 3. Spalte enthält die relativen Häufigkeiten der in den ersten N Versuchen aufgetretenen Indizien. Daneben stehen die Folgen der durch die Indizien nach Bayes revidierten subjektiven Wahrscheinlichkeiten für die fünf genannten Hypothesen. Diese Folgen sind für die drei Schüler verschieden, da sie anfangs verschiedene a priori Verteilungen besaßen.

Ergebnisse der Simulation (S. 42):

(a) Unabhängig voneinander räumen die drei Kandidaten nach gewisser Zeit der Hypothese von Laplace, die den relativen Häufigkeiten am nächsten kommt, die größte Wahrscheinlichkeit ein. Nach 2000 Versuchen ist sie auf über 99% angewachsen.

(b) für $N > 800$ stimmen die subjektiven Wahrscheinlichkeitsverteilungen aller drei Kandidaten überein, die Übereinstimmung ist besser als 1%. Das bedeutet aber: Die subjektiven Wahrscheinlichkeiten sind objektiv geworden.

Bezüglich der eingangs gestellten Frage halten wir fest:

(c) Das intuitive Angleichen der subjektiven Wahrscheinlichkeiten an die relativen Häufigkeiten ist mit einer iterierten Revision nach Bayes gut vereinbar.

(d) Die Laplace-Hypothese kann (zumindest für die im Computer simulierten Münzen) aufrechterhalten werden.

1.5.2 Variation des Beispiels (S. 43)

Eben haben sich drei Personen zwischen 5 Hypothesen über die Wahrscheinlichkeit von 2 Köpfen in einem Vierfachmünzwurf entschieden (linke Spalte untenstehender Tabelle). Wir können auch folgende Deutung aussprechen: Sie haben sich zwischen den in 1.2.1 aufgestellten "komplizierteren" Hypothesen über die gesamte Verteilung der Kopfzahl (rechte Tabellenseite) entschieden, dabei allerdings zur Entschei-

Kopfzahl	2	0	1	2	3	4
H_{Fr}	$w=0,2$	0,2	0,2	0,2	0,2	0,2
H_{Je}	$w=0,3$	0,1	0,25	0,3	0,25	0,1
H_{Ex}	$w=0,4$	0,06	0,24	0,4	0,24	0,06
H_{La}	$w=0,375$	0,0625	0,25	0,375	0,25	0,0625
H_{Ar}	$w=0,5$	0,05	0,2	0,5	0,2	0,05

dung nur die Indizien A (zwei der 4 Münzen zeigen Kopf) und B (nicht zwei der 4 Münzen zeigen Kopf) herangezogen. Bei dieser Interpretation haben sie also auf Informationen verzichtet. Wir erwarten, daß sich die Hypothese von Laplace

Zum Beispiel 1.5.1

5 Hypothesen über die Wahrscheinlichkeit von 2 Köpfen beim Vierfachmünzwurf:

Table with 5 columns: H Fr, H Je, H La, H Ex, H Ar. Values: w=0,2; w=0,3; w=0,375; w=0,4; w=0,5

Als Indiz zur Abwägung zwischen den Hypothesen beobachtet man
A: 2 von 4 Münzen zeigen Kopf oder
B: nicht 2 von 4 Münzen zeigen Kopf

Main data table with columns for N, I, VERT, FRANK, JEANETTE, ARNOT, and various probability values. It contains a large grid of numerical data points.

Zum Beispiel 1.5.2 (Variation von 1.5.1)

5 Hypothesen über die Verteilung der Kopffzahl beim Vierfachmünzwurf:

	H Fr	H Je	H La	H Ex	H Ar
0 Köpfe:	0,2	0,1	0,0625	0,06	0,05
1 Kopf:	0,2	0,25	0,25	0,24	0,2
2 Köpfe:	0,2	0,3	0,375	0,4	0,5
3 Köpfe:	0,2	0,25	0,25	0,24	0,2
4 Köpfe:	0,2	0,1	0,0625	0,06	0,05

Als Indiz zur Abwägung zwischen den Hypothesen beobachtet man die gefallene Kopffzahl

N	VERTEILUNG DER INDIZIEN I:					FRANK: WK-VERTEILUNG FUER					JEANETTE: SUBJ. WK-VERTEILUNG FUER					ARMDT: WK-VERTEILUNG FUER					
	D	1	2	3	4	H Fr	H Je	H La	H Ex	H Ar	H Fr	H Je	H La	H Ex	H Ar	H Fr	H Je	H La	H Ex	H Ar	
	a priori					0.600	0.100	0.100	0.100	0.100	0.100	0.600	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.600
1	0.000	1.000	0.000	0.000	0.000	0.561	0.117	0.117	0.112	0.093	0.084	0.628	0.105	0.100	0.084	0.093	0.117	0.117	0.112	0.561	
2	0.000	0.500	0.500	0.000	0.000	0.397	0.124	0.155	0.159	0.165	0.051	0.577	0.120	0.123	0.128	0.044	0.083	0.104	0.106	0.663	
3	0.000	0.330	0.345	0.291	0.069	0.247	0.116	0.147	0.198	0.252	0.007	0.566	0.132	0.138	0.140	0.020	0.375	0.242	0.242	0.995	
4	0.000	0.250	0.250	0.000	0.000	0.137	0.097	0.189	0.220	0.358	0.017	0.423	0.138	0.161	0.261	0.009	0.336	0.071	0.082	0.803	
5	0.000	0.200	0.200	0.000	0.000	0.343	0.121	0.147	0.165	0.224	0.043	0.550	0.112	0.125	0.170	0.013	0.066	0.080	0.090	0.732	
6	0.000	0.167	0.167	0.000	0.167	0.203	0.107	0.164	0.195	0.331	0.025	0.471	0.120	0.143	0.242	0.014	0.043	0.066	0.079	0.799	
7	0.000	0.143	0.143	0.143	0.000	0.184	0.123	0.167	0.212	0.291	0.007	0.590	0.127	0.140	0.206	0.013	0.028	0.038	0.090	0.766	
8	0.000	0.250	0.500	0.125	0.125	0.164	0.135	0.206	0.227	0.267	0.018	0.527	0.134	0.147	0.174	0.012	0.062	0.094	0.103	0.729	
9	0.000	0.222	0.556	0.111	0.111	0.088	0.108	0.206	0.242	0.356	0.010	0.442	0.141	0.165	0.243	0.005	0.040	0.076	0.089	0.789	
10	0.000	0.300	0.500	0.100	0.100	0.078	0.120	0.229	0.258	0.319	0.008	0.469	0.149	0.168	0.206	0.005	0.048	0.091	0.102	0.754	
11	0.000	0.333	0.500	0.083	0.083	0.068	0.134	0.247	0.271	0.271	0.007	0.500	0.157	0.170	0.217	0.005	0.079	0.108	0.116	0.714	
12	0.000	0.385	0.500	0.077	0.077	0.035	0.100	0.239	0.275	0.352	0.004	0.408	0.162	0.187	0.239	0.002	0.037	0.087	0.101	0.773	
13	0.000	0.432	0.412	0.118	0.059	0.030	0.110	0.262	0.290	0.308	0.003	0.432	0.172	0.190	0.202	0.002	0.044	0.104	0.115	0.735	
14	0.000	0.418	0.409	0.227	0.061	0.002	0.072	0.333	0.350	0.244	0.000	0.349	0.173	0.208	0.272	0.001	0.028	0.084	0.099	0.788	
15	0.000	0.400	0.467	0.067	0.067	0.013	0.089	0.265	0.300	0.353	0.002	0.372	0.185	0.209	0.232	0.001	0.033	0.100	0.113	0.753	
16	0.000	0.438	0.438	0.063	0.063	0.011	0.097	0.288	0.314	0.290	0.001	0.394	0.196	0.213	0.197	0.001	0.040	0.118	0.129	0.713	
17	0.000	0.412	0.412	0.118	0.059	0.010	0.104	0.311	0.325	0.250	0.001	0.414	0.205	0.215	0.165	0.001	0.047	0.139	0.145	0.669	
18	0.000	0.399	0.364	0.227	0.056	0.000	0.126	0.280	0.321	0.309	0.001	0.336	0.208	0.232	0.223	0.000	0.030	0.113	0.126	0.730	
19	0.000	0.368	0.421	0.158	0.053	0.004	0.068	0.312	0.333	0.267	0.000	0.355	0.220	0.236	0.189	0.000	0.037	0.134	0.143	0.687	
20	0.000	0.350	0.450	0.150	0.050	0.002	0.061	0.285	0.325	0.326	0.000	0.282	0.219	0.249	0.250	0.000	0.023	0.108	0.124	0.744	
21	0.000	0.333	0.429	0.190	0.048	0.002	0.065	0.310	0.339	0.283	0.000	0.300	0.233	0.255	0.213	0.000	0.028	0.128	0.140	0.704	
22	0.000	0.318	0.409	0.227	0.040	0.002	0.072	0.333	0.350	0.244	0.000	0.318	0.246	0.268	0.180	0.000	0.032	0.150	0.158	0.660	
23	0.000	0.348	0.391	0.217	0.043	0.004	0.076	0.355	0.359	0.208	0.000	0.332	0.257	0.260	0.151	0.000	0.037	0.174	0.176	0.612	
24	0.042	0.333	0.375	0.208	0.042	0.001	0.123	0.358	0.347	0.168	0.000	0.458	0.222	0.215	0.104	0.000	0.057	0.195	0.189	0.548	
25	0.089	0.320	0.360	0.200	0.040	0.013	0.190	0.346	0.322	0.130	0.001	0.588	0.178	0.166	0.067	0.001	0.116	0.211	0.196	0.475	
26	0.177	0.308	0.385	0.192	0.038	0.007	0.149	0.339	0.336	0.169	0.001	0.514	0.195	0.193	0.097	0.001	0.081	0.184	0.183	0.582	
27	0.107	0.333	0.370	0.185	0.037	0.006	0.157	0.356	0.339	0.142	0.001	0.528	0.200	0.191	0.080	0.001	0.092	0.209	0.199	0.500	
28	0.071	0.321	0.357	0.179	0.071	0.017	0.235	0.334	0.306	0.107	0.001	0.653	0.195	0.141	0.049	0.002	0.155	0.220	0.201	0.422	
29	0.069	0.310	0.345	0.207	0.069	0.014	0.244	0.347	0.305	0.089	0.001	0.664	0.197	0.138	0.040	0.002	0.171	0.242	0.213	0.372	
30	0.107	0.333	0.333	0.200	0.067	0.012	0.253	0.359	0.303	0.074	0.001	0.673	0.159	0.134	0.033	0.001	0.186	0.265	0.223	0.325	
31	0.065	0.323	0.323	0.226	0.065	0.010	0.260	0.370	0.299	0.061	0.001	0.681	0.161	0.131	0.026	0.001	0.201	0.286	0.231	0.281	
32	0.106	0.344	0.313	0.219	0.063	0.008	0.267	0.380	0.295	0.050	0.001	0.688	0.163	0.127	0.021	0.001	0.215	0.306	0.238	0.240	
33	0.061	0.333	0.303	0.242	0.061	0.007	0.274	0.389	0.290	0.041	0.000	0.695	0.165	0.123	0.017	0.001	0.228	0.325	0.242	0.204	
34	0.059	0.324	0.324	0.235	0.059	0.004	0.225	0.399	0.317	0.056	0.000	0.636	0.188	0.150	0.026	0.000	0.176	0.313	0.249	0.262	
35	0.107	0.343	0.314	0.229	0.057	0.003	0.230	0.409	0.312	0.046	0.000	0.643	0.190	0.145	0.021	0.000	0.188	0.333	0.255	0.224	
36	0.056	0.333	0.333	0.222	0.056	0.002	0.185	0.414	0.337	0.052	0.000	0.579	0.214	0.175	0.032	0.000	0.143	0.316	0.258	0.283	
37	0.059	0.351	0.359	0.216	0.054	0.001	0.191	0.425	0.335	0.051	0.000	0.587	0.237	0.170	0.026	0.005	0.153	0.308	0.265	0.243	
38	0.053	0.342	0.316	0.211	0.079	0.004	0.280	0.388	0.291	0.037	0.000	0.701	0.162	0.122	0.015	0.001	0.237	0.328	0.246	0.188	
39	0.051	0.333	0.308	0.231	0.077	0.003	0.285	0.396	0.285	0.030	0.000	0.706	0.163	0.118	0.012	0.000	0.248	0.345	0.248	0.158	
40	0.050	0.325	0.325	0.225	0.075	0.002	0.235	0.408	0.314	0.042	0.000	0.649	0.186	0.144	0.019	0.000	0.195	0.305	0.260	0.207	
41	0.049	0.317	0.341	0.220	0.073	0.001	0.191	0.413	0.339	0.056	0.000	0.586	0.212	0.174	0.029	0.000	0.149	0.323	0.265	0.253	
42	0.048	0.310	0.357	0.214	0.071	0.000	0.152	0.412	0.361	0.075	0.000	0.519	0.234	0.205	0.042	0.000	0.111	0.300	0.263	0.326	
43	0.047	0.302	0.372	0.209	0.070	0.000	0.120	0.405	0.378	0.098	0.000	0.489	0.253	0.236	0.061	0.000	0.080	0.272	0.254	0.394	
44	0.045	0.295	0.364	0.227	0.068	0.000	0.124	0.419	0.376	0.081	0.000	0.459	0.259	0.232	0.050	0.000	0.088	0.299	0.287	0.346	
45	0.044	0.289	0.378	0.222	0.067	0.000	0.096	0.408	0.390	0.105	0.000	0.390	0.276	0.263	0.071	0.000	0.063	0.286	0.256	0.413	
46	0.043	0.283	0.370	0.239	0.065	0.000	0.100	0.424	0.389	0.087	0.000	0.400	0.283	0.259	0.058	0.000	0.070	0.295	0.271	0.365	
47	0.043	0.277	0.362	0.255	0.065	0.000	0.100	0.438	0.386	0.072	0.000	0.409	0.289	0.254	0.049	0.000	0.076	0.322	0.294	0.318	
48	0.042	0.271	0.375	0.260	0.063	0.000	0.080	0.426	0.400	0.094	0.000	0.344	0.304	0.285	0.067	0.000	0.055	0.290	0.273	0.383	

schneller durchsetzt, wenn die ganze Information, also die tatsächlich beobachtete Anzahl von Köpfen zur Revision nach Bayes benutzt wird. Die zugehörige Computersimulation bestätigt diese Erwartung, wenn auch die Unterschiede zur ersten Simulation mit wachsender Versuchszahl verschwinden. Das ist nicht verwunderlich, denn die Hypothesen sind in diesem Beispiel symmetrisch. Deswegen kommt im wesentlichen nur die Information über die Indizien 0 und 4 hinzu, welche aber für eine Entscheidung zwischen H_{Ex} und H_{La} viel weniger relevant ist als die (schon in der ersten Simulation verwendete) Information über das Indiz 2.

In beiden Simulationen wurde die gleiche Folge von Zufallszahlen benutzt. Immer, wenn bei der ersten Simulation das Indiz A auftaucht, beobachtet man in der zweiten Simulation das Indiz 2.

1.6 Testen von Hypothesen

Mit 1.4 und 1.5 befinden wir uns schon mitten in der Statistik und der Schritt zum Testen von Hypothesen ist bestens vorbereitet. Im folgenden Abschnitt soll er der Vollständigkeit halber an einem Beispiel skizziert werden.

Wir messen die "Weichheit" eines Lehrers durch einen Softy-Parameter

$$s = \frac{n_{1-2}}{n_{1-2} + n_{5-6}} \quad 0 \leq s \leq 1 ,$$

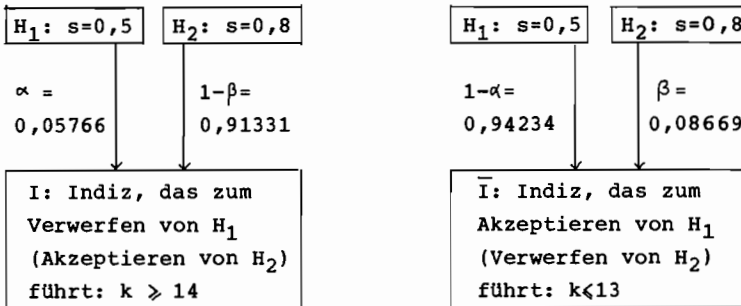
dabei ist n_{1-2} die Anzahl der Noten im Bereich 1 und 2, die der Lehrer verteilt, analog n_{5-6} , und der Softy-Parameter s gibt den Anteil der guten Randnoten an der Gesamtheit der Randnoten an. $s=0$ bedeutet, der Lehrer ist extrem hart, er gibt nur 5 und 6, aber keine 1 und 2. Bei $s=1$ ist er extrem weich, er gibt nur 1 und 2, aber keine 5 und 6. Natürlich kann der Parameter für verschiedene Klassen schwanken, aber Lehrer, die gleichzeitig in verschiedenen Fächern einer Klasse oder nacheinander im gleichen Fach unterrichten, lassen sich durchaus miteinander vergleichen.

Von einem Deutschlehrer A ist bekannt, daß er den Softy-Parameter 0,5 besitzt. (Das ist schon recht hart). Es ist kein Geheimnis, daß Schüler trotz anders ausgeprägter Interessen oft den Kurs des Lehrers mit dem höheren Softy-Parameter wählen. Ein Kollege B (mit dem gleichen Fach) behauptet, er besitze den Parameter 0,8. Die Schüler haben zu entscheiden, ob B tatsächlich besser benotet als A oder das nur vorgibt. Sie haben zwischen folgenden Hypothesen zu entscheiden

H_1 : B hat auch nur den Parameter $s=0,5$
(eine Umwahl würde sich nicht "lohnen")

H_2 : B hat den Parameter $s=0,8$
(man sollte eine Umwahl versuchen).

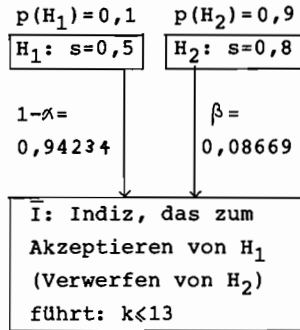
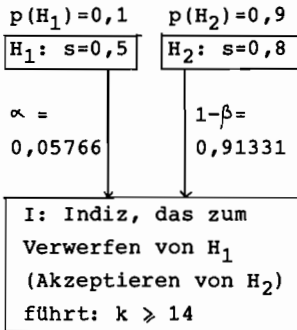
Vor der Entscheidung stellt man einige Recherchen an, man macht (etwa 20) "Randnoten" (1,2,5,6) des Lehrers B ausfindig und beschließt, bei einem hohen Anteil guter Noten (1/2), etwa $k > 14$, die Hypothese H_1 zu verwerfen. Schüler, die mit dem umgekehrten Wahrscheinlichkeitsbaum gearbeitet haben, skizzieren die Situation wie folgt: ¹⁾



Man erkennt nach der induktiven Logik, daß das Eintreten des Indizes I tatsächlich sehr für H_2 und gegen H_1 , das Eintreten von \bar{I} sehr für H_1 und gegen H_2 spricht. Eine Aussage über die Wahrscheinlichkeit von H_1 bzw. H_2 ist jedoch nicht möglich. Besitzen wir aber Informationen über den Kollegen B, so besitzen wir eine subjektive a priori Verteilung p für die Gültigkeit von H_1 bzw. H_2 . Stammt das Gerücht über den Parameter $s=0,8$ aus einer Klasse, in der B lange unterrichtet hat, gilt vielleicht $p(H_1)=0,1$, $p(H_2)=0,9$. Der Schulleiter, der die Meinung zu vertreten hat, die Noten würden von allen Kollegen in gleicher Weise vergeben, d.h. auch der Parameter von B liegt in der Nähe von 0,5, muß sich die Verteilung $p(H_1)=0,9$, $p(H_2)=0,1$ zueigen machen. Ein unvoreingenommener Beobachter bewertet beide Hypothesen vielleicht mit 0,5.

1) Die bedingten Wahrscheinlichkeiten, die zu den Pfeilen gehören, werden mit der Binomialverteilung berechnet; α und β haben die Bedeutung von Fehlerwahrscheinlichkeiten erster und zweiter Art (s.u.).

Wir können jetzt unsere Diagramme ergänzen



und je nach Auftreten der Indizien I bzw. \bar{I} unsere subjektiven Wahrscheinlichkeitsverteilungen nach Bayes revidieren. Für die drei erwähnten a priori Verteilungen sehen die numerischen Ergebnisse folgendermaßen aus: ¹⁾

ω	H_1	H_2	H_1	H_2	H_1	H_2
$p(\omega)$	0,1	0,9	0,5	0,5	0,9	0,1
falls I eintritt:						
$P_{\text{revidiert}}(\omega)$	0,007	0,993	0,059	0,941	0,362	0,638
falls \bar{I} eintritt:						
$P_{\text{revidiert}}(\omega)$	0,547	0,543	0,916	0,084	0,990	0,010
$p(H_1)\alpha + p(H_2)\beta$	0,084		0,072		0,061	

Die qualitative Aussage, daß I sehr für H_2 , \bar{I} sehr für H_1 spricht, bestätigt sich natürlich auch quantitativ. Wichtig

1) Die prinzipielle Symmetrie zwischen den Hypothesen wird durch die unterschiedliche Vorbewertung aufgehoben. Hier drängt sich der Begriff der Nullhypothese auf. Es ist diejenige Hypothese, deren fälschliches Verwerfen die schlimmeren Folgen hat, die man also vom Bayesschen Standpunkt tunlichst mit der höheren a priori Wahrscheinlichkeit bewerten sollte. So hat in der linken Spalte H_2 die Bedeutung der Nullhypothese, in der rechten Spalte ist es umgekehrt. In der linken Spalte wird H_1 gegen H_2 , in der rechten Spalte H_2 gegen H_1 "getestet".

ist, daß der begriffliche Unterschied zwischen Bayesschem Prinzip und Hypothesentesten deutlich wird:

- (a) Das Indiz, auf dessen Beobachtung es ankommt, hat eine völlig andere Form bekommen, aus einem Punkt ist ein Intervall (kritische Region) geworden.
- (b) Es ist vor der Untersuchung festzulegen.
- (c) Es entscheidet jetzt unmittelbar über Annahme bzw. Ablehnung von H_1 bzw. H_2 und dient nicht mehr bloß zum Abwägen zwischen Alternativen.

Wegen (c) erhalten die vertrauten, bei Anwendung der Bayesschen Regel anfallenden Wahrscheinlichkeiten neue und überraschende Interpretationen, es sind Irrtumswahrscheinlichkeiten geworden:

- (d) α ist die Wahrscheinlichkeit eines Irrtums unter der Voraussetzung, daß H_1 gilt (=Signifikanzniveau), β ist die Wahrscheinlichkeit eines Irrtums unter der Voraussetzung, daß H_2 gilt.
- (d) Deswegen ist $p(H_1)\alpha + p(H_2)\beta$ die a priori Gesamtwahrscheinlichkeit eines Irrtums.
- (e) Im Gegensatz dazu haben die eben berechneten Größen $P_{rev.}(H_1)$ bzw. $P_{rev.}(H_2)$ die Bedeutung von a posteriori Irrtumswahrscheinlichkeiten erhalten. Genauer ist:
 $P_{rev.}(H_1)$ die Wahrscheinlichkeit eines Irrtums unter der Voraussetzung, daß H_1 abgelehnt wurde,
 $P_{rev.}(H_2)$ die Wahrscheinlichkeit eines Irrtums unter der Voraussetzung, daß H_2 abgelehnt wurde.
 Die Zahl $p_{rev.}(H_1)$ gibt nämlich die Wahrscheinlichkeit für die Gültigkeit von H_1 an, nachdem I eingetreten ist, also H_1 verworfen wurde. Noch umgangssprachlicher ist $p_{rev.}(H_1)$ die Wahrscheinlichkeit, daß die nach Eintreten von I gefällte Entscheidung, H_1 zu verwerfen, falsch war. Analog für $P_{rev.}(H_2)$.

Es ist außerordentlich schwierig, Schülern, die mit der Bayesschen Regel nicht vertraut sind, den Unterschied zwi-

schen den Irrtumswahrscheinlichkeiten (d) bis (e), insbesondere die Bedeutung des Signifikanzniveaus, zu vermitteln. Diese bedingte Irrtumswahrscheinlichkeit wird ohne den Bayesschen Rahmen von Schülern intuitiv oft als Wahrscheinlichkeit für die Gültigkeit von Hypothesen gedeutet, wie das eingangs erwähnte Zitat belegt. Ebenso stößt der Unterschied, den die Testtheorie zwischen Nullhypothese und Alternative macht, ohne Vorbereitung oft auf Unverständnis.

Das umgekehrte Baumdiagramm ist zur Vorbeugung gegen solche Fehlvorstellungen insofern hilfreich, als es erstens die Diskussion über die Bedeutung von α und β auf visuell eingängige Weise unterstützt. Die funktionalen Abhängigkeiten vom kritischen Bereich werden deutlich und nach dem Prinzip der induktiven Logik sieht man: Je kleiner α und β , desto mehr spricht das Auftreten von I für H_2 , gegen H_1 und das Auftreten von \bar{I} für H_1 gegen H_2 . Zweitens wird klar, daß α und β nichts mit Wahrscheinlichkeiten zu tun haben, mit denen Hypothesen gelten. Aussagen über solche Wahrscheinlichkeiten sind nach Eintreten der Indizien nur möglich, wenn schon subjektive a priori Verteilungen gegeben waren. Das Eintreten von Indizien kann nur zur Revision schon vorhandener Wahrscheinlichkeitsverteilungen benutzt werden, nicht aber zu ihrer Bestimmung. Drittens läßt sich die in der Testtheorie auftretende Asymmetrie zwischen Hypothese und Alternative im Bayesschen Rahmen als unterschiedliche Vorbewertung deuten.

Resumé:

Unsere Darstellung darf nicht als Plädoyer gegen die klassische Testtheorie mißverstanden werden, im Gegenteil: Erst die Bayessche Betrachtungsweise schafft eine intuitive Grundlage, auf der ein fundiertes Verständnis der Testtheorie wachsen kann. Zu dieser Grundlage gehört, daß Entscheidungsprobleme früh angesprochen werden. Während man bei der Testtheorie die Verteilungen "komplizierter" Testgrößen benötigt, weswegen sich der Unterricht auch auf höheren

Stufen meist auf Binomialtests beschränkt, braucht man zum quantitativ fundierten Abwägen zwischen verschiedenen Hypothesen nach Bayes ausschließlich die Pfadregel. Die Hypothesen können dabei "beliebig kompliziert" sein (1.4.2, 1.5.2). Aus diesem Grunde eignet sich die Bayessche Regel bei geeigneter Darstellung auch schon auf unteren Stufen sehr gut zur Vorbereitung entscheidungstheoretischer Grundgedanken.

Abschließend sei mitgeteilt, daß bei der erwähnten Untersuchung der Kollegen A und B das Indiz I auftrat. Spätere Nachforschungen lieferten für den Softy-Parameter von Herrn A das 80% Konfidenzintervall $[0,52;0,64]$, für Herrn B das Intervall $[0,87;0,94]$. Das 80% Intervall des Autors betrug $[0,79;0,89]$. Alle Größen beziehen sich auf die gleiche Klasse.

2 Primäre Intuitionen, sekundäre Intuitionen und heuristische Modelle aus der Sicht Fischbeins

Der Aufbau der Stochastik auf der Grundlage eines subjektiven Wahrscheinlichkeitsbegriffs wurde schon im vorigen Kapitel unter Berufung auf das Brunersche Spiralprinzip mit seinen Forderungen nach Präfiguration und Fortsetzbarkeit begründet. Diese Theorien sind so bekannt, daß wir auf sie nicht näher einzugehen brauchen. Weniger Beachtung scheinen unterdessen Bruners Ausführungen zu Fragen des intuitiven Denkens zu finden. Sie haben ihre Fortsetzung in den lernpsychologisch orientierten Arbeiten Fischbeins zu Intuitionen und heuristischen Modellen. Wir wollen sie heranziehen, um das heuristische Vorgehen bei Begründung der Bayesschen Regel unter Einsatz des umgekehrten Wahrscheinlichkeitsbaumes zu rechtfertigen. Gleichzeitig werfen die Ausführungen Licht auf die Bedeutung der Modelle in den folgenden Kapiteln.

2.1 Primäre Intuitionen

Der Mensch besitzt auf Grund seiner Alltagserfahrung eine große Zahl von Intuitionen, die es ihm erlauben, auf Umweltsituationen adäquat zu reagieren. Was wir dabei mit dem Begriff der Intuition inhaltlich verbinden wollen, wird von Fischbein wie folgt umrissen:

"Our main hypothesis is that human cognition is fundamentally unitary. It is unitary by virtue of its original adaptive character, and by virtue of the fact that on any level it envisages a single reality and provides essentially concordant sets of information, even though they may be expressed in different codes. Consequently our starting-point is that intuition and intelligence (or intuition and reasoning) tell us about the same reality... ."

"The feeling of conviction, of certitude, which accompanies intuition, expresses a stabilized, and therefore long-verified mental organisation. This certitude or immediacy which characterizes intuition does not result, as Kant held, from

its a priori nature, but from the stability of patterns broadly verified through active contact with reality."

"It should be stressed that it is their close relation to effective action that causes intuitions to be so specific, as compared to other cognitive modalities. It is not their global nature, as such, that is the distinctive feature of intuitions, but the fact that they synthesize experience in the service of action. Intuitions is the means whereby cognition meets the requirements of speed, fluency and coherence of effective action. Intuitions, then, to this point of view, are nothing but intelligence in action. They develop along with the individual; they are changed and modified through contact with the reality which otherwise could not be effectively coped with." [Fischbein 1, S. 6 bis 14].

Wie sich diese Intuitionen im Bereich der Stochastik mit wachsendem Lebensalter (4 bis 12 Jahre) entwickeln, wurde in einer Reihe empirischer Untersuchungen erforscht ¹⁾. Sie haben in Ergänzung zu den ersten Ansätzen durch Piaget ergeben, daß die Entwicklung im Rahmen gewisser endogen bedingter Grenzen durch gezielte Umweltanregungen gefördert werden kann. In manchen Fällen sind solche Anregungen zur Vermeidung von später schwer korrigierbaren Fehlintuitionen unerlässlich ²⁾. Das läßt das Einsetzen des Stochastikunterrichts schon auf der Primarstufe ratsam erscheinen. Es ist klar, daß ein solcher Unterricht "unter Verzicht auf jeden formalen Apparat" [Müller/Wittmann, S. 241] "von guten Spielen lebt" [Winter zitiert nach Heitele 2, S.229], weil ja zu fördernde Intuition und Handlung in engstem Zusammen-

1) Vgl. etwa [Davies], [Doherty], [Falk/Falk/Levin], [Fischbein/Pampu/Minzat], [Goldberg], [Gratch], [Hoemann/Ross], [Leake], [Leffin], [Yost/Siegel/Andrews] und andere. Eine Übersicht über die Forschungsergebnisse findet sich in [Heitele 2, S. 146f] oder in [Kütting, S. 70ff].

2) Vgl. z.B. die Analyse zu sampling-effect und recency-effect in [Fischbein 1, S.11, S. 126] oder die Ausführungen zur Unabhängigkeit in [Heitele 2, S. 242ff].

hang miteinander stehen. Der an diesen Fragen interessierte Leser sei auf [Kütting] und die Arbeit "Didaktische Ansätze in Grundschule und Förderstufe" [Heitele 2, S. 229ff] verwiesen. Da sich unsere Ausführungen auf Kinder der anschließenden Entwicklungsstufe (Stadium der formalen Operationen) beziehen, wollen wir nur festhalten:

"Alles das, was das Kind im Alter von 12 Jahren über Zufall, Wahrscheinlichkeit, Kombinatorik ... weiß, ist gegründet auf vom Kind erlebte konkrete Handlungen. Die stochastischen Erfahrungen wurden an irgendwelchen Situationen (innerhalb oder außerhalb des Schulunterrichts, Anm. d. Verf.) gewonnen und führen zu Intuitionen, die man als Restsubstrat von vergessenen konkreten Situationen ansehen mag [Heitele 2, S. 160].

Auf Grund dieser Intuitionen sind die Schüler (etwa in den Beispielen 1.2.1, 1.3.1, 1.4.1) spontan bereit, subjektive Wahrscheinlichkeitsverteilungen zu formulieren bzw. nach Kenntnisnahme von Indizien zu revidieren. Da diese Intuitionen ohne systematische Schulung durch langfristigen Kontakt mit der Realität erworben wurden, bezeichnet man sie als primäre Intuitionen.

2.2 Sekundäre Intuitionen

Davon zu unterscheiden sind die sekundären Intuitionen. Sie werden in systematischen Lehr- und Lernprozessen erworben und erlauben es dem Individuum, die engen Grenzen zu überschreiten, die ihm die primären Intuitionen setzen, bzw. sie verfeinern die primären Intuitionen und geben ihnen den Charakter wissenschaftlicher Wahrheiten. "The formation of a natural intuitive substrate must be distinguished from the development of secondary intuitions which are the result of systematic instruction. Since the intuitive substrate of probabilistic thinking is relatively poor (and contradictory) the problem of the formation of secondary probabilistic intuitions is particularly important from the point of view of mathematics curricula." "Secondary intuitions ... enable

the individual to transcend primary cognitive acquisitions. They convey the products of social experience in the form, mostly, of scientific truths." [Fischbein 1, S. 18 u. S. 9]. Die Bayessche Regel

$$P(A_i|B) = P(A_i)P(B|A_i) / \sum_k P(A_k)P(B|A_k)$$

hat den Charakter einer solchen "wissenschaftlichen Wahrheit" und unsere heuristische Behandlung (1.2 bis 1.6) hat zum Ziel, sie in den Rang einer sekundären Intuition zu erheben, die in der Lage ist, die primären Intuitionen, welche in den subjektiven Revisionen zum Vorschein kommen, bewußt zu machen und zu präzisieren. Durch Deduktion aus den Axiomen kann der Bayesschen Regel jedenfalls nicht der Status einer sekundären Intuition gegeben werden.¹⁾ Beispiel:

Bei der Entwicklung einer Formel für die Berechnung von $P(A|B)$ aus den Größen $P(A)$, $P(\bar{A})$, $P(B|A)$ und $P(B|\bar{A})$ – man spricht hier von einem Umkehrproblem – geht man von der Zerlegung des Ergebnisraums Ω in die Ereignisse A und \bar{A} aus: $\Omega = A \cup \bar{A}$. Dann gilt:

$$\begin{aligned} B &= B \cap \Omega = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A}) \Rightarrow \\ \Rightarrow P(B) &= P(A \cap B) + P(\bar{A} \cap B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A}) \end{aligned}$$

Daraus gewinnt man die nach Bayes¹ benannte Formel für die Zerlegung von Ω in zwei Ereignisse:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}$$

Allgemein gilt

Satz 6.2.

Sei $\{A_1, A_2, \dots, A_n\}$ eine Zerlegung eines Ergebnisraums Ω , so gilt

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{k=1}^n P(A_k) \cdot P(B|A_k)} \quad (i \in \{1, 2, \dots, n\}; P(B) > 0).$$

(Aufgaben: siehe Anhang, S. 248)

¹ Thomas Bayes, 1702–1761

1) Natürlich schließt unsere Darstellung eine Deduktion nicht aus, doch sollte diese erst nach der intuitiven Einsicht in die Gültigkeit der Bayesschen Regel erfolgen.

Bruner denkt wohl an solche Beispiele, wenn er schreibt: "In den vorhergehenden Kapiteln war oft die Rede davon, wie wichtig es ist, daß ein Schüler die Gegenstände, mit denen er in Berührung kommt, intuitiv - im Gegensatz zu formal - versteht." "Leider hat der Formalismus im Schulunterricht die Intuition etwas abgewertet. Es kann von höchster Wichtigkeit sein, für ein intuitives Verständnis von Unterrichtsstoffen den Grund zu legen, ehe wir unsere Schüler mit den mehr traditionellen und formalen Deduktions- und Beweismethoden bekannt machen. [Bruner 1, S. 64 u. 67]. Und Fischbein präzisiert:

"Simple theoretical explanation is not sufficient to convert the information into stabilized acquisitions, involving the basic features of intuition." [Fischbein 1, S. 11].

Wie kann man nun erreichen, daß für fundamentale erachtete mathematische Zusammenhänge (hier das Bayessche Prinzip und seine Bedeutung für den Prozeß des induktiven Schließens) den Status von sekundären Intuitionen erhalten?

2.3 Modelle

"In our view, generative models (for example, tree diagrams, in the case of combinatorial operations) are the best teaching devices for the construction of secondary intuitions." [Fischbein 1, S. 129].

Das ist - schon vorweggenommen - die Antwort, die Fischbein auf die oben gestellte Frage gibt. Die Rolle von didaktischen Modellen umreißt Fischbein in seinem Artikel "Image and concept in learning mathematics" wie folgt: "The essential roles of a model are:

- (a) to facilitate the interpretation of certain given facts;
- (b) to help solve problems according to the original facts."

Besonders wertvoll sind solche Modelle, die intuitive Züge tragen - d.h. deren Beziehung zur Realität selbst den Charakter einer Intuition besitzt - und die gleichzeitig heuristische Eigenschaften haben. Damit ist gemeint, daß sie ein

Problem nicht nur darstellen, sondern implizit auch Lösungswege generieren, die möglichst allgemein, also auch auf benachbarte Probleme übertragbar sind. (Deswegen nennt Fischbein heuristische Modelle auch generativ.)

"If a model is an intuitive one, it has an ... important advantage. The human mind is inclined, naturally, to visualize facts. We think more easily with images because we are used to thinking of material objects. An intuitive model offers a spatial version of different types of facts... . In order to be useful, the model must be essentially heuristic."

"It must have the capacity to stimulate the process of reasoning and permit its progress by their own means. In other words, in order to be really efficient, the model must be a generative one. Such a generative capacity is based on: consistent isomorphism with the original, natural inner consistency (which is equivalent to autonomy with regard to the original) and intrinsic structure." [Fischbein 2, S. 155 u. 164].

"A good generative model possesses also a specific proliferative capacity. A good model is open and flexible enough to be able to inspire the invention of new related models adapted to new types of problems. [Fischbein 2, S.159].

Wir machen in diesem Buch von solchen Modellen wiederholt Gebrauch. So ist die Münzdarstellung von Wahrscheinlichkeitsverteilungen im nächsten Kapitel ein enaktives generatives Modell, das Engel in abgewandelter Form für seinen Wahrscheinlichkeitsabakus einsetzt [Engel 4, 5]. Bei uns führt es auf den Faltungsabakus und wegen seiner generativen Eigenschaften in sehr naheliegender Weise auf den Varianzabakus. Auch die Ratterkiste trägt als Modell für die Brownsche Bewegung generative Züge, denn sie legt das mathematische Irrfahrtmodell nahe, erzwingt die sonst oft willkürlich erscheinende Normierung der Binomialverteilungen und den stark verallgemeinerungsfähigen Übergang von der Differenzgleichung zur Differentialgleichung. Der elastische Stab hat als Modell zwar intuitive Züge, ist aber in Bezug auf die Stochastik weniger generativ als das Münzmodell.

2.4 Der umgekehrte Wahrscheinlichkeitsbaum

Nach diesem Vorgriff auf folgende Kapitel wollen wir auf die Bayessche Regel und das umgekehrte Baumdiagramm zurückkommen. Zum normalen Baumdiagramm äußert sich Fischbein: "It is, in fact, more than a graphical transcription of given facts. ... It is a heuristic model because it is internally consistent and generative." "... it can correctly represent an unlimited number of different situations, using a limited number of elements and rules." [Fischbein 2, S. 158 u. 155].

Bei praktischen Erprobungen zu den Kapiteln 1, 3, 5 und 7 gewinnt man den Eindruck, daß auch der umgekehrte Wahrscheinlichkeitsbaum ein generatives Modell darstellt, und daß es sehr hilfreich ist, ihn im Unterricht begrifflich klar vom normalen Wahrscheinlichkeitsbaum zu trennen. Der Unterschied besteht in einem Blickpunktwechsel:

Während man beim normalen Wahrscheinlichkeitsbaum die "Ursache" als feststehend auffaßt und von ihr ausgehend die möglichen "Wirkungen" und deren Wahrscheinlichkeiten betrachtet (deduktiver Standpunkt), wendet man beim umgekehrten Wahrscheinlichkeitsbaum den Blick von einer festen "Wirkung" auf mögliche "Ursachen" und bestimmt bzw. vergleicht die Wahrscheinlichkeiten, mit denen sie die "Wirkung" zur Folge haben (induktiver Standpunkt). Die "gedanklichen Pfeile" führen nicht mehr von der Spitze des Baumes weg, sondern zu ihr hin. Weiter geht - obwohl die im letzten Zitat angesprochenen "Elemente" die gleichen sind - in den umgekehrten Wahrscheinlichkeitsbaum zusätzlich zur Multiplikationsregel auch die Additionsregel ein. Und was Fischbein über den generativen Charakter des normalen Wahrscheinlichkeitsbaumes schreibt, gilt auch für den umgekehrten:

"The model is actually active as an intellectual tool: it solves the problem, and not only describes the solution. With such a model we can learn to think effectively and understand actively." [Fischbein 2, S. 158].

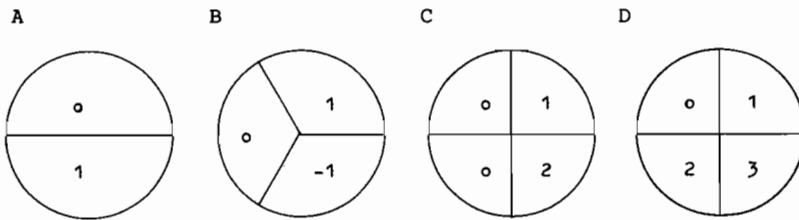
Am eindrucksvollsten wird das belegt durch die in 1.3 festgehaltene Beobachtung, daß Schüler in der Lage sind, die Bayessche Regel aus dem Baumdiagramm herauszulesen. Auch der Forderung nach Universalität wird das umgekehrte Baumdiagramm gerecht. Es repräsentiert nicht nur, wie wir sahen,

- (a) das Bayessche Prinzip,
 - (b) die Idee des Hypothesentestens,
- sondern leistet auch ausgezeichnete Dienste bei
- (c) der Berechnung von Faltungen und
 - (d) der rekursiven Berechnung endlicher Markoffketten.

Wir wollen (c) und (d) nur kurz erläutern, da wir hierauf später näher eingehen werden.

Zu (c): Rekursive Berechnung von Faltungen

Ein Glücksspieler dreht nacheinander die Räder A bis D, dann noch einmal A bis D und summiert seine 8 Punktzahlen. Die Summe wird zwischen -2 und 14 schwanken.

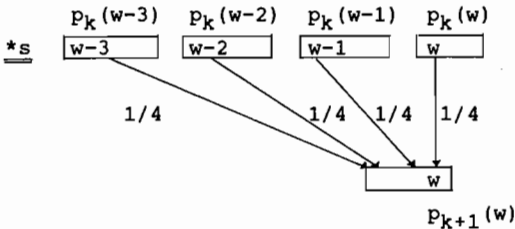


Verteilung p Verteilung q Verteilung r Verteilung s

Welche Wahrscheinlichkeitsverteilung p_8 haben wir für die Summe aus den 8 Punktzahlen anzusetzen, wenn für die Verteilungen p, q, r und s Laplace-Hypothesen gerechtfertigt sind? Würden wir einen normalen achtstufigen Wahrscheinlichkeitsbaum benutzen, hätten wir $2 \cdot 3 \cdot 4 \cdot 4 \cdot 2 \cdot 3 \cdot 4 \cdot 4 = 9216$ Äste zu verarbeiten. Bei einem Abstand von 0,5 cm wären das 46 m Papier für die letzte Stufe. Die rekursive Berechnung ist wesentlich effektiver: Ihr Kernstück ist die Frage: Auf welchen Wegen entsteht aus den Ergebnissen des k-fachen Summen-

experimentes im $k+1$ -fachen Summenexperiment die Augenzahl w ?
Welche Wahrscheinlichkeit haben diese Wege?

Sie wird (etwa für $k+1=4$ und $k+1=8$, wenn der $k+1$ -te Summand durch das Glücksrad D gegeben ist) durch den umgekehrten Wahrscheinlichkeitsbaum



beantwortet, der die Rekursionsformel

$$\begin{aligned}
 P_{k+1}(w) &= (1/4)P_k(w-3) + (1/4)P_k(w-2) + (1/4)P_k(w-1) + (1/4)P_k(w) \\
 &= (1/4)(P_k(w-3) + P_k(w-2) + P_k(w-1) + P_k(w))
 \end{aligned}$$

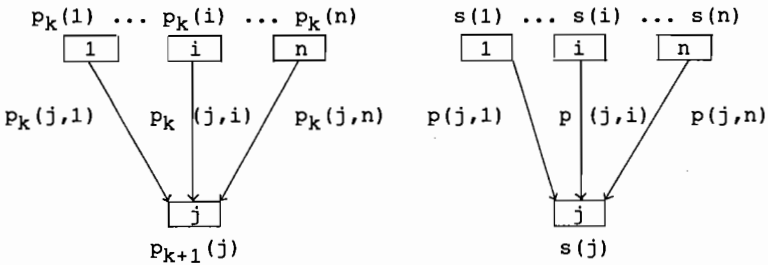
symbolisiert. Besonders einfach ist sein Einsatz, wenn Laplace-Hypothesen vorliegen. Wie unser Beispiel zeigt, kann man sich dann auf Verarbeitung der Zähler zu $p_k(w)$ beschränken und die Nenner gesondert notieren. Anschaulich gesprochen (vgl. die folgende Tabelle) erhält man die Zähler der $k+1$ -ten Summenverteilung einfach durch "Entlangsschieben" des umgekehrten Wahrscheinlichkeitsbaumes zum $k+1$ -ten Summanden an der k -ten Summenverteilung und entsprechender Addition. Optisch eindrucksvoll demonstriert man das mit verschiebbaren Folien am Tageslichtprojektor. Die umgekehrten Wahrscheinlichkeitsbäume werden zu ikonischen "Faltungsmaschinen". In Kapitel 7 wird die rekursive Berechnung den Ansatz zur Herleitung des zentralen Grenzwertsatzes liefern. Zur Veranschaulichung seiner Aussage enthält die untenstehende Tabelle neben p_8 auch die Werte der zugehörigen Normalverteilung $\varphi_{\mu, \sigma}$.

Rekursive Berechnung der Verteilung $p_8 = p^*p^*q^*r^*s^*p^*q^*r^*s$ als achtfache Faltung

	Nenner	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
$p_1 = p$	2				1	1													
$* q$																			
$p_2 = p^*q$	6		1	2	2	1													
$* r$																			
$p_3 = p^*q^*r$	24		2	5	7	6	3	1											
$* s$																			
$p_4 = p^*q^*r^*s$	96		2	7	14	20	21	17	10	4	1								
$* p$																			
$p_5 = p^*q^*r^*s^*p$	192		2	9	21	34	41	38	27	14	5	1							
$* q$																			
$p_6 = p^*q^*r^*s^*p^*q$	576	2	11	32	64	96	113	106	79	46	20	6	1						
$* r$																			
$p_7 = p^*q^*r^*s^*p^*q^*r$	2304	4	24	77	171	288	386	421	377	277	165	78	28	7	1				
$* s$																			
$p_8 = p^*q^*r^*s^*p^*q^*r^*s$	9216	4	28	105	276	560	922	1266	1472	1461	1240	897	548	278	114	36	8	1	
p_8		.000	.003	.011	.030	.061	.100	.137	.160	.159	.135	.097	.059	.030	.012	.004	.001	.000	
$\varphi_{SS}, 2,33$.001	.004	.012	.028	.057	.097	.137	.163	.163	.137	.097	.057	.028	.012	.004	.001	.000	

Zu (d):

$\{1, \dots, n\}$ sei der Zustandsraum einer endlichen Markoffkette, $p_k(i)$, $i=1, \dots, n$ sei die Wahrscheinlichkeit, daß sich das System zur Zeit k im Zustand i befindet. $p_k(i, j)$ sei die Übergangswahrscheinlichkeit vom Zustand i in den Zustand j zur Zeit k . Zur Berechnung von $p_{k+1}(j)$ verwendet man analog zu (c) das linksstehende Diagramm:



Es stellt die Multiplikation des Verteilungsvektors p_k mit einer Zeile der Übergangsmatrix dar. Die stationäre Verteilung $s(i)$, $i=1, \dots, n$ ergibt sich bei nicht zeitabhängigen Übergangswahrscheinlichkeiten ebenso aus dem rechtsstehenden Diagramm. Es besagt, daß s ein "Fixvektor" der Übergangsmatrix ist (vgl. auch 5.6).

Zu (a), (c) und (d) gehören die Formeln

$$(a) \quad p(A_i|B) = \frac{p(B|A_i) \cdot p(A_i)}{\sum_j p(B|A_j) \cdot p(A_j)}$$

$$(c) \quad p_{k+1}(i) = (p_k * p_1)(i) = \sum_{l+m=i} p_k(l) p_1(m)$$

$$(d) \quad p_{k+1}(i) = \sum_j p_k(i, j) p_k(j) \quad \text{und}$$

$$s(i) = \sum_j p(i, j) s(j).$$

Sie leiten sich aus den entsprechenden Baumdiagrammen her und formalisieren nur die am kognitiven Schema des Baumdiagramms intuitiv einsichtig gewordenen Zusammenhänge.

Was Fischbein über normale Baumdiagramme schreibt, trifft auch auf unsere umgekehrten zu:

"What the diagram provided was a principle of construction which synthesized induction and deduction in a single operation - a working formula which was at the same time a conceptual, cognitive schema. Extrapolation follows naturally from such a schema; it follows from the working formula itself but not from the mathematical formula which derives from it, and which only formalizes an accepted truth."
[Fischbein 1, S. 116].

Aus dem Blickwinkel der Brunerschen Darstellungstheorie kann der umgekehrte Wahrscheinlichkeitsbaum umschrieben werden als ein ikonisches Hilfsmittel, das die enaktive Berechnung von Markoffketten und Faltungen durch Engels Wahrscheinlichkeitsabakus bzw. unseren Faltungsabakus (3.2) verbindet mit der formalen Berechnung durch Matrizenalgebra und Faltungsprodukt. Sein Einsatz macht für Lernende von Anfang an deutlich, daß etwa das Pascalsche Dreieck nur der Spezialfall einer sehr allgemeinen Berechnungsmethode für Verteilungen zu Summenexperimenten ist und daß dieser Berechnungsmethode die gleiche Struktur zugrunde liegt wie der rekursiven Berechnung endlicher Markoffketten.

3 Der Varianzabakus

"Die Aussagen von Fachwissenschaftlern, u. a. Freudenthal und Dinges legen es überhaupt nahe, auf einen deduktiven Aufbau der Stochastik mit Hilfe der Mengenalgebra zu verzichten und stattdessen zu versuchen, das, was man als fundamental ansieht, unmittelbar zu elementarisieren." [Heitele 2, S. 220 f]. Kapitel 1 hat gezeigt, daß dieser Ratsschlag nicht notwendig zu "didaktischem Flickwerk" führt. Im Gegenteil gab er Anlaß zu einem curricularen Rahmen, der durch inhaltlich tragende Ideen strukturiert und in sich konsistent ist. Der Leser mag jedoch einwenden, daß - trotz des schönen Rahmens - Stochastik nicht nur aus ihren induktiven Teilen besteht, und daß Mengenalgebra, Produkträume und Zufallsvariable spätestens beim Themenkreis Varianz, Tschebyscheff-Ungleichung, schwaches Gesetz der großen Zahlen unverzichtbar sind. Das folgende Kapitel zeigt jedoch, daß es durchaus möglich ist, auch diese zentralen Themen schon auf tiefen Curriculumstufen ¹⁾ ohne Formalismus anzugehen und Brunerschen Leitsätzen entsprechend "intellektuell ehrlich" zu präfigurieren.

3.1 Vorbemerkung

Während es sehr einleuchtend ist, den Mittelwert einer relativen Häufigkeitsverteilung oder den Erwartungswert einer Wahrscheinlichkeitsverteilung zu definieren als

$$\mu = \sum_{i=1}^n x_i q_i ,$$

erscheint es recht willkürlich, die Varianz

$$V = V_2 = \sum_{i=1}^n (x_i - \mu)^2 q_i$$

1) Im Anschluß an die Orientierungsstufe, wenn Erfahrungen mit der Bruchrechnung vorliegen.

als Maß für die Streuung zu wählen. Ist nicht

$$V_1 = \sum_{i=1}^n |x_i - \mu| q_i$$

viel naheliegender? Tatsächlich könnte man ebensogut

$$V_p = \sum_{i=1}^n |x_i - \mu|^p q_i$$

oder sogar die Spannweite

$$V_\infty = \sup_{i=1}^n |x_i - \mu|$$

als Maß für die Streuung heranziehen. Daß man sich in der Mathematik für V_2 entscheidet, hat zwei Gründe, einen gewichtigen, der durch das Stichwort "Additivität" gekennzeichnet ist und einen ästhetischen, der als Minimalitätsbedingung charakterisiert werden kann. Wir wollen das erläutern und erinnern dazu an einige gängige Notationen:

Für einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, q) (Ω Grundmenge, \mathcal{A} σ -Algebra, q Wahrscheinlichkeitsmaß) bezeichnet $(L_p, \|\cdot\|_p)$ den Banachraum der p -integrierbaren Funktionen $X: \Omega \rightarrow \mathbb{R}$ mit $\|X\|_p = \sqrt[p]{\int |X|^p d q} = \sqrt[p]{\int |X|^p d q}$ ($1 \leq p < \infty$). Für $p=2$ hat man speziell $\|X\|_2 = \sqrt{\langle X|X \rangle} = \sqrt{\int X^2 d q}$, wobei $\langle X|Y \rangle = \int XY d q$ das Skalarprodukt in L_2 ist. Für $p=\infty$ sei $(L_\infty, \|\cdot\|_\infty)$ der Banachraum der beschränkten \mathcal{A} -meßbaren Funktionen mit $\|X\|_\infty = \sup_{\omega \in \Omega} |X(\omega)|$.

Die Argumente für V_2 können wir nun wie folgt formulieren:

3.1.1 Additivität

Unabhängige Zufallsvariable X und Y sind unkorreliert, d.h. $X - \mu(X)$ und $Y - \mu(Y)$ stehen bezüglich des L_2 -Skalarproduktes orthogonal. Der Satz des Pythagoras liefert daher:

$$\begin{aligned} V_2(X+Y) &= \|X - \mu(X) + Y - \mu(Y)\|^2 = \|X - \mu(X)\|^2 + \|Y - \mu(Y)\|^2 \\ &= V_2(X) + V_2(Y) . \end{aligned}$$

Für $p \neq 2$ geht die Additivität $V_p(X+Y) = V_p(X) + V_p(Y)$ verloren, wie ein einfaches Gegenbeispiel zeigt: Sind nämlich X und Y $B_{1,1/2}$ verteilt, so ist $X+Y$ verteilt nach $B_{2,1/2}$. Man hat

dann $V_p(X) = V_p(Y) = (1/2)^p$, $V_p(X+Y) = 1/2$ und $V_p(X+Y) = V_p(X) + V_p(Y)$
 $\Leftrightarrow 2(1/2)^p = 1/2 \Leftrightarrow p=2$. Dies macht V_p für $p \neq 2$ als Streuungsmaß
 unbrauchbar. (Man denke nur an die Normierung bei Formulie-
 rung von Grenzwertsätzen für Summen unabhängiger Zufallsvari-
 ablen.)

3.1.2 Minimalitätseigenschaft des Erwartungswertes¹⁾

Wir identifizieren Zahlen $k \in \mathbb{R}$ mit konstanten Zufallsvariab-
 len $k: \Omega \rightarrow \mathbb{R}, \omega \mapsto k$ und suchen zu einer beliebigen Zufalls-
 variablen $X: \Omega \rightarrow \mathbb{R}$ nach einer konstanten, die von X minima-
 len Abstand hat, also $\|X-k\|_p$ minimiert.

Für $p=2$ stammt die Norm $\|\cdot\|_p$ von einem Skalarprodukt und k
 ist die eindeutige Projektion von X auf den eindimensionalen
 Unterraum $U = \langle 1 \rangle$ der konstanten Funktionen, also wegen

$$k = \langle X | 1 \rangle = \int X \cdot 1 \, d\mathbb{Q} = \int X \, d\mathbb{Q} = \mu(X)$$

mit dem Erwartungswert identisch. Der Abstand ist dann

$$\|X - \mu(X)\|_2 = \sqrt{\int (X - \mu(X))^2 \, d\mathbb{Q}} = \sqrt{V_2(X)} = \tilde{\sigma}_2(X),$$

also gerade die Standardabweichung.

Für $p \neq 2$ geht dieser Zusammenhang zwischen μ und $\tilde{\sigma}_p$ verloren.
 Wegen der Vollständigkeit von L_p und der Abgeschlossenheit
 von U gibt es zwar immer noch konstante Zufallsvariable k ,
 die $\|X-k\|_p$ minimieren; auch ist für $p \in]1; \infty[$ k immer noch
 eindeutig bestimmt²⁾, aber i.a. von μ verschieden. Für $p = 1$
 geht auch die Eindeutigkeit verloren³⁾.

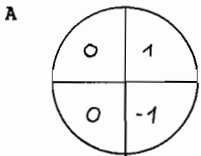
1) Diese Überlegung wurde angeregt durch [Scheid 1].

2) Für $1 < p < \infty$ folgt dies aus der gleichmäßigen Konvexität von
 L_p (vgl. z. B. [Hirzebruch/Scharlau, S. 74]).

3) k ist der nicht notwendig eindeutig bestimmte Median.

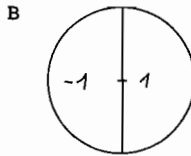
3.2 Der Faltungsabakus

Spieler X dreht nacheinander folgende Glücksräder



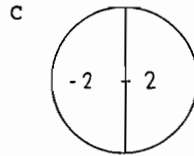
Laplace-Hypothese

ω	-1	0	1
$p(\omega)$	1/4	1/2	1/4



Laplace-Hypothese

ω	-1	1
$q(\omega)$	1/2	1/2



Laplace-Hypothese

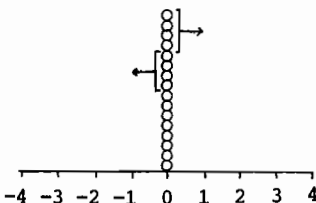
ω	-2	2
$r(\omega)$	1/2	1/2

und addiert die drei Punktzahlen, die er erhalten hat. Y dreht viermal hintereinander das Glücksrad B und addiert ebenfalls seine Ergebnisse. Beide Spieler erhalten Punktzahlen zwischen -4 und 4, beide erhalten im Mittel 0 Punkte. Wer eine Punktzahl erhält, die um 3 oder mehr von 0 abweicht, ($\pm 3, \pm 4$), bekommt einen Preis. Möchtest Du lieber Spieler X oder Y sein?

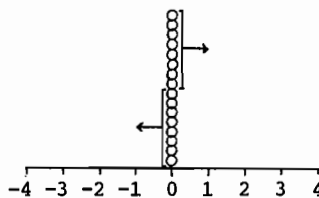
Um diese Frage beantworten zu können, benötigen wir die Wahrscheinlichkeitsverteilungen zum Summenexperiment von A, B und C bzw. zum Vierfach-Summenexperiment von B. Wir bezeichnen sie mit $p*q*r$ bzw. $q*q*q*q$ und benutzen zu ihrer Bestimmung 16 Münzen, von denen jede einzelne die Wahrscheinlichkeit 1/16 repräsentieren soll.

Vor Spielbeginn haben beide Spieler sicher (mit Wahrscheinlichkeit 16/16) 0 Punkte:

Spieler X (Verteilung n)



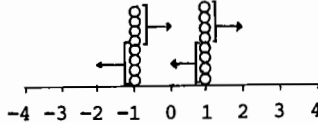
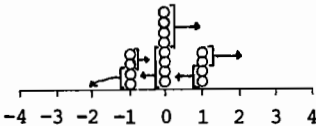
Spieler Y (Verteilung n)



Beim Drehen des ersten Glücksrades A bzw. B teilt sich dieser Münzturm (d. h. die Wahrscheinlichkeit 1) gemäß der Verteilung p bzw. q auf. Das wird durch die Pfeile in den obigen Skizzen angedeutet.

Spieler X (Verteilung p)

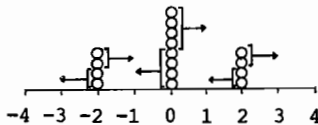
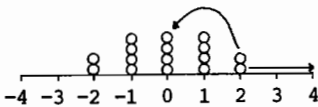
Spieler Y (Verteilung q)



Hat ein Spieler durch Drehen des ersten Glücksrades die Zahl 1 erreicht (X mit Wahrscheinlichkeit $1/4$, Y mit Wahrscheinlichkeit $1/2$), so wird er durch das Zusatzexperiment B in der Hälfte aller Fälle einen Punkt mehr (2), in der Hälfte der Fälle einen Punkt weniger (0) erhalten, d. h. beim Drehen des Glücksrades teilt sich der Turm der Stelle 1 auf gemäß der Verteilung q . Weil das Ergebnis das Glücksrad B liefert, unabhängig ist vom Ergebnis des ersten Glücksrades, gilt eine analoge Überlegung auch für die übrigen Türme. Für die Verteilungen $p*q$ bzw. $q*q$ erhält man (wie oben durch Pfeile angedeutet):

$p*q$

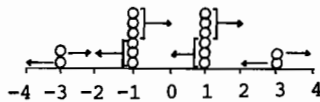
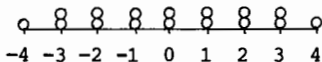
$q*q$

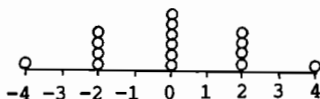


Nochmaliges Drehen und Aufteilen gemäß r bzw. q liefert:

$p*q*r$

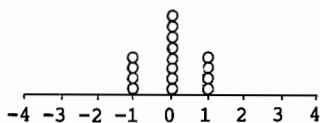
$q*q*q$



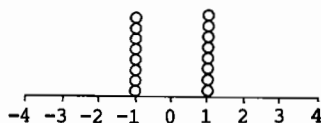
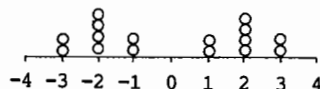
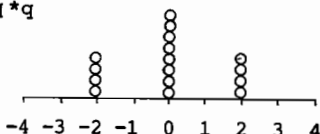
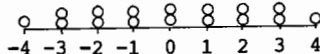
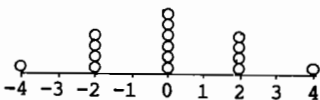
$q*q*q*q$ 

Anmerkung: Würde man zuerst das Experiment A, dann C und zuletzt B ausführen, so ergäbe sich linksstehende Folge von Wahrscheinlichkeitsverteilungen. Ein Vergleich mit obigen Skizzen liefert $(p*q)*r = (p*r)*q$.

p



q

 $p*r$  $q*q$  $(p*r)*q$  $(q*q)*(q*q)$ 

Wie die rechtsstehende Folge von Verteilungen belegt, liefert auch $(q*q)*(q*q)$ dasselbe Ergebnis wie $((q*q)*q)*q$. Kommutativität und Assoziativität der Faltungsoperation sind durch solche Beispiele auch auf unteren Curriculumstufen leicht zu entdecken.

Natürlich lassen sich all diese Operationen auch ikonisch unter Benutzung des umgekehrten Wahrscheinlichkeitsbaumes (2.5) durchführen:

	-4	-3	-2	-1	0	1	2	3	4		-4	-3	-2	-1	0	1	2	3	4
p				1/4	1/2	1/4				q				1/2		1/2			
*r	$\swarrow \quad \searrow$ 1/2 1/2									*q	$\swarrow \quad \searrow$ 1/2 1/2								
p*r			1/8	1/4	1/4	1/4	1/8			q*q			1/4		2/4		1/4		
*q	$\swarrow \quad \searrow$ 1/2 1/2									*(q*q)	$\swarrow \quad \searrow$ 1/4 1/2 1/4								
(p*r)*q	1/16	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/16	(q*q)*(q*q)	1/16		4/16		6/16		4/16		1/16

Die Münzen übernehmen bisher nur die Rolle einer simultanen "Divisions- und Additionsmaschine". Ihre generative Kraft in Bezug auf ein Streuungsmaß werden sie erst in 3.3 entfalten.

Doch zurück zur eingangs gestellten Frage: "Möchtest Du lieber Spieler X oder Y sein?" Lieber Spieler X, denn er erhält mit 6/16 gegenüber 2/16 dreimal so oft die erwünschten großen Abweichungen vom Mittelwert 0. Modifiziert man die Gewinnregel, und verteilt man einen Preis bei Abweichungen um mindestens 2 Punkte vom Erwartungswert 0, dann haben X und Y gleiche Gewinnchancen; bei Abweichungen um 1 oder mehr Punkte ist jedoch wieder X im Vorteil. Daher scheint die Aussage gerechtfertigt, daß die Punktzahlen des Spielers X eine größere "Streuung" ¹⁾ aufweisen als diejenigen des Spielers Y.

1) Wir verwenden den Ausdruck "Streuung" zunächst rein intuitiv. Später stellt sich heraus, daß das Streuungsmaß, welches durch den folgenden Abakus definiert wird, mit der Varianz identisch ist. Daher der Name Varianzabakus.

3.3 Der Varianzabakus

Wir wollen nun versuchen, ein Maß zum Vergleich von Streuungen beliebiger Verteilungen zu definieren. Benutzen wir dafür den Buchstaben S , so sollte nach der eben gemachten Bemerkung gelten

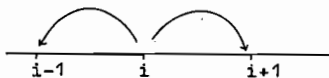
$$(1) \quad S(q*q*q*q) < S(p*q*r).$$

Weiter sollte S die Beobachtung widerspiegeln, daß rein intuitiv die Streuung von Verteilungen zu Summenexperimenten (wie bei den Spielern X und Y oder den Binomialverteilungen) mit wachsender Summandenzahl zunimmt. Wir hätten neben (1) die Gültigkeit folgender Ungleichungen zu fordern:

$$(2) \quad 0=S(n) < S(p) < S(p*q) < S(p*q*r)$$

$$(3) \quad 0=S(n) < S(q) < S(q*q) < S(q*q*q) < S(q*q*q*q).$$

Aber um welchen Betrag nimmt S bei jedem Schritt dieser Ungleichungskette zu? Eine plausible Antwort finden wir, wenn wir bedenken, daß die Verteilungen in (2) und (3) über den Faltungsabakus durch eine eindeutig bestimmte Anzahl von Münz-Doppelsprüngen auseinander hervorgehen, oder - was das gleiche ist - durch eine wachsende Zahl solcher Sprünge aus der Anfangsverteilung n "aufgebaut" werden können. So entsteht etwa $q*q$ aus q durch insgesamt 8 und aus n durch insgesamt 16 Doppelsprünge von der Form

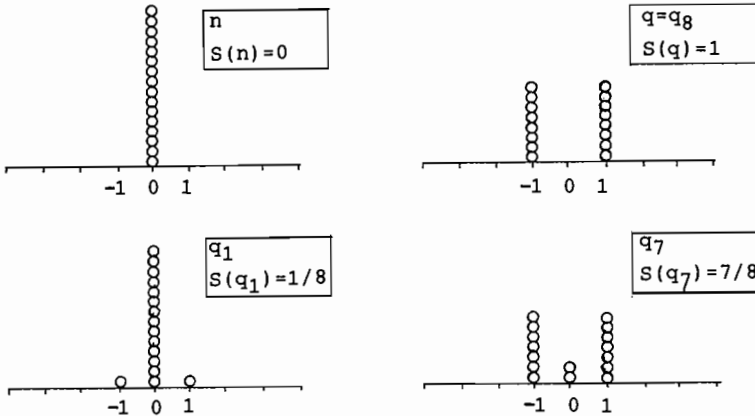


um je eine Einheit. p entsteht aus n durch 4, $p*q$ entsteht aus n durch 12 Doppelsprünge der gleichen Form. $p*q*r$ entsteht aus $p*q$ durch 2 Doppelsprünge von der Form



um je zwei Einheiten. Was liegt näher, als unser Streu-

ungsmaß S in Übereinstimmung mit (2) und (3) über diese Sprunzzahlen zu definieren? Wenn wir eine Einheit festlegen durch $S(q)=1$, dann haben wir der unten aufgeführten Verteilung q_1 den Streuungswert $1/8$ zuzuordnen:



Während nämlich q aus n durch 8 Doppelsprünge um je eine Einheit hervorgeht, entsteht q_1 aus n nur durch einen dieser Doppelsprünge. Allgemeiner sollte gelten $S(q_i)=i/8$ (insbesondere wegen $p=q_4$: $S(p)=1/2$), und wir halten als Ergebnis unserer bisherigen Überlegungen fest:

Provisorische Regel: Ein Doppelsprung um je eine Einheit erhöht den Streuungswert S (einer durch 16 Münzen dargestellten Wahrscheinlichkeitsverteilung) um $1/8$.

Sie liefert mit $S(q)=8/8=1$, $S(q*q)=16/8=2$, $S(q*q*q)=24/8=3$, $S(q*q*q*q)=32/8=4$, $S(p)=4/8=0,5$, $S(p*q)=12/8=1,5$ die ersten Indizien für die Additivität des Streuungsmaßes S .

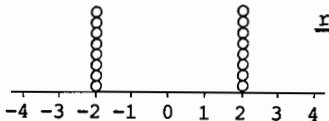
Zur Feststellung der Streuungsmaßzahl einer beliebig vorgegebenen Verteilung, von der wir nicht wissen, ob sie überhaupt durch Faltungen entstanden ist, wird man natürlich nicht mit n starten, sondern mit dieser Verteilung. Man wird versuchen, durch Doppelsprünge in umgekehrter Richtung - jeweils eine Einheit aufeinander zu - die Streuung abzu-

bauen, bis sie zu Null geworden ist. Dabei ist darauf zu achten, daß die beiden Münzen des umgekehrten Doppelsprunges jeweils auf derselben Stelle landen.

3.3.1 Beispiel 1

Als Beispiel legen wir nach der eben aufgestellten Regel $S(p*q*r)$ durch eine Folge von umgekehrten Doppelsprüngen um je eine Einheit fest. Der umseitigen Tabelle entnimmt man, daß $p*q*r$ durch 44 umgekehrte Doppelsprünge um je eine Einheit in n überführt werden kann. Mit $S(p*q*r)=44/8=5,5$ erhalten wir ein Ergebnis, das mit der eingangs erhobenen Forderung $S(p*q*r) > S(q*q*q*q)$ in Einklang steht. (Der Leser überzeuge sich davon, daß $p*q*r$ tatsächlich die Varianz 5,5 hat.)

Die Anwendung der Regel ist aber durch Beschränkung auf Doppelsprünge um je eine Einheit recht langwierig, ja der Verteilung



können wir auf diese Weise gar keinen Streuungswert $S(r)$ zuordnen, wir bräuchten Doppelsprünge um 2 Einheiten. Doch um welchen Betrag reduzieren sie S ?

Kreative Schüler können durchaus auf die Idee kommen, die in den Beispielen beobachtete Additivität zu nutzen: Da wir für p, q und $p*q*r$ die Streuungsmaße kennen, erhalten wir mit

$$S(p*q*r) = S(p) + S(q) + S(r), \quad \text{also} \\ 5,5 = 0,5 + 1 + S(r)$$

für r den Streuungswert 4. Da r durch 8 Doppelsprünge um je 2 Einheiten auf n reduziert werden kann, würde jeder einzelne dieser Sprünge den Streuungswert S um $(1/8) \cdot 4 = (1/8) \cdot 2^2$ reduzieren. Dieses Resultat weist den Weg zu einer "kühnen" Verallgemeinerung der provisorischen Regel:

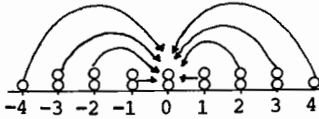
Zu Beispiel 1: Bestimmung von $S(p^*q^*r)$

Verteilung	Sprungbewegung	Veränderung von S
<p><u>p^*q^*r</u></p>	2 Doppelsprünge um 1 Einheit	-2 (1/8)
	4 Doppelsprünge um 1 Einheit	-4 (1/8)
	2 Doppelsprünge um 1 Einheit	-2 (1/8)
	2 Doppelsprünge um 1 Einheit	-2 (1/8)
	4 Doppelsprünge um 1 Einheit	-4 (1/8)
	6 Doppelsprünge um 1 Einheit	-6 (1/8)
	4 Doppelsprünge um 1 Einheit	-4 (1/8)
	4 Doppelsprünge um 1 Einheit	-4 (1/8)
	8 Doppelsprünge um 1 Einheit	-8 (1/8)
	8 Doppelsprünge um 1 Einheit	-8 (1/8)
		Summe: -44/8 = -5,5

Ergebnis: $S(p^*q^*r) = 5,5$

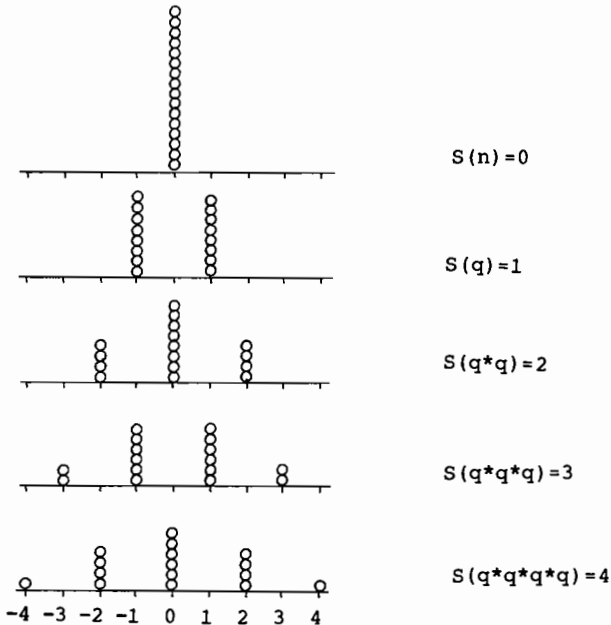
Ein Doppelsprung um k Einheiten reduziert den Streuungswert S einer (durch 16 Münzen dargestellten) Wahrscheinlichkeitsverteilung um $(1/8) \cdot k^2$.

Und tatsächlich liefert auch diese Verallgemeinerung (wie in folgender Skizze angedeutet) für p^*q^*r :



$$S(p^*q^*r) = (1/8)4^2 + (2/8)3^2 + (2/8)2^2 + (2/8)1^2 = 5,5.$$

Wir geben noch einen einfacheren Weg zur Verallgemeinerung unserer provisorischen Regel an. Betrachten wir die Verteilungen q , q^*q , q^*q^*q , $q^*q^*q^*q$, deren Streuungswerte schon bestimmt wurden,



so fällt uns auf, daß $q*q$ nicht nur durch 16 Doppelsprünge um je 1 Einheit, sondern auch durch 4 Doppelsprünge um je 2 Einheiten in n überführt werden kann:

Ein Doppelsprung um je 2 Einheiten reduziert S um
 $4/8 = (1/8) \cdot 2^2$.

Ebenso kann man $q*q*q$ durch 6 Doppelsprünge um je 1 Einheit und 2 Doppelsprünge um je 3 Einheiten in n überführen. Da die 6 Doppelsprünge um je eine Einheit die Streuung $S(q*q*q)$ um $6(1/8)$ auf $18/8$ reduzieren, bleiben für die beiden Doppelsprünge um je 3 Einheiten nur noch $18/8$ übrig, also:

Ein Doppelsprung um je 3 Einheiten reduziert S um
 $9/8 = (1/8) \cdot 3^2$.

Das Überführen von $q*q*q*q$ in n durch 4 Doppelsprünge um je 2 Einheiten und einen Doppelsprung um je 4 Einheiten liefert völlig analog:

Ein Doppelsprung um je 4 Einheiten reduziert S um
 $16/8 = (1/8) \cdot 4^2$.

Wenn wir q nicht durch 16 sondern allgemein durch N Münzen darstellen, dann haben wir zur Überführung von q in n (statt 8) $N/2$ Doppelsprünge um je eine Einheit durchzuführen. Wegen Unabhängigkeit des Streuungswertes $S(q)=1$ von der Münzdarstellung muß jeder dieser Sprünge S um $(2/N) \cdot 1^2$ (statt $(1/8) \cdot 1^2$) reduzieren. Entsprechendes gilt für die anderen Sprungweiten und wir fixieren als Ergebnis unserer bisherigen Überlegungen die Spielregeln des Varianzabakus, eines Spiels zur Festlegung eines Streuungswertes S für eine Wahrscheinlichkeitsverteilung, die durch N Münzen dargestellt ist:

- (1) Versuche durch eine Folge von symmetrischen Doppelsprüngen (die jeweils auf derselben Stelle enden müssen), alle Münzen auf einem Punkt zu konzentrieren.
- (2) Jeder Doppelsprung um k Einheiten reduziert S um $(2/N) \cdot k^2$.
- (3) Addiere alle diese Beträge und Du erhältst den Streuungswert S der ursprünglichen Verteilung.

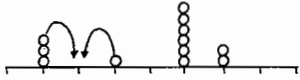


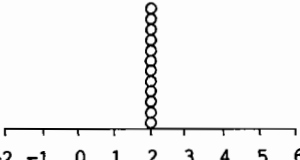
Wir wollen das Funktionieren des Abakus, insbesondere die Unabhängigkeit seines Ergebnisses vom Spielverlauf testen an einer nicht symmetrischen Verteilung, die durch $N=12$ Münzen dargestellt ist:

3.3.2 Beispiel 2 (erster Spielverlauf)

Verteilung	Sprungbewegung	Veränderung von S
	3 Doppelsprünge um 2 Einheiten	$3(1/6) 2^2$
	3 Doppelsprünge um 1 Einheit	$-3(1/6) 1^2$
	2 Doppelsprünge um 1 Einheit	$+2(1/6) 1^2$
	2 Doppelsprünge um 2 Einheiten	$-2(1/6) 2^2$
		Summe: $-21/6 = -3,5$

Ergebnis: Die Streuung der ursprünglichen Verteilung beträgt $S=3,5$.

Andere Spielverläufe führen zum gleichen Ergebnis:

Verteilung	Sprungbewegung	Veränderung von S
	1 Doppelsprung um 1 Einheit	$-(1/6) 1^2$
	4 Doppelsprünge um 2 Einheiten	$-4(1/6) 2^2$
	4 Doppelsprünge um 1 Einheit	$-4(1/6) 1^2$
		Summe: $-21/6 = -3,5$

Das Ergebnis $S=3,5$ stimmt mit demjenigen des ersten Spielverlaufs überein.

Der Erwartungswert der ursprünglichen Verteilung liegt bei

$$\mu = -1(3/12) + 1(1/12) + 3(6/12) + 4(2/12) = 24/12 = 2,$$

also gerade dort, wo der letzte Münzturm steht, und für den Leser sei angemerkt, daß ihre Varianz

$$\begin{aligned} v &= (-1-2)^2(3/12) + (1-2)^2(1/12) + (3-2)^2(6/12) + (4-2)^2(2/12) \\ &= 42/12 = 3,5 \end{aligned}$$

tatsächlich wieder mit dem Ergebnis des Abakus übereinstimmt. (Die Schüler führen diese Kontrolle erst später (3.4) durch, wenn sie die Formel für die Varianz aus den Spielregeln hergeleitet haben.)

Unser Abakus "geht für jede diskrete Verteilung (über \mathbb{Z}) auf", d.h. es ergibt sich zum Schluß ein Turm auf der Stelle des Erwartungswertes, sofern dieser eine ganze Zahl ist. Andernfalls kann man das Spiel stets so weit spielen, bis zwei eindeutig bestimmte Türme nebeneinander (auf zwei ganzzahligen Stellen a , $a+1$ des Zahlenstrahls) stehen.

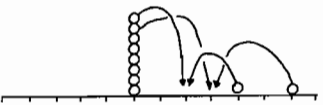
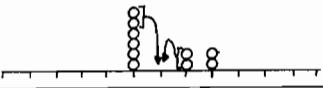
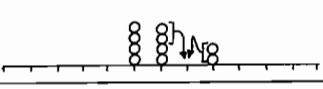
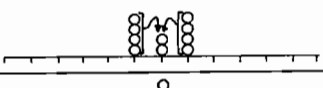
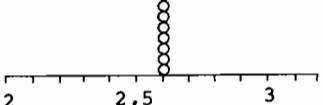
3.3.3 Beispiel 3 (mit $N=10$ Münzen)

Verteilung	Sprungbewegung	Veränderung von S
	1 Doppelsprung um 2 Einheiten	$-(1/5) 2^2$
	1 Doppelsprung um 2 Einheiten	$-4(1/5) 2^2$

Unterteilt man das Intervall $[a; a+1]$ in N gleichlange Teile (in unserem Beispiel $N=10$), so kann man das Spiel unverändert nach obiger Regel fortsetzen, nur daß die Sprungweite k keine natürliche Zahl mehr ist, sondern ein Vielfaches von $1/N$.

Wir beenden das oben abgebrochene Spiel wie folgt:

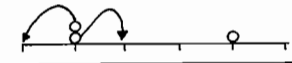

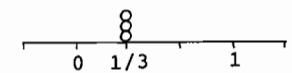
	4 Doppelsprünge um $5/10$ Einheiten	$-4(1/5) (5/10)^2$
	1 Doppelsprung um $1/10$ Einheiten	$+ (1/5) (1/10)^2$

	je 1 Doppelsprung um 2/10 Einheiten 3/10 Einheiten	$-(1/5)(2/10)^2$ $-(1/5)(3/10)^2$
	2 Doppelsprünge um 1/10 Einheiten	$-2(1/5)(1/10)^2$
	2 Doppelsprünge um 1/10 Einheiten	$-2(1/5)(1/10)^2$
	4 Doppelsprünge um 1/10 Einheiten	$-4(1/5)(1/10)^2$
		Summe: $-1,84$

Der letzte Turm der Münzverteilung steht auf der Stelle 2,6, und unser Abakus liefert eine Streuungsmaßzahl $S=1,84$. Wie man nachrechnet, hat die Verteilung, mit der wir gestartet sind, tatsächlich einen Erwartungswert von 2,6 und die Varianz 1,84.

3.3.4 Beispiel 4

Auch für die Binomialverteilung $B_{1;1/3}$ stimmen die Ergebnisse des Abakus mit dem Erwartungswert $\mu=1/3$ und Varianz $V=2/9$ überein:

	1 Doppelsprung um 1/3 Einheiten	$+(2/3)(1/3)^2$
	1 Doppelsprung um 2/3 Einheiten	$-(2/3)(2/3)^2$
		Summe: $-6/27$ $= -2/9$

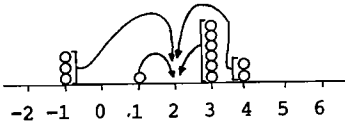
3.4 Herleitung der Formel für die Varianz aus den Spielregeln des Abakus

Die Beispiele zeigen, daß für eine beliebige Anfangsverteilung der letzte Turm des Abakus stets auf der Stelle μ stehen wird, die sich durch die Formel

$$\mu = x_1 p(x_1) + x_2 p(x_2) + \dots + x_n p(x_n)$$

des Erwartungswertes vorhersagen läßt.

Ist es möglich, eine Vorhersage auch für die Streuungsmaßzahl S zu machen, die der Abakus liefern wird? Ja! Wenn der letzte Turm sowieso auf der Stelle μ stehen wird, könnten wir versuchsweise auf symmetrische Sprünge verzichten und jede Münze für sich sofort auf die Stelle μ hüpfen lassen. Für das Beispiel 3.3.2 mit $\mu=2$



wären 3 Einzelsprünge um $k=3$ Einheiten, 2 Einzelsprünge um $k=2$ Einheiten und 7 Einzelsprünge um $k=1$ Einheiten nötig. Da ein Doppelsprung um k Einheiten den Streuungswert S um $(2/N)k^2$ reduziert, vermuten wir, daß S durch einen Einzelsprung zur Stelle μ um die Hälfte $(1/N)k^2$ verkleinert wird. In der Tat liefert diese Vermutung mit

$$(3/12)3^2 + (1/12)1^2 + (6/12)2^2 = 42/12 = 3,5$$

das gleiche Resultat wie der Abakus. Andere Beispiele bestätigen die Vermutung, und es ist kaum zu bezweifeln, daß S bei bekanntem Erwartungswert stets durch die Formel

$$S = (x_1 - \mu)^2 p(x_1) + (x_2 - \mu)^2 p(x_2) + \dots + (x_n - \mu)^2 p(x_n)$$

vorhersagbar ist, also mit der Varianz übereinstimmt.

3.5 Beweise zum Varianzabakus

Beim Spiel des Abakus faszinieren immer wieder die Tatsachen,

- (a) daß er stets aufgeht, egal wie "krumm" die Anfangsverteilung auch aussehen mag,
 (b) daß die Ergebnisse, also die Stelle des letzten Turmes, insbesondere aber die Streuungsmaßzahl S vom Spielverlauf unabhängig - ja durch die Formeln $\mu = \sum_{i=1}^n x_i p(x_i)$ und $S=V = \sum_{i=1}^n (x_i - \mu)^2 p(x_i)$ vorhersagbar - sind.

Wir wollen sie mit (b) beginnend beweisen:

Behauptung 1:

Wenn der Abakus aufgeht, d.h. wenn sich die Ausgangsverteilung durch symmetrische Doppelsprünge in einen Turm überführen läßt, dann steht dieser auf der Stelle des Erwartungswertes μ und das Spielergebnis S stimmt mit der Varianz V der Ausgangsverteilung überein.

Beweis: In jedem Schritt unseres Spiels werden zwei Münzen (vom Gewicht $1/N$) symmetrisch um je k Einheiten aufeinander zubewegt. Haben sie sich vorher auf den Stellen $i+k$ und $i-k$ befunden, liegen sie nachher auf der Stelle i .

Ihr Beitrag zum Erwartungswert ist

vorher: $(1/N)(i-k) + (1/N)(i+k)$

nachher: $(1/N) i + (1/N) i$.

Der Erwartungswert der Verteilung ändert sich also durch diese Bewegung nicht. Die Stelle, auf der der Turm schließlich steht, ist daher der Erwartungswert.

Der Beitrag der betrachteten Münzen zur Varianz V ist

$$\begin{aligned} \text{vorher: } & (1/N)((i-k)-\mu)^2 + (1/N)((i+k)-\mu)^2 = \\ & (1/N)((i-\mu)-k)^2 + (1/N)((i-\mu)+k)^2 = \\ & (2/N)(i-\mu)^2 + (2/N)k^2 + \underbrace{(2/N)(i-\mu)k - (2/N)(i-\mu)k}_{=0 (*)} = \\ & (2/N)(i-\mu)^2 + (2/N)k^2 \end{aligned}$$

nachher: $(2/N)(i-\mu)^2$.

Wie man sieht, ist die Varianz V der Verteilung mit $(2/N)k^2$ um den gleichen Betrag abgebaut worden wie die Streuungsmaßzahl S . Die Gleichheit von S und V folgt nun aus der Tat-

sache, daß bei Spielende (Konzentration der Münzen auf einem Turm) sowohl S als auch V verschwinden.

Der "Trick" des Abakus besteht darin, daß sich bei der vorgeschriebenen Bewegung die Veränderungen (*) der Varianz aufheben, die von der Lage des (noch unbekanntes) Erwartungswertes abhängen.

Zu (a): Wir beweisen nun, daß der Abakus stets aufgeht, d.h. bis zur Konzentration aller Münzen in einem Punkt gespielt werden kann. Gegeben sei eine Anfangsverteilung A von N Münzen ($N \geq 3$)¹⁾ auf ganzzahligen Stellen des Zahlenstrahls. Springt man zunächst nur um ganze Einheiten, so hat man

Behauptung 2:

Man kann die Münzen durch symmetrische Doppelsprünge soweit konzentrieren, daß entweder ein Turm entsteht (falls $\mu \in \mathbb{Z}$) oder zwei benachbarte Türme (falls $\mu \notin \mathbb{Z}$).

Beweis: Wir betrachten eine Verteilung M mit minimaler Varianz V, welche aus der anfänglichen Verteilung A durch symmetrische Doppelsprünge um ganze Einheiten zu konstruieren ist. Es gilt

(i) M besteht aus höchstens zwei Türmen.

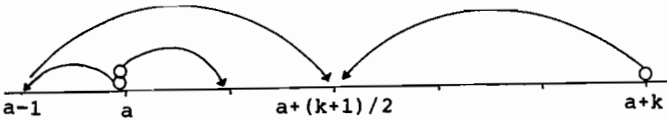
Beweis (indirekt): Angenommen, M hätte mehr als zwei Türme, dann wähle man drei Türme aus. Mindestens zwei dieser drei Türme haben einen geradzahligen Abstand auf der Zahlengeraden. Mit den Münzen dieser Türme ist daher ein ganzzahliger symmetrischer Doppelsprung möglich, welcher die Varianz der Verteilung M reduziert. Das steht aber im Widerspruch zur Minimalität von M.

(ii) Falls M aus zwei Türmen besteht, liegen diese benachbart (auf den Stellen a, $a+1 \in \mathbb{Z}$).

Beweis (indirekt): Wir nehmen an, die Türme befänden sich auf den Punkten a und $a+k$ des Zahlenstrahls mit $k > 1$. Nach

1) Die Fälle $N=1$ und $N=2$ sind trivial.

der Minimalitätseigenschaft der Verteilung M muß k ungerade sein, insbesondere muß $k > 2$ gelten. Bei geradem k wäre nämlich ein symmetrischer Sprung von zwei Münzen auf die Stelle $a+k/2$ möglich, der die Varianz sofort reduzieren würde. Es sei also $k > 2$ ungerade. Da M nach Voraussetzung aus mindestens drei Münzen besteht, enthält einer der Türme, etwa derjenige auf der Stelle a , mindestens zwei Münzen. Die folgenden symmetrischen Münzbewegungen



reduzieren die Varianz um

$$\Delta V = -\frac{2}{N} 1^2 + \frac{2}{N} \left(\frac{k+1}{2}\right)^2 = \frac{(k+1)^2 - 4}{2N} > \frac{16-4}{2N} = \frac{6}{N},$$

wobei die Ungleichung aus $k > 3$ folgt. Damit haben wir wieder einen Widerspruch zur Minimalität von M erhalten.

Wir nehmen an, daß sich beim Spiel über ganzzahligen Stellen der Zahlengeraden zwei Türme in den Punkten a und $a+1$ ergeben haben ($a, a+1 \in \mathbb{Z}$).

Behauptung 3:

Nach Unterteilung des Intervalls $[a; a+1]$ in N Teile kann man das Spiel bis zur Konzentration aller Münzen auf einem Turm im Erwartungswert fortsetzen.

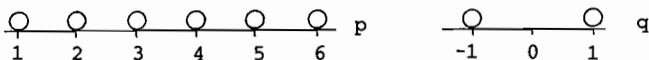
Beweis: Nach Behauptung 2 (nun für Schrittweite $1/N$ statt 1) kann man das Spiel soweit fortsetzen, bis entweder ein Turm entsteht oder zwei benachbarte Türme, jetzt in Punkten von der Form p/N und $(p+1)/N$. Die zweite Möglichkeit entfällt jedoch hier, da der Erwartungswert einer Verteilung von N Münzen auf ganzzahligen Stellen des Zahlenstrahls von der Form k/N sein muß, also nicht zwischen p/N und $(p+1)/N$ liegen kann. Der Beweis für das Funktionieren des Varianzabakus ist damit erbracht.

3.6 Beweis der Additivität von Erwartungswert und Varianz über den Faltungsabakus und den Varianzabakus

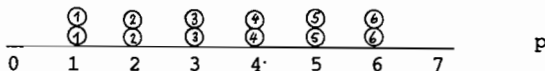
Die Additivität der Varianz unter Faltungen $V(p*q)=V(p)+V(q)$ wurde bereits bei Einführung des Varianzabakus in 3.3 demonstriert. Sie soll nun anhand zweier einfacher Beispiele bewiesen werden, die wiederum auf die Sekundarstufe I zugeschnitten sind, aber alle wesentlichen Gedanken offenlegen.

Erstes Beispiel mit $\mu(q)=0$:

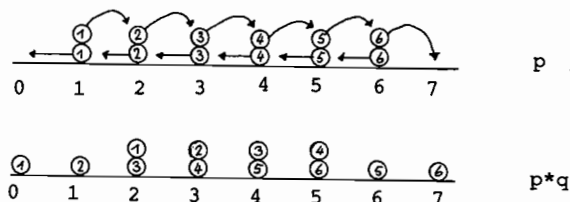
Ein Mensch Ärgere Dich Nicht Spieler versucht, seine gewürfelte Punktzahl durch anschließenden Münzwurf aufzubessern. Fällt Kopf ($\hat{=}$ 1), dann geht er zusätzlich zur gewürfelten Augenzahl einen Schritt weiter, fällt Zahl ($\hat{=}$ -1), so muß er einen Schritt zurück. Wir stellen die erste Wahrscheinlichkeitsverteilung p (Würfel) durch $N=6$ und die zweite Wahrscheinlichkeitsverteilung q (Münze) durch $M=2$ Münzen dar:



Die Wahrscheinlichkeitsverteilung der gesamten Punktzahl (Summenexperiment) entsteht durch Faltung von p und q , indem wir jeden Münzturm von p der Verteilung q entsprechend aufteilen. Zur experimentellen Durchführung dieser Aufteilung stellen wir p durch $N \cdot M=12$ Münzen dar und erhalten:

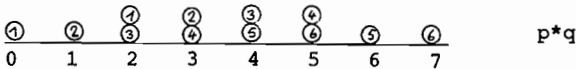


Falten von p mit q :

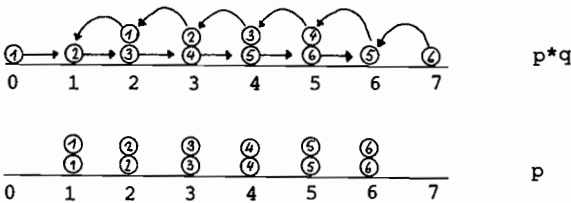


Denkt man sich die Faltungsoperation, die von der Verteilung p zur Verteilung $p*q$ führte, in umgekehrter Richtung durchgeführt, dann sieht man ein, daß gilt $V(p*q) = V(p) + V(q)$.

Falls $\mu(q) = 0$, wie in unserem Beispiel 1, kann man nämlich diese Rückführung sofort auch als N -fache Anwendung des Varianzabakus zur Verteilung q deuten. Da er jeweils mit Münzen vom Gewicht $1/(N \cdot M)$ (und nicht $1/M$) ausgeführt wird, verringert sich dadurch die Varianz um $(1/N)V(q) \cdot N = V(q)$.



$N (=6)$ -fache Anwendung des Varianzabakus für q :



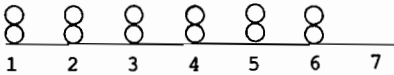
Nun liefert die Anwendung des Abakus auf p die gesuchte Gleichung $V(p*q) = V(p) + V(q)$.

Falls $\mu(q) \neq 0$, kann man die Rückführung auch als N -fache Anwendung des Varianzabakus zur Verteilung q und anschließende Verschiebung der Verteilung um $-\mu(q)$ deuten. Zusätzlich zur Formel $V(p*q) = V(p) + V(q)$ ist damit die Formel $\mu(p*q) = \mu(p) + \mu(q)$ bewiesen.

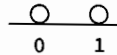
Zweites Beispiel mit $\mu(q) \neq 0$:

Wie oben, nur daß der Spieler beim Wurf einer Zahl nicht um eine Einheit zurückzugehen braucht ($K \hat{=} 1, Z \hat{=} 0$).

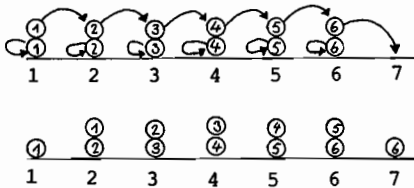
Verteilung p dargestellt
durch $N \cdot M = 12$ Münzen



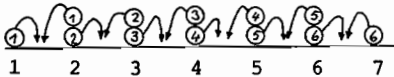
Verteilung q dargestellt
durch $M = 2$ Münzen



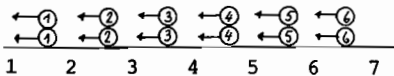
Falten von p mit q



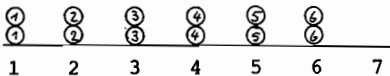
Umkehrung der Faltung durch $N=6$ -fache Anwendung des Varianz-
abakus zu q



und anschließende Verschiebung um $\mu(q) = -1/2$



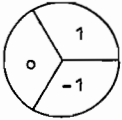
liefert die Verteilung p



Die Varianz des Mittelwertes

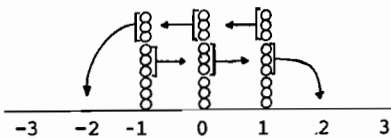
Warum macht man zur Bestimmung von Größen oft mehrere Messungen und bestimmt deren Mittelwert? Das Wissen um die Additivität der Varianz erlaubt eine Antwort:

Simulieren wir Meßfehler (Wahrscheinlichkeitsverteilung p) durch ein Glücksrad

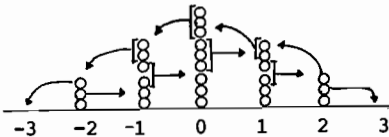


0 bedeutet kein Meßfehler
 1 bedeutet eine Einheit zu viel gemessen
 -1 bedeutet eine Einheit zu wenig gemessen,

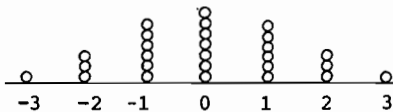
so erhalten wir die Wahrscheinlichkeiten der Fehler zur Summe aus zwei und drei unabhängig ausgeführten Versuchen wieder durch Faltung von p mit sich selbst:



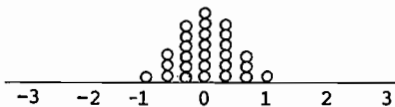
(1) Verteilung p der Fehler zu einer Messung (27 Münzen)



(2) Verteilung $p*p$ der Fehler zur Summe zweier Messungen, $V(p*p) = 2V(p)$



(3) Verteilung $p*p*p$ der Fehler zur Summe dreier Messungen, $V(p*p*p) = 3V(p)$



(4) Verteilung $(1/3) \circ p*p*p$ des Mittelwertes aus drei Meßfehlern, $V((1/3) \circ p*p*p) = (1/3)V(p)$

Auch ist leicht einzusehen, daß die Verteilung des Mittelwertes aus drei Messungen (Schreibweise $(1/3) \circ p * p * p$) durch Kontraktion der Türme von $p * p * p$ (3) mit dem Faktor $1/3$ in Richtung der Längsachse hervorgeht (4). Der Varianzabakus für (4) ist mit demjenigen für (3) identisch - bis auf die Tatsache, daß die Sprungweite bei (4) um den Faktor $1/3$ kleiner ist als bei (3). Das Gewicht der Sprünge zu (4) beträgt also jeweils nur $1/9$ des Gewichtes der entsprechenden Sprünge zu (3). Damit ist

$$V((1/3) \circ p * p * p) = (1/9)V(p * p * p) = (1/3)V(p) \quad , \text{ analog}$$

$$V((1/k) \circ \overset{k}{\underset{1 \dots 1}{*}} p) = (1/k)V(p) .$$

Die Varianz der Verteilung des Mittelwertes aus k Messungen nimmt also mit dem Faktor $1/k$ ab. Und das ist die Antwort auf die oben gestellte Frage.

Nebenbei fördert die enaktive Darstellungsweise das Entstehen eines intuitiven Gefühls für den zentralen Grenzwertsatz: Egal wie die Anfangsverteilung aussieht, die Verteilungen zu den Summenexperimenten erhalten Glockenform.

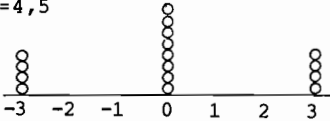
3.7 Die Ungleichung von Tschebyscheff

Kehren wir zum Ausgangspunkt unserer Überlegungen, dem Glücksspieler X aus 3.2 zurück. Seine Glücksmaschine hatte den Streuungswert $V(p * q * r) = 5,5$ und sie lieferte mit Wahrscheinlichkeit $6/16$ die Punktzahlen $|x| \geq 3$. Kann er durch Konstruktion eines neuen Glücksspiels mit gleichem Erwartungswert 0 seine Gewinnchancen verbessern, ohne die Varianz $5,5$ zu überschreiten?

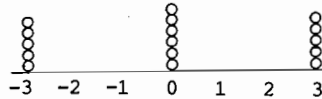
Wenn man mit den 16 Münzen an der Wahrscheinlichkeitsverteilung des gewünschten Glücksspiels herumexperimentiert, dann erkennt man schnell, daß die Gewinnchancen optimal werden, wenn man so viele Münzen wie die Varianz gerade zuläßt, auf den Stellen 3 und -3 placiert und die restlichen auf der Stelle des Erwartungswertes $-$ wo sie keinen zusätzlichen Beitrag zur Varianz liefern.

8 Münzen, zu gleichen Teilen auf die Stellen 3 und -3 verteilt, ergeben einen Beitrag von $V=(8/16) \cdot 3^2=4,5 < 5,5$. 10 Münzen, zu gleichen Teilen auf die Stellen 3 und -3 verteilt, liefern aber schon einen Beitrag von $V=(10/16) \cdot 3^2=5,625 > 5,5$, der zu groß ist.

$V=4,5$

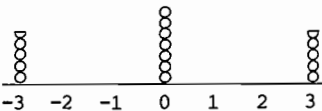


$V=5,625$

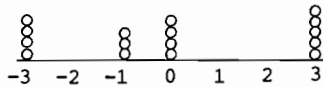


Vielleicht sollte man es mit 9 Münzen versuchen? Wo bringt man die neunte unter? Entweder man teilt sie in zwei Hälften, die ebenfalls in 3 und -3 deponiert werden (tatsächlich hat man dann $V=(9/10) \cdot 3^2=5,0625 < 5,5$) oder man legt sie ganz auf die Stelle 3 und sorgt durch 3 Münzen in -1 für einen Erwartungswert 0. Auch dann gilt noch $V=(9/10) \cdot 3^2 + (3/10) \cdot 1^2=5,25 < 5,5$,

$V=5,0625$



$V=5,25$



aber die erste Position ist günstiger, denn sie liefert bei gleichen Gewinnchancen geringere Varianz.

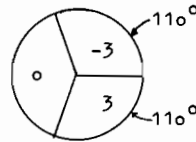
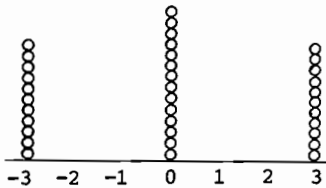
Deuten wir $8/16$, $10/16$ und $9/16$ wieder als Wahrscheinlichkeiten $P=P(|x-\mu| \geq 3)$ für eine Abweichung um 3 oder mehr Einheiten, so erkennt man, daß gelten muß $P \cdot 3^2 < 5,5$ oder $P < 5,5/9=22/36$ bzw. allgemein:

$$P \cdot d^2 \leq V \quad \text{bzw.} \quad P \leq V/d^2 \quad (\text{Ungleichung von Tschebyscheff}).$$

Dabei bezeichnet d die gewünschte Abweichung vom Erwartungswert, V die gegebene Varianz, die nicht überschritten werden darf und $P=P(|x-\mu| \geq d)$ die Wahrscheinlichkeit der Abweichung um mehr als d Einheiten.

Unser Glücksspieler X wird also bei der gegebenen Gewinnregel seine Zufallsmaschine so einrichten, daß sie folgende Verteilung erzeugt:

$V=5,5$



Sie wird realisiert durch das rechtsstehende Glücksrad und liefert mit Wahrscheinlichkeit $22/36=0,6\bar{1}$ fast doppelt so viele Abweichungen $|x| \geq 3$ wie das alte Zufallsexperiment $p * q * r$.

Natürlich kann man die Ungleichung von Tschebyscheff auch in den Zusammenhang mit 3.6 stellen: Da die Varianz V der Verteilung des Mittelwertes aus k Messungen mit dem Faktor $1/k$ abnimmt, nimmt auch die obere Schranke V/d^2 der Wahrscheinlichkeit für Abweichungen um mehr als d vom Erwartungswert mit dem Faktor $1/k$ ab. Und das ist das schwache Gesetz der großen Zahlen, das wir hiermit auf weitgehend enaktivem Niveau hergeleitet haben.

4 Der elastische Stab

Dieses Kapitel beschreibt einen im Vergleich zu 3. nicht weniger interessanten heuristischen Zugang zum Themenkreis Varianz und Tschebyscheff-Ungleichung. Diesmal wird ein physikalisches Modell verwendet.

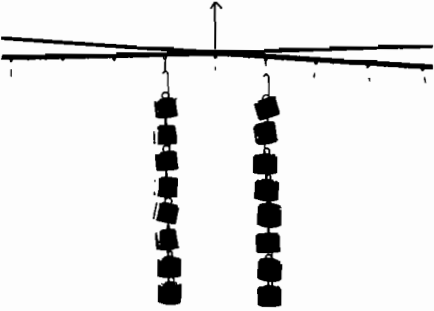
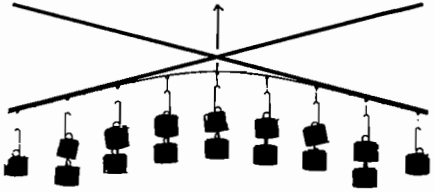
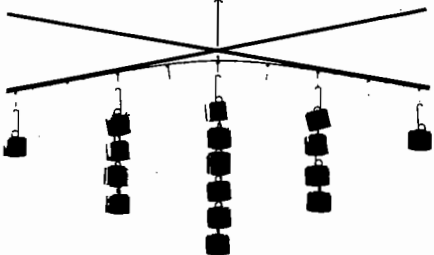
4.1 Das Modell

Wie bekannt, kann der Schwerpunkt eines mit Gewichten belasteten Stabes zur Veranschaulichung des Erwartungswertes herangezogen werden. Wenn man statt eines starren Stabes einen elastischen verwendet, dann nimmt dessen Durchbiegungswinkel mit weit vom Schwerpunkt (Erwartungswert) entfernten Gewichten (Wahrscheinlichkeiten) zu, so daß er intuitiv naheliegend als Maß für die Streuung einer Verteilung dienen kann. Überraschenderweise erweist sich dieses Maß als proportional zur Varianz.

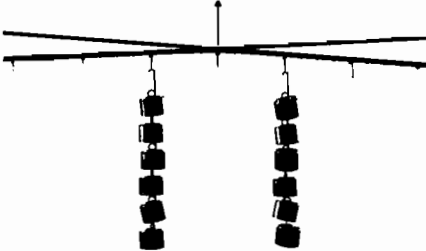
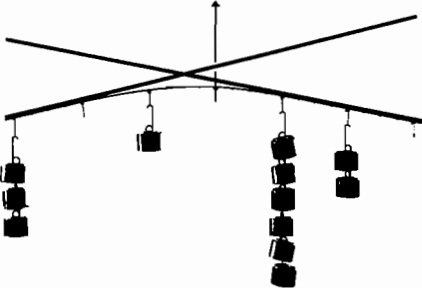
Den Durchbiegungswinkel des elastischen Stabes kann man über Lichtstrahlen und auf dem Stab befestigte Spiegel sehr genau bestimmen. Folgende Methode ist jedoch einfacher und eindrucksvoller: An den Enden des Stabes werden tangential dünne Aluminiumrohre befestigt; sie spreizen sich bei Belastung auseinander und gestatten an ihrem Schnittpunkt das Ablesen des Durchbiegungswinkels (am besten in Schattenprojektion).

Die Fotoserie 1 auf der folgenden Seite zeigt die symmetrischen Verteilungen q , $p \cdot q \cdot r$ und $q \cdot q \cdot q \cdot q$ der Spieler X und Y aus 3.2. Sie werden am elastischen Stab durch 16 Gewichtsstücke vom Gesamtgewicht 16 N dargestellt. (Die Einheit auf dem Stab beträgt 6 cm.) Die Fotoserie 2 auf der übernächsten Seite zeigt die unsymmetrische Verteilung aus 3.3.2, dargestellt durch 12 Gewichtsstücke vom Gesamtgewicht 12 N. (Die Einheit auf dem Stab beträgt hier 8 cm.)

Fotoserie 1 zum elastischen Stab

<p>Einheitsverteilung q Durchbiegungswinkel $\alpha (q) = 5,5^\circ$</p>	
<p>Verteilung $p*q*r$ Durchbiegungswinkel $\alpha (p*q*r) = 29^\circ$</p>	
<p>Verteilung $q*q*q*q$ Durchbiegungswinkel $\alpha (q*q*q*q) = 21^\circ$</p>	

Fotoserie 2 zum elastischen Stab

<p>Einheitsverteilung q Durchbiegungswinkel $\alpha(q) = 7,0^\circ$</p>	
<p>Verteilung t Durchbiegungswinkel $\alpha(t) = 25,0^\circ$</p>	

Wie den Varianzabakus (3.3) so kann man auch den elastischen Stab sofort zum quantitativen Vergleich zweier Streuungen benutzen. Man stellt die Verteilungen, die verglichen werden sollen, durch eine gleiche Anzahl von Gewichtsstücken dar und "hängt sie an den Stab", welcher im Schwerpunkt unterstützt wird. Die Verteilung, welche den größeren Durchbiegungswinkel liefert, hat die größere Streuung. Wenn wir eine Verteilung mit Einheitsstreuung festlegen (die Verteilung q aus 3.2 mit $q(-1)=q(1)=1/2$), dann können wir auf diese Weise jeder beliebigen Verteilung eine Streuungsmaßzahl S zuordnen.

So liefert etwa Fotoserie 1 mit $\alpha(p*q*r)=29^{\circ}$, $\alpha(q*q*q*q)=21^{\circ}$ und $\alpha(q)=5,5^{\circ}$ die Streuungsmaßzahlen

$$S(p*q*r) = \alpha(p*q*r)/\alpha(q) = 5,3,$$

$$S(q*q*q*q) = \alpha(q*q*q*q)/\alpha(q) = 3,8.$$

Fotoserie 2 liefert mit $\alpha(t)=25,0^{\circ}$, $\alpha(q)=7^{\circ}$ die Streuungsmaßzahl

$$S(t) = \alpha(t)/\alpha(q) = 3,6.$$

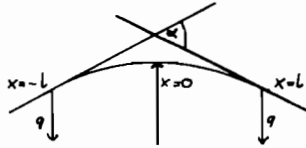
Diese Ergebnisse stimmen mit den Varianzen $V(p*q*r)=5,5$, $V(q*q*q*q)=4$, $V(t)=3,5$ recht gut überein.

4.2 Herleitung der Formel für die Varianz

Wir wollen versuchen, für eine beliebige Verteilung das experimentelle Ergebnis vorherzusagen. Zu diesem Zwecke studieren wir, wie sich der Durchbiegungswinkel unseres Stabes

- a) bei konstantem Hebelarm l und variabler Belastung q
 - b) bei konstanter Belastung q und variablem Hebelarm l
- ändert:

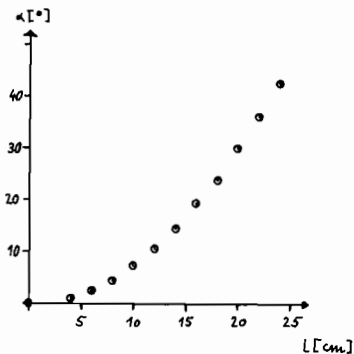
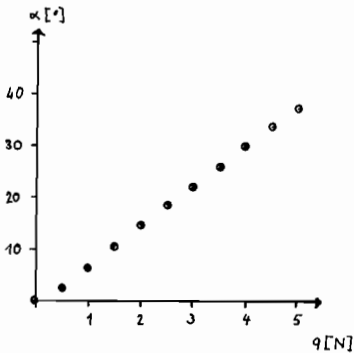
Meßreihen zur Abhängigkeit des Durchbiegungswinkels α von der Belastung q und dem Hebelarm l für einen symmetrisch belasteten Stab:



1. Meßreihe: $l = 20 \text{ [cm]}, q$ variabel			
q [N]	α [°]	$h_1 = \alpha/q$ [°/N]	$h_2 = \alpha/2ql^2$ [°/N cm ²]
0.5	2.6	5.2	0.0065
1.0	6.6	6.6	0.0083
1.5	10.6	7.1	0.0088
2.0	14.5	7.3	0.0091
2.5	18.5	7.4	0.0093
3.0	22.0	7.3	0.0092
3.5	26.0	7.4	0.0093
4.0	30.0	7.5	0.0094
4.5	33.5	7.4	0.0093
5.0	37.0	7.4	0.0093

2. Meßreihe: $q = 4 \text{ [N]}, l$ variabel			
l [cm]	α [°]	$h_1 = \alpha/l^2$ [°/cm ²]	$h_2 = \alpha/2ql^2$ [°/N cm ²]
4.0	1.0	0.063	0.0078
6.0	2.5	0.069	0.0087
8.0	4.5	0.070	0.0088
10.0	7.5	0.075	0.0094
12.0	10.8	0.075	0.0094
14.0	14.5	0.074	0.0092
16.0	19.5	0.076	0.0095
18.0	24.0	0.074	0.0093
20.0	30.0	0.075	0.0094
22.0	36.0	0.074	0.0093
24.0	42.5	0.074	0.0092

Mittlerer Wert: $k = 0.0093 \text{ [°/N cm}^2\text{]}$



Den Tabellen entnimmt man, daß sich bei Verdoppelung (Verdreifachung) von q auch α verdoppelt (verdreifacht). Wenn man dagegen l verdoppelt (verdreifacht), dann vervierfacht (verneunfacht) sich α . α hängt also linear von q und quadratisch von l ab, was in Übereinstimmung mit unserem Experiment beschrieben wird durch die Funktion

$$\begin{aligned} q &\longrightarrow \alpha = k_1 \cdot q & , & \quad k_1 \approx 7,3 \\ l &\longrightarrow \alpha = k_2 \cdot l^2 & , & \quad k_2 \approx 0,074. \end{aligned}$$

Faßt man die Abhängigkeiten in einer Gleichung $\alpha = k_3 \cdot q \cdot l^2$ zusammen, dann liefern unsere Meßreihen für $k_3 = \alpha / (q \cdot l^2)$ einen mittleren Wert von 0,019. Da der Stab mit zwei Gewichten ¹⁾ belastet wurde, trägt ein einzelnes im Abstand l vom Schwerpunkt mit der Hälfte, nämlich $\alpha = k \cdot q \cdot l^2$, $k = \alpha / (2q l^2) = k_3 / 2 \approx 0,0093$ zur Durchbiegung bei. Deswegen sollte sich bei Belastung mit mehreren Gewichten, die (nicht notwendig symmetrisch) in Abständen $l_i = x_i - \mu$ vom Schwerpunkt μ aufgehängt sind, insgesamt ein Durchbiegungswinkel von

$$\alpha = k \sum_i q_i \cdot l_i^2 = k \sum_i q_i (x_i - \mu)^2, \quad k \approx 0,0093$$

ergeben. Tatsächlich liefert diese Formel für die unsymmetrische Verteilung t der Fotoserie 2 mit

3 N bei $x = -3$ E = -24 cm,

1 N bei $x = -1$ E = - 8 cm,

6 N bei $x = 1$ E = 8 cm,

2 N bei $x = 2$ E = 16 cm

einen Durchbiegungswinkel

$\alpha = 0,0093 [3 \cdot (-24)^2 + 1 \cdot (-8)^2 + 6 \cdot 8^2 + 2 \cdot 16^2] = 24,9^\circ$, der mit dem gemessenen Wert von $\alpha = 25^\circ$ gut übereinstimmt. Die Streumaßzahl einer beliebigen Verteilung p erhalten wir, indem wir den Durchbiegungswinkel $\alpha(p)$ durch den Durchbiegungswinkel $\alpha(q)$ der zugehörigen Einheitsverteilung teilen. Ist p

¹⁾ Man kann den Stab natürlich auch im Auflagepunkt horizontal einspannen und das Experiment mit einseitiger Belastung durchführen. Zweiseitige Belastung liefert wegen des doppelten Durchbiegungswinkels aber höhere Meßgenauigkeit.

durch Gewichtsstücke p_i vom Gesamtgewicht P dargestellt, dann wird q durch die Gewichte $P/2$ an den Stellen -1 und 1 repräsentiert. Unsere Formel liefert $\alpha(q) = k \cdot P$, und die Streuungsmaßzahl

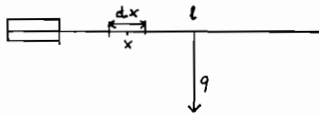
$$s(p) = \frac{\alpha(p)}{\alpha(q)} = \frac{k \sum p_i (x_i - \mu)^2}{k P} = \sum \frac{p_i}{P} (x_i - \mu)^2$$

entpuppt sich wiederum als die Varianz.

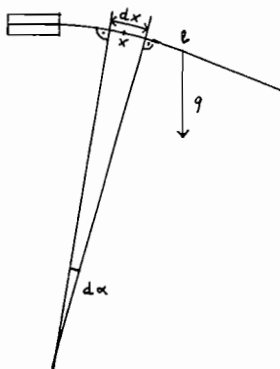
4.3 Physikalischer Hintergrund

In der Sekundarstufe I wird der Zugang zum Gesetz $\alpha = k \sum q_i (x_i - \mu)^2$ rein experimentell sein, in der Sekundarstufe II ist eine Herleitung aus dem Hookeschen Gesetz möglich. Ein elastisches Material ist dadurch definiert, daß die Verformung der Belastung proportional ist (Hookesches Gesetz). Betrachten wir einen einseitig horizontal eingespannten Stab, der an der Stelle l durch ein Gewicht q belastet ist. Er wird sich solange verformen, bis wieder Gleichgewicht besteht:

Stab vor
Verformung



Stab nach
Verformung



Auf einen Punkt des Stabes, der bei x liegt, übt die Kraft q ein Drehmoment $q(1-x)$ aus. Nach dem Hookeschen Gesetz erfährt der Stab auf dem Abschnitt dx daher eine Verbiegung um den Winkel ¹⁾

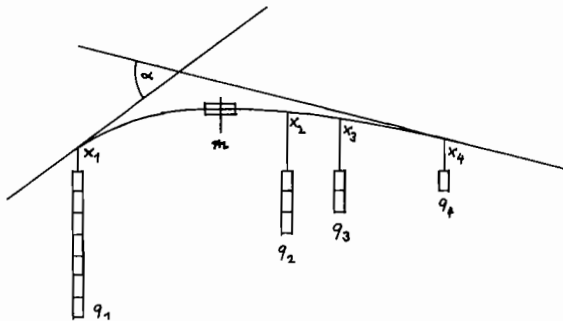
$$d\alpha = \lambda q(1-x) dx.$$

Der gesamte Durchbiegungswinkel ist dann

$$\alpha = \int_0^l d\alpha = \lambda \int_0^l q(1-x) dx = \frac{\lambda}{2} q \cdot l^2.$$

Er ist der Belastung q direkt proportional und nimmt quadratisch mit der Länge l des "Hebelarmes" zu. Belastet man den Stab mit Gewichten q_1, q_2, \dots, q_n an den Stellen x_1, x_2, \dots, x_n und spannt man ihn an der Stelle m horizontal ein, so erhält man durch Überlagerung den Durchbiegungswinkel

$$\alpha = k \sum_{i=1}^n (x_i - m)^2 q_i, \text{ wobei } k = \lambda/2.$$



1) Die Proportionalitätskonstante λ ist gegeben durch $\lambda = \frac{1}{EJ}$, wobei E den Elastizitätsmodul des verwendeten Stabmaterials, J das Trägheitsmoment des Stabquerschnittes bezeichnet. So gilt für Stahl $E = 2,1 \cdot 10^6$ kp/cm², für einen Flachträger mit rechteckigem Querschnitt der Breite b und Höhe h : $J = \frac{b h^3}{12}$ [Hütte, S. 873, 688, 1003].

Wählt man als Einspannpunkt den Schwerpunkt $m=\mu$, dann kann man statt der Einspannung auch eine drehbare Lagerung vornehmen, der Stab hat dann im Lager automatisch horizontale Neigung.

Analyse der systematischen Fehler

Aus den Meßreihen entnimmt man, daß sich für sehr kleine und sehr große Belastungen experimentell etwas zu kleine Winkel (zu kleine Werte für $k=\alpha/2ql^2$) ergeben.

(a) Bei starker Belastung ist nämlich die durch die Neigung des Stabes (gegen die Horizontale) verursachte Änderung der Drehmomente zu berücksichtigen. Der Hebelarm $l-x$, mit dem ein bei l sitzendes Gewicht an der Stelle x angreift, ist zu ersetzen durch $\int_x^l \cos\alpha(t)dt$. Die sofort integrierbare Differentialgleichung

$$\frac{d\alpha(x)}{dx} = \mathcal{K} q (l-x)$$

geht dadurch über in die Integro-Differentialgleichung

$$\frac{d\alpha(x)}{dx} = \mathcal{K} q \int_x^l \cos\alpha(t)dt,$$

aus dem man durch Differenzieren die Differentialgleichung zweiter Ordnung

$$\frac{d^2\alpha(x)}{dx^2} = -\mathcal{K} q \cos\alpha(x)$$

erhält, welche nur noch numerisch zu lösen ist.

(b) Bei schwacher Belastung, also kleinen Drehmomenten ist das Gegendrehmoment, das die tangential befestigten Aluminiumstäbe erzeugen, nicht mehr zu vernachlässigen. Messungen des Durchbiegungswinkels über die Ablenkung von Lichtstrahlen durch Spiegel, welche auf dem Stab befestigt werden, liefern bessere Ergebnisse.

In der Praxis sind die systematischen Fehler (a) und (b) unbedeutend, da sie sich erst bei sehr kleinen und sehr großen Winkeln ($\alpha < 4^\circ$, $\alpha > 40^\circ$) auswirken. Bei der experimentellen Bestimmung von Varianzen kann man sie durch geeignete Wahl der Einheiten auf dem Stab leicht umgehen. Trotzdem mag man sich fragen, welchen didaktischen Vorzug der elastische Stab gegenüber der bekannten und durch keinen systematischen Fehler behafteten Veranschaulichung der Varianz als Trägheitsmoment eines belasteten starren Stabes hat. Die Antwort fällt leicht: Der Durchbiegungswinkel ist sofort abzulesen, das Trägheitsmoment läßt sich nur über eine gleichmäßig beschleunigte Drehbewegung als Proportionalitätskonstante zwischen Winkelbeschleunigung und Drehmoment ermitteln. Es ist dem Erlebnishorizont kleiner Schüler nicht zugänglich.

Die Gleichung $\sum_i (x_i - m)^2 q_i = \sum_i (x_i - \mu)^2 q_i + (m - \mu)^2 Q$, aus der die Minimalitätseigenschaft des Erwartungswertes (3.1.1) folgt, ist in beiden Modellen überprüfbar. Für den elastischen Stab besagt sie, daß der Durchbiegungswinkel minimal wird, wenn man den Stab im Schwerpunkt horizontal einspannt, für das Trägheitsmoment ist sie als der Steinersche Satz bekannt.

4.4 Die Ungleichung von Tschebyscheff

Wir führen einen im Schwerpunkt $\mu=0$ unterstützten elastischen Stab vor, dessen Massenbelegung (etwa 10 gleichartige Massenstücke vom Gesamtgewicht Q) verdeckt ist. Der Durchbiegungswinkel ist an den gespreizten Stäben der Messung zugänglich. Wie viele Massenstücke mit Gesamtgewicht q können sich im Bereich B außerhalb des Intervalles $]-d;d[$ aufhalten? Durch Experimente unterstützt finden Schüler schnell heraus, daß bei einem festen, nicht zu überschreitenden Durchbiegungswinkel am meisten Massenstücke dann in dem fraglichen Bereich unterzubringen sind, wenn man gleich

viele an den Stellen $-d$ und d aufhängt, die restlichen im Schwerpunkt, so daß sie zur Durchbiegung keinen Beitrag mehr leisten. Wegen $k \cdot d^2 \cdot q \ll \mu$ muß dann gelten

$$q \ll \frac{\alpha}{k d^2}$$

und das ist eine "physikalische Form" der Tschebyscheff-Ungleichung. Erinnern wir uns, daß der Streuungswert S einer Verteilung mit Durchbiegungswinkel α und Gesamtgewicht Q gegeben war durch $S = \alpha/kQ$, so erhalten wir

$$\frac{q}{Q} \ll \frac{S}{d^2}$$

also die "mathematische Form" der Tschebyscheff-Ungleichung.

4.5 Curriculare Aspekte

Vergleicht man die beiden Methoden zur Einführung eines Streuungsmaßes (Varianzabakus und elastischer Stab), dann wird man auf der Grundlage der Ausführungen von Kapitel 2 geneigt sein, dem Münzmodell den Vorzug zu geben, weil es sich um das Musterbeispiel eines generativen Modells handelt. Es ist - mit dem Faltungsabakus beginnend - heuristisch und in sich konsistent. Das generative Potential gestattet nicht nur die Einführung eines Streuungsmaßes, sondern auch die Erklärung für das Auftreten quadratischer Summanden $(x_i - \mu)^2 \cdot q_i$ und damit zusammenhängend die Additivität von Streuungsmaß und Erwartungswert unter Faltungen. Auch der elastische Stab ist im Sinne von Fischbein und Polya ein heuristisches Modell, doch sind seine generativen Eigenschaften in Bezug auf die Stochastik nicht so ausgeprägt wie die des Münzmodells; sie erschöpfen sich in der Herleitung der Varianzformel, der Tschebyscheff-Ungleichung und der Minimalitätseigenschaft des Erwartungswertes. Eine intrinsische Begründung der Additivität unter Faltungen liefert nur das Münzmodell. Dafür läßt sich der elastische Stab auch ohne den Zusammenhang mit der Stochastik hervorragend zu einer beziehungshaltigen Einführung quadratischer

Funktionen in der Sekundarstufe I verwenden:

Dreisatzrechnungen haben zusammen mit einer propädeutischen Behandlung des Funktionsbegriffs einen festen Platz im Curriculum am Ende der Orientierungsstufe:

Es werden Fahrzeiten bei variablen Weglängen

$x \mapsto t=kx$, direkte Proportionalität

und variablen Geschwindigkeiten

$x \mapsto t=k(1/x)$, umgekehrte Proportionalität

berechnet.

Auch der Durchbiegungswinkel eines symmetrisch belasteten Stabes ist bei festgehaltenen Angriffspunkten der Belastung proportional. Variiert man statt der Gewichte ihre Aufhängungspunkte, dann gilt offensichtlich wieder: Je größer die Entfernung vom Schwerpunkt ist, desto größer ist auch der Durchbiegungswinkel. Wieder tippt man auf Proportionalität. Doch welche Überraschung, wenn man statt der erwarteten linearen Abhängigkeit $x \mapsto kx$ die quadratische $x \mapsto kx^2$ entdeckt.

Daß dieses Modell später für den Ausbau der Analogie Erwartungswert - Schwerpunkt zur Analogie Varianz - Durchbiegungswinkel herangezogen werden kann, steigert die Anwendungsmöglichkeit. Es wird didaktisch noch beziehungshaltiger, wenn Schüler das entsprechende Modell mit einem starren Stab zur Lösung von Produktgleichungen eingesetzt und dabei (neben der Mathematik) ein intuitives Gefühl für Schwerpunkt und Drehmomente gewonnen haben [Bruner 2, S. 66f].

5 Brownsche Molekularbewegung im elastischen Zentralkraftfeld, der Grenzwertsatz von de Moivre-Laplace

5.1 Vorbemerkung

Es ist üblich, die Binomialverteilung mittels kombinatorischer Überlegungen einzuführen und den Grenzwertsatz von de Moivre-Laplace mit der Stirlingschen Formel zu beweisen. Dieser Beweis, der auf die Originalarbeit von Laplace zurückgeht, gilt als recht schwierig. Eine Begründung, die dem Schulniveau eher angemessen ist, geben [Böhme], [Bangen], [Heigel] und [Strehl], indem sie die Binomialverteilungen logarithmieren und plausibel machen, daß die entstandenen Funktionen durch Parabeln angenähert werden.

Hier soll eine projektorientierte Alternative zu diesem Themenkreis ausgearbeitet werden, in deren Zentrum der Prozeß des Mathematisierens steht. Der Mathematisierungsprozeß liefert die Binomialverteilung (ohne Kombinatorik) als stationäre Verteilung einer Irrfahrt. Er findet seinen Abschluß durch eine genetische Begründung des Grenzwertsatzes von de Moivre-Laplace, die durch das zu mathematisierende Problem strukturiert ist.

Daß sich die Überlegungen auch konsequent an die Ausführungen der Kapitel 1 bis 3 zum umgekehrten Wahrscheinlichkeitsbaum (vgl. 2.5) anschließen lassen, wird 5.6 am Ende dieses Kapitels belegen. Doch zunächst ein kurzer Überblick:

Zu 5.2: Ausgangspunkt ist die "Ratterkiste" ¹⁾, ein Modell zur Brownschen Molekularbewegung in einem elastischen Zentralkraftfeld. Die quantitative Auswertung der Moleküldichte in Abhängigkeit vom Ort liefert eine glockenförmige Verteilungsfunktion. Dieses Phänomen soll qualitativ und quantitativ erklärt werden.

1) Der Name kommt vom hohen Geräuschpegel beim Betrieb.

5.3: Dazu konstruieren wir in einem ersten Schritt ein diskretes mathematisches Irrfahrtmodell, das für die (im Experiment beobachtete) stationäre Verteilung s die Differenzgleichung $s_{k+1} = \frac{N-k}{k+1} s_k$ liefert. Sie führt unmittelbar auf explizite Formeln für die Binomialkoeffizienten und die binomiale Grenzverteilung.

5.4: Bei Verfeinerung der Diskretisierung zur besseren Anpassung des Modells an die kontinuierliche Realität geht die Differenzgleichung über in die (sofort integrierbare) Differentialgleichung $0 = \varphi'(x) + x\varphi(x)$ der Normalverteilung. Damit sind sowohl das experimentelle Ergebnis wie auch der Grenzwertsatz von de Moivre-Laplace (für $p=q=1/2$) begründet.

In 5.5 wird die Begründung auf den Fall $p \neq 1/2$ übertragen.

Es sei erwähnt, daß sich die Irrfahrt 5.3 auch bei dem bekannten Ehrenfestmodell zur Diffusion ergibt, doch ist in unserem Zusammenhang wichtig, daß nur die Brownsche Interpretation einen intrinsisch motivierten Übergang zum Grenzwertsatz von de Moivre-Laplace ermöglicht. Nur hier ist nämlich die Verfeinerung der Diskretisierung und der Zwang zur Normierung der Binomialverteilung problemimmanent, d.h. durch den Mathematisierungsprozeß bedingt. In der Ehrenfest'schen Interpretation hat sie keinen Sinn.

Man erkennt: die Ratterkiste ist mehr als ein bloßes Demonstrationsobjekt. Ihre Mathematisierung steuert vergleichsweise komplexe und verallgemeinerungsfähige ¹⁾ Überlegungen. In diesem Sinne scheint es durchaus gerechtfertigt, ihr Züge eines generativen Modells zuzusprechen.

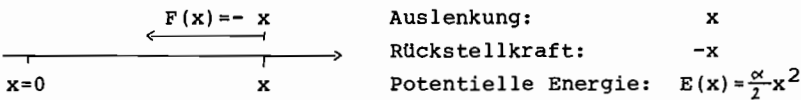
1) Vgl. Kapitel 6, [Feller 1, S. 354 f, 377 f], [Kac, S. 396 f]. Die Arbeiten der genannten Autoren zum zeitlichen Ablauf von Diffusionsvorgängen gaben den Anstoß für das vorliegende Kapitel.

5.2 Simulation der Brownschen Bewegung im elastischen Zentralkraftfeld

5.2.1 Physikalischer Hintergrund

Die eindimensionale Brownsche Molekularbewegung in einem elastischen Zentralkraftfeld ist dadurch definiert, daß auf ein durch zufällige Zusammenstöße gestörtes Teilchen eine Kraft wirkt, die proportional zur Auslenkung aus der Ruhelage ist ¹⁾.

Skizze:



Nach dem Boltzmannschen Verteilungssatz wird die Wahrscheinlichkeit dafür, daß sich das Teilchen an der Stelle x befindet, beschrieben durch die Dichte ²⁾

$$\varphi(x) = \varphi_0 e^{-\frac{E(x)}{kT}} = \varphi_0 e^{-\frac{1}{2} \frac{\alpha}{kT} x^2}, \quad \varphi_0 = \frac{1}{\sqrt{2\pi} \sqrt{kT/\alpha}},$$

die eine Normalverteilung mit $\mu=0$ und $\sigma^2=kT/\alpha$ darstellt. In einem Gas kann man diese Wahrscheinlichkeit auch als Teilchendichte interpretieren.

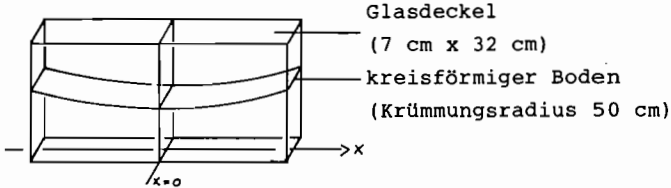
1) Z. B. der Spiegel eines empfindlichen Galvanometers in der thermisch bewegten Luft [Pohl, S. 287]. Obwohl wir uns im folgenden mit der Ratterkiste auf Gaspartikel beschränken werden, haben die anhand des Irrfahrtmodells gewonnenen Ergebnisse Gültigkeit auch für beliebige gestörte, elastisch gegengekoppelte Systeme wie stabilisierte Spannungsquellen, drehzahlstabilisierte Motoren, zielende Schützen u.s.w.. Sie liefern neben dem Grenzwertsatz von de Moivre-Laplace, die Erkenntnis, daß solche Systeme normalverteilte Zustandsdichten besitzen.

2) k ist die Boltzmannkonstante, T die absolute Temperatur.

5.2.2 Experiment

Die Brownsche Bewegung wird durch einen Kasten mit kreisförmig gebogenem Boden simuliert, in dem sich mehrere Kugeln eines Modellgases befinden. Jede einzelne dieser Kugeln kann als mathematisches Pendel aufgefaßt werden, das einer zur Auslenkung proportionalen Rückstellkraft unterliegt.

Skizze:



Der Kasten wird auf einen Vibrator geschraubt, welcher für ständige "thermische" Bewegung der Modellgaskugeln sorgt. Eine quantitative Auswertung der Kugeldichte in Abhängigkeit vom Ort ist durch Auszählen einiger Momentaufnahmen leicht möglich. Darüberhinaus kann man durch Markieren einer einzigen Kugel ihre Bahn verfolgen und erkennen, daß die räumliche Verteilung der Kugeldichte gleichzeitig die Aufenthaltswahrscheinlichkeit für eine festgehaltene Kugel angibt.

5.2.3 Quantitative Ergebnisse

Zur Auszählung von 10 Momentaufnahmen mit je 100 Kugeln wurde das Intervall $[-16; 16]$ der x -Achse in 16 Teilintervalle der Breite 2 zerlegt. Die quantitativen Ergebnisse sind der Tabelle auf der folgenden Seite zu entnehmen. Die relative Häufigkeitsverteilung r_{1000} der insgesamt 1000 Kugeln liefert ein Stichprobenmittel $\mu = -0,002$, eine Stichprobenstandardabweichung $\sigma = 4,175$, und ein Vergleich mit χ^2 belegt eindrucksvoll, wie gut die experimentellen Ergebnisse mit der theoretisch zu erwartenden Normalverteilung übereinstimmen. Es sei angemerkt, daß der χ^2 Test die Testgröße 6,9 liefert. Die 95% Grenze liegt bei 21, so daß die experimentellen Ergebnisse auch im Sinne der Statistik der Normalverteilungshypothese nicht widersprechen.

Auswertung von 10 Aufnahmen mit je 100 Kugeln

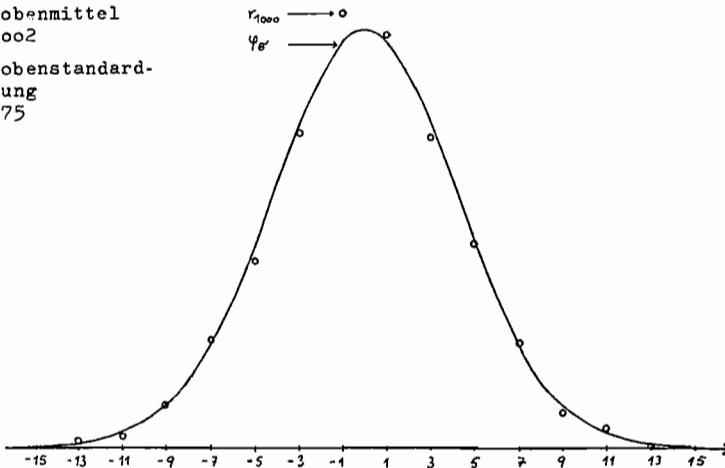
		Mittelpunkte der Teilintervalle														
Nr. der Aufnahme	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11	13	15
1			1	1	3	8	16	25	19	11	7	5	2	1	1	
2				4	7	10	11	21	25	9	10	3				
3				3	4	8	16	10	21	16	8	4	1	1		
4		1		2	2	9	11	22	21	19	8	4	1			
5		1		2	5	9	19	16	15	15	9	7	2			
6				1	8	8	12	23	21	10	10	3	2	2		
7		1		1	5	9	16	16	21	14	11	4	1	1		
8			2	3	4	11	11	19	12	16	11	5	4	2		
9			1	1	5	4	16	20	16	14	10	9	2	2		
10			1	2	6	9	15	19	17	17	9	4	1			
a_{1000}		3	5	20	49	85	143	199	188	141	93	48	16	9	1	
r_{1000}	.000	.003	.005	.020	.049	.085	.143	.199	.188	.141	.093	.048	.016	.009	.001	.000
ψ_{σ}	.000	.001	.006	.019	.047	.093	.148	.186	.186	.142	.093	.047	.019	.006	.001	.000

Stichprobenmittel

$$\mu = -0.002$$

Stichprobenstandard-
abweichung

$$\sigma = 4.175$$



5.3 Diskretes Modell der Brownschen Bewegung, die Binomialverteilung als stationäre Verteilung

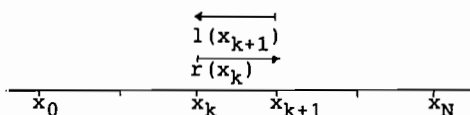
Wie läßt sich das experimentelle Ergebnis ohne Rückgriff auf den Boltzmannschen Verteilungssatz erklären?

5.3.1 Diskretes Modell

Wir greifen eine Kugel unseres Gases heraus und machen über ihre Bewegung folgende Modellannahmen:

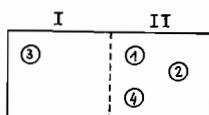
- Die Kugel kann sich nur an diskreten Stellen $x_k = k\delta, \delta > 0, k=0, \dots, N$ des Zahlenstrahls aufhalten.
- Sie soll je Zeiteinheit einen Sprung der Weite δ machen, und zwar von der Stelle x_k aus mit der Wahrscheinlichkeit $l_k = l(x_k)$ um δ nach links, mit $r_k = r(x_k)$ um δ nach rechts. Für $k=0$ kann sie nur nach rechts springen, also $r(x_0)=1, l(x_0)=0$, entsprechend $r(x_N)=0, l(x_N)=1$.
- Je weiter rechts die Kugel sich befindet, desto größer wird die Wahrscheinlichkeit sein, daß sie (wegen der elastischen Kraft) nach links springt. Auf Grund des linearen Kraft-Weg Gesetzes nehmen wir an, daß diese Wahrscheinlichkeit linear vom Ort des Teilchens abhängt. Mit obigen Randbedingungen ergibt sich $l_k = k/N$ und $r_k = (N-k)/N$.

Skizze:



Anmerkung: Diese Irrfahrt im Zustandsraum $\{0, \dots, N\}$ ($\delta=1$)

liegt auch dem Ehrenfestmodell zu Grunde: Gegeben sind N Kugeln in einem Behälter, der in zwei Hälften unterteilt ist. In jeder Zeiteinheit wechselt eine Kugel, deren Nummer zufällig ausgewählt wurde, die Seite des



$N=4$: Zustand $k=1$

Behälters. k sei der Zustand, der dadurch beschrieben ist,

daß sich k Kugeln in der Hälfte I befinden und $N-k$ Kugeln in der Hälfte II. Bezeichnet man mit l_k bzw. r_k die Übergangswahrscheinlichkeiten vom Zustand k in den Zustand $k-1$ bzw. $k+1$, dann gilt wieder $l_k = k/N$ (denn mit Wahrscheinlichkeit k/N liegt die ausgewählte Kugel, die ihren Platz wechseln muß, in der Hälfte I) und $r_k = (N-k)/N$.

Wie [Kac] mitteilt, wurde das Ehrenfestmodell von [P. und T. Ehrenfest] 1907 im Zusammenhang mit dem Boltzmannschen H-Theorem konstruiert, der Zusammenhang mit der Brownschen Bewegung im elastischen Zentralkraftfeld wurde erstmalig 1926 von [Schrödinger und Kohlrausch] erwähnt.

5.3.2 Binomialverteilung

Es sei $s_k = s(x_k)$, $k=0, \dots, N$, die "stationäre" Wahrscheinlichkeitsverteilung dafür, daß sich unser Teilchen an der Stelle $x_k = k\delta$ aufhält. Da es im Mittel genauso oft von der Stelle x_k zur Stelle x_{k+1} springen wird wie zurück, hat man das Gleichungssystem

$$s_{k+1} \cdot l_{k+1} = s_k \cdot r_k,$$

das wegen $r_k = (N-k)/N$, $l_{k+1} = (k+1)/N$ übergeht in

$$s_{k+1} = \frac{N-k}{k+1} \cdot s_k \quad (0 \leq k \leq N-1).$$

Die Lösung kann man sofort hinschreiben:

$$\begin{aligned} s_1 &= \frac{N}{1} s_0 \\ s_2 &= \frac{N-1}{2} s_1 = \frac{N(N-1)}{1 \cdot 2} s_0 \\ s_3 &= \frac{N-3}{3} s_2 = \frac{N(N-1)(N-2)}{1 \cdot 2 \cdot 3} s_0 \\ &\vdots \\ s_k &= \frac{N-k+1}{k} s_{k-1} = \frac{N(N-1) \dots (N-k+1)}{1 \cdot 2 \cdot \dots \cdot k} s_0. \end{aligned}$$

Die Faktoren sind die Binomialkoeffizienten, deren explizite Form sich hier "gleichsam von selbst" ergibt. Aus der Be-

dingung $s_0 + \dots + s_N = 1$ folgt $s_0 = (1/2)^N$, und als stationäre Verteilung erhält man die Binomialverteilung

$$s_k = s(x_k) = \binom{N}{k} (1/2)^N.$$

5.3.3 Normierung

Das diskrete Modell paßt noch nicht zur Ratterkiste, denn der Erwartungswert der empirischen Verteilung aus 5.2 liegt bei $\mu = 0$, die Varianz ist σ^2 . Der Erwartungswert unserer (durch den Faktor δ gestauchten) Binomialverteilung liegt bei $(N/2)\delta$, die Varianz ist $(N/4)\delta^2$.

Wir wählen deshalb in unserem Modell δ (bzw. N) so, daß gilt $(N/4)\delta^2 = \sigma^2$ ($\delta = \sigma / (1/2) \sqrt{N}$) und verschieben unsere gestauchte Binomialverteilung um $(N/2)\delta$ nach links. Es entsteht eine Verteilung $s(x)$, die jetzt an den Stellen

$$x = x_k = (k - (N/2))\delta = \sigma \frac{k - N/2}{(1/2)\sqrt{N}}$$
 definiert ist mit

$$s(x_k) = s\left(\sigma \frac{k - N/2}{(1/2)\sqrt{N}}\right) = \binom{N}{k} (1/2)^N.$$

Wir drücken auch die Übergangswahrscheinlichkeiten durch die Variable x_k aus:

$$l(x_k) = \frac{k}{N} = \frac{1}{2} + \frac{1}{\delta N} x_k = \frac{1}{2} + \frac{\delta}{4\delta^2} x_k$$

$$r(x_k) = \frac{N-k}{N} = \frac{1}{2} - \frac{1}{\delta N} x_k = \frac{1}{2} - \frac{\delta}{4\delta^2} x_k.$$

Aus der Gleichgewichtsbedingung

$$r(x_k) s(x_k) = l(x_{k+1}) s(x_{k+1}) \quad \text{ergibt sich}$$

$$s(x_k) \left(\frac{1}{2} - \frac{\delta}{4\delta^2} x_k\right) = s(x_{k+1}) \left(\frac{1}{2} + \frac{\delta}{4\delta^2} x_{k+1}\right).$$

1) Für $\delta = 1$ ist dies die übliche Normierung der Binomialverteilung durch Verschiebung um $N/2$ (für den Erwartungswert 0) und Stauchung um $(1/2)\sqrt{N}$ (für die Varianz 1). Hier ist die Normierung durch den Modellbildungsprozeß motiviert.

5.4 Verfeinerung der Diskretisierung, Normalverteilung

Um unser diskretes Modell der kontinuierlichen Realität besser anzupassen, verkleinern wir die Sprungweite δ . Ist $\varphi_\delta(x)$ die Wahrscheinlichkeitsdichte für den Aufenthaltsort eines Teilchens, so gilt $s(x_k) \approx \varphi_\delta(x_k) \delta$, und die Genauigkeit der Annäherung wird mit kleiner werdendem δ immer größer. Die Gleichgewichtsbedingung geht über in

$$\delta \varphi_\delta(x) \left(\frac{1}{2} - \frac{\delta}{4\delta^2} x \right) = \delta \varphi_\delta(x+\delta) \left(\frac{1}{2} + \frac{\delta}{4\delta^2} (x+\delta) \right) \quad \text{bzw.}$$

$$\varphi_\delta(x) \left(\frac{1}{2} - \frac{\delta}{4\delta^2} x \right) = \varphi_\delta(x+\delta) \left(\frac{1}{2} + \frac{\delta}{4\delta^2} (x+\delta) \right) ,$$

und wir können wegen $\delta \rightarrow 0$ eine Entwicklung von φ_δ vornehmen. Dabei vernachlässigen wir alle Terme zweiter Ordnung:

$$\begin{aligned} \frac{\varphi_\delta(x)}{2} - \frac{\delta}{4\delta^2} x \varphi_\delta(x) &= (\varphi_\delta(x) + \delta \varphi_\delta'(x)) \left(\frac{1}{2} + \frac{\delta}{4\delta^2} (x+\delta) \right) \\ - \frac{\delta}{4\delta^2} x \varphi_\delta(x) &= \delta \left(\frac{1}{2} \varphi_\delta'(x) + \frac{1}{4\delta^2} x \varphi_\delta(x) \right) \\ - \frac{1}{\delta^2} x \varphi_\delta(x) &= \varphi_\delta''(x) . \end{aligned}$$

Ergebnis: Die Wahrscheinlichkeitsdichte $\varphi_\delta(x)$ für den Aufenthaltsort des Teilchens wird der Differentialgleichung $\varphi_\delta''(x) = -\frac{1}{\delta^2} x \varphi_\delta(x)$ bzw. $\frac{\varphi_\delta''(x)}{\varphi_\delta(x)} = -\frac{1}{\delta^2} x$ genügen.

Integration liefert

$$\ln(\varphi_\delta(x)) = -\frac{1}{2\delta^2} x^2 + k, \quad \text{also } \varphi_\delta(x) = c \cdot e^{-\frac{1}{2} \left(\frac{x}{\delta}\right)^2} .$$

Wegen $\int \varphi_\delta(x) dx = 1$ muß gelten $c = 1/\sqrt{2\pi} \delta$. Dies ist die Dichte der Normalverteilung mit Erwartungswert 0 und Varianz δ^2 . Das im Experiment beobachtete Auftreten der Normalverteilung ist damit begründet. Gleichzeitig haben wir die Formel von de Moivre-Laplace hergeleitet:

$$\begin{aligned} \underline{\underline{B_{N, \frac{1}{2}}(k) = s(x_k) \approx \delta \varphi_\delta(x_k) = \left(\frac{\delta}{\sqrt{2\pi} N}\right) \frac{1}{\sqrt{2\pi} \delta} e^{-\frac{1}{2} \left(\frac{k - N/2}{\delta}\right)^2}}} \\ = \underline{\underline{\frac{1}{\sqrt{N\pi}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{k - N/2}{\sqrt{N/4}}\right)^2}}} . \end{aligned}$$

5.5 Grenzwertsatz von de Moivre-Laplace, allgemeiner Fall

Natürlich kann man die Differentialgleichung $\varphi'(x) = -x\varphi(x)$ (jetzt $\varrho=1$) auch ohne Rückgriff auf die Brownsche Molekularbewegung herleiten, und zwar auch für $p \neq 1/2$: Die Binomialverteilung

$$B_{N,p}(k) = \binom{N}{k} p^k q^{N-k} \quad \text{mit } \mu = Np, \quad \sigma = \sqrt{Npq}$$

ist nach Normierung (Verschiebung um μ nach links, Stauchung mit Faktor $1/\sigma =: \delta$ in x-Richtung, Streckung mit Faktor $\delta = 1/\delta$ in y-Richtung) definiert an den Stellen

$$x_k = \frac{k - \mu}{\delta} = \delta(k - \mu) \quad \text{mit } s(x_k) = \binom{N}{k} p^k q^{N-k}.$$

Sie genügt der Rekursionsformel

$$s(x_{k+1}) \frac{k+1}{p} = \frac{N-k}{q} s(x_k),$$

die wir für $p=q=1/2$ als Gleichgewichtsbedingung interpretiert haben. Wegen

$$k = \frac{x_k}{\delta} + \mu = \frac{x_k}{\delta} + Np, \quad N-k = Nq - \frac{x_k}{\delta}, \quad x_{k+1} = x_k + \delta$$

kann man die Rekursionsformel umformen zu

$$s(x_k + \delta) \left[\frac{x_k + \delta}{\delta p} + N \right] = \left[N - \frac{x_k}{\delta q} \right] s(x_k)$$

bzw. nach Division durch N:

$$s(x_k + \delta) \left[1 + \frac{1}{\delta p N} (x_k + \delta) \right] = s(x_k) \left[1 - \frac{1}{\delta q N} x_k \right].$$

Wegen $\delta^2 = 1/Npq$, also $1/\delta p N = \delta q$ und $1/\delta q N = \delta p$ folgt

$$s(x_k + \delta) \left[1 + \delta q (x_k + \delta) \right] = s(x_k) \left[1 - \delta p x_k \right].$$

Wir bezeichnen die Grenzfunktion zu den normierten Binomialverteilungen, deren Existenz wir voraussetzen, mit $\varphi(x)$. Es

gilt $\varphi(x) \approx s(x)$, wobei die Näherung mit wachsendem N ($N \rightarrow \infty$, $\delta \rightarrow 0$) immer besser wird. Aus der letzten Formel folgt dann nach Entwicklung von $\varphi(x)$ und Vernachlässigung von Termen höherer Ordnung in δ :

$$[\varphi(x) + \delta \varphi'(x)] [1 + \delta qx] = \varphi(x) [1 - \delta px]$$

$$\varphi(x) + \delta [\varphi'(x) + qx\varphi(x)] = \varphi(x) - \delta px\varphi(x)$$

und hieraus wegen $p+q=1$

$$\varphi'(x) = -x \varphi(x)$$

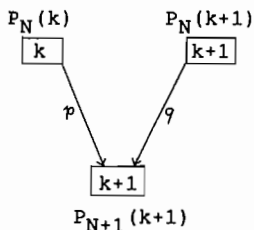
mit der Standardnormalverteilung $\varphi(x) = c e^{-\frac{1}{2}x^2}$, $c = 1/\sqrt{2\pi}$ als Lösung.

5.6 Der umgekehrte Wahrscheinlichkeitsbaum und Einführung der Binomialverteilung ohne Kombinatorik

Wie in 5.2 und 5.3 gezeigt, läßt sich die Binomialverteilung sehr einfach und ohne Rückgriff auf kombinatorische Überlegungen als stationäre Verteilung einer Irrfahrt konstruieren. Die Gleichgewichtsbedingung, die sofort auf die explizite Formel führt, ist dem Graphen der Irrfahrt unmittelbar zu entnehmen. In diesem Abschnitt soll abschließend skizziert werden, wie man diesen Zugang in einen größeren Zusammenhang stellen und zu einer beziehungshaltigen Einführung der Binomialverteilung ausbauen kann, in dessen Mittelpunkt rekursive Berechnungen stehen. Die Darstellung durch den umgekehrten Wahrscheinlichkeitsbaum schließt dabei nahtlos an den Faltungsabakus aus Kapitel 3 an. Sie läßt, wie schon in 2.4 angedeutet, die strukturellen Gemeinsamkeiten zwischen Faltungen (Spezialfall: Pascalsches Dreieck) und der rekursiven Berechnung beliebiger Markoffketten hervortreten.

Wir erinnern uns an den Vierfachmünzwurf, der in 1.2 auf der

Grundlage des subjektiven Wahrscheinlichkeitsbegriffs behandelt wurde. Die zugehörigen Laplace-Wahrscheinlichkeiten $P_N(k)$ (für k Köpfe in N Würfeln) können wir durch Faltung mit einem umgekehrten Wahrscheinlichkeitsbaum rekursiv berechnen:



"Zentrale Frage:"

Welche Wege führen zu $k+1$ Köpfen in $N+1$ Versuchen?

Antwort:

1. Weg: k Köpfe in N Würfeln und ein weiterer im $N+1$ ten Wurf, Wahrscheinlichkeit $p \cdot P_N(k)$.
2. Weg: $k+1$ Köpfe in N Würfeln und kein weiterer im $N+1$ ten Wurf, Wahrscheinlichkeit $q \cdot P_N(k+1)$.

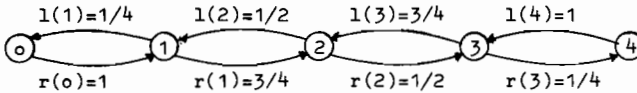
Er repräsentiert die Rekursionsformel

$$(1) \quad P_{N+1}(k+1) = p P_N(k) + q P_N(k+1),$$

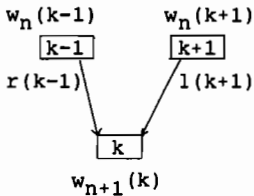
die folgende Tabelle liefert (Pascalsches Dreieck):

	faire Münze ($p=q=1/2$)					gezinkte Münze				
	Anzahl der Köpfe					Anzahl der Köpfe				
	0	1	2	3	4	0	1	2	3	4
P_1	1/2	1/2				$1q$	$1p$			
* P_1	$1/2 \swarrow \searrow 1/2$					$p \swarrow \searrow q$				
$P_2 = P_1 * P_1$	1/4	1/2	1/4			$1q^2$	$2qp$	$1p^2$		
* P_1	$1/2 \swarrow \searrow 1/2$					$p \swarrow \searrow q$				
$P_3 = P_1 * P_1 * P_1$	1/8	3/8	3/8	1/8		$1q^3$	$3q^2p$	$3qp^2$	$1p^3$	
* P_1	$1/2 \swarrow \searrow 1/2$					$p \swarrow \searrow q$				
$P_4 = P_1 * P_1 * P_1 * P_1$	1/16	4/16	6/16	4/16	1/16	$1q^4$	$4q^3p$	$6q^2p^2$	$4qp^3$	$1p^4$

Die Übertragung der rekursiven Berechnungsmethode auf das Irrfahrtmodell, hier für $N=4$



stellt ein angemessenes Problem dar. Bezeichnet $w_n(k)$ die Wahrscheinlichkeit, sich zur Zeit n im Zustand k aufzuhalten, dann hat man



Zentrale Frage:

Welche Wege führen zu einem Aufenthalt im Punkt k zur Zeit $n+1$?

Antwort:

1. Weg: Aufenthalt im Punkt $k-1$ zur Zeit n und Sprung nach rechts
2. Weg: Aufenthalt im Punkt $k+1$ zur Zeit n und Sprung nach links.

Rekursionsformel:

$$(2) \quad w_{n+1}(k) = r(k-1)w_n(k-1) + l(k+1)w_n(k+1)$$

$$= \frac{N-k+1}{N} w_n(k-1) + \frac{k+1}{N} w_n(k+1) .$$

Für einen Start in 2 bzw. 0 erhält man die Tabellen auf der folgenden Seite.

Der Unterschied zum vertrauten Faltungsprinzip besteht nur darin, daß sich die Wahrscheinlichkeiten des umgekehrten "Faltungsbaumes" beim Entlangsschieben an einer Zeile verändern. Der Tabelle kann man entnehmen,

- daß der Einfluß der Anfangsverteilung sehr schnell abnimmt
- daß man (hier) zwischen zwei Verteilungen w_g, w_u hin- und herpendelt, je nachdem ob n gerade oder ungerade ist ¹⁾.

1) Es liegt eine irreduzible periodische Markoffkette vor;

	Start in 2					Start in 0				
	0	1	2	3	4	0	1	2	3	4
w_0			1.00			1.00				
	$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$				
w_1		.500		.500			1.00			
	$\begin{array}{c} \swarrow 3/4 \quad \searrow 3/4 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 3/4 \quad \searrow 3/4 \\ \downarrow \end{array}$				
w_2	.125		.750		.125	.250		.750		
	$\begin{array}{c} \swarrow 1/2 \quad \searrow 1 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 1/2 \quad \searrow 1 \\ \downarrow \end{array}$				
w_3		.500		.500			.625		.375	
	$\begin{array}{c} \swarrow 1/4 \quad \searrow 0 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 1/4 \quad \searrow 0 \\ \downarrow \end{array}$				
w_4	.125		.750		.125	.156		.750		.094
	$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$				
w_5		.500		.500			.531		.469	
	$\begin{array}{c} \swarrow 0 \quad \searrow 1/4 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 0 \quad \searrow 1/4 \\ \downarrow \end{array}$				
w_6	.125		.750		.125	.133		.750		.117
	$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$					$\begin{array}{c} \swarrow 1 \quad \searrow 1/2 \\ \downarrow \end{array}$				
w_7		.500		.500			.508		.492	
w_g	1/8		3/4		1/8					
w_u		1/2		1/2						
s	1/16	1/4	3/8	1/4	1/16					

$(s = \frac{1}{2} w_g + \frac{1}{2} w_u)$

Stellt man eine Beobachtung in einem zufällig ausgewählten Zeitpunkt n an, dann wird n in der Hälfte aller Fälle gerade, in der anderen Hälfte ungerade sein, woraus sich für die zeitunabhängige Aufenthaltswahrscheinlichkeit der Kugel an der Stelle k ergibt: $s(k) = 0,5w_g(k) + 0,5w_u(k)$.

Man kann sich die Überraschung beim Vergleich der stationären Verteilung $s(k) = s_4(k)$ mit der Verteilung $P_4(k)$ zum Vierfachmünzwurf einer Laplace-Münze vorstellen. Besteht die Gleichheit $s_4(k) = P_4(k)$ auch für andere Werte $N \neq 4$?

solche Ketten besitzen eindeutige stationäre Verteilungen s , aber nicht für alle Anfangsverteilungen w_0 hat man Konvergenz $w_0 \rightarrow s$.

Zur Beantwortung dieser Frage bedenkt man, daß die Rekursionsformel (2)

$$w_{n+1}(k) = \frac{N-k+1}{N} w_n(k-1) + \frac{k+1}{N} w_n(k+1)$$

für die stationäre Verteilung $s(k) = s_N(k)$ (N bezeichnet im Gegensatz zu n die Anzahl der Zustände) übergeht in

$$(3) \quad s_N(k) = \frac{N-k+1}{N} s_N(k-1) + \frac{k+1}{N} s_N(k+1)$$

oder ausgeschrieben:

$$\left. \begin{array}{l} s_N(0) = \frac{1}{N} s_N(1) \\ s_N(1) = \frac{N}{N} s_N(0) + \frac{2}{N} s_N(2) \\ s_N(2) = \frac{N-1}{N} s_N(1) + \frac{3}{N} s_N(3) \\ \cdot \\ \cdot \\ s_N(N) = \frac{1}{N} s_N(N-1) \end{array} \right\} (3) \quad \left. \begin{array}{l} s_N(1) = N s_N(0) \\ s_N(2) = \frac{N-1}{2} s_N(1) \\ s_N(3) = \frac{N-2}{3} s_N(2) \\ \cdot \\ \cdot \\ s_N(N) = \frac{1}{N} s_N(N-1) \end{array} \right\} (4)$$

Durch Einsetzen einer Gleichung in die jeweils folgende erhält man das rechtsstehende Gleichungssystem (4) für die stationäre Verteilung, das in 5.3 unmittelbar dem Graphen der Irrfahrt entnommen wurde, und das die explizite Darstellung der Binomialverteilung $s_N(k) = \binom{N}{k} 0,5^N$ liefert¹⁾.

Die vermutete Gleichheit von $s_N(k) = P_N(k)$ folgt nun daraus, daß $s_N(k)$ auch die Rekursionsformel (1) der $P_N(k)$ erfüllt.

Zur experimentellen Überprüfung der eben durchgeführten Berechnungen simulieren wir die Irrfahrt für $N=4$:

Die erste Tabelle gibt ihren Verlauf über 100 Zeiteinheiten wider (Start im Zustand 2). Die zweite Tabelle zeigt die relativen Besuchshäufigkeiten der einzelnen Stationen nach

1) Hinter den Rekursionsformeln (1) und (4) für $P_N(k)$ bzw. $s_N(k)$ stehen die Rekursionsformeln $\binom{N+1}{k+1} = \binom{N}{k} + \binom{N}{k+1}$ bzw. $\binom{N}{k+1} = \frac{N-k+1}{k+1} \binom{N}{k}$.

100 bzw. 1000 Zeiteinheiten - und die Binomialverteilung zum Vergleich:

Protokoll der Irrfahrt:

2 3 4 3 2 3 2 3 2 1 0 1 2 3 2 3 2 3 4 3 2 3 4 3 4
 3 2 1 2 3 2 3 4 3 4 3 2 1 2 1 2 3 2 1 2 3 2 3 2 3
 2 3 2 3 2 1 2 1 2 3 2 3 2 1 0 1 2 3 2 1 0 1 0 1 2
 3 2 3 2 1 0 1 2 1 2 3 4 3 2 3 2 1 2 1 2 1 0 1 2 3

relative Besuchshäufigkeiten und Binomialverteilung:

k	0	1	2	3	4
$r_{100}(k)$	0.0600	0.200	0.370	0.300	0.0700
$r_{1000}(k)$	0.0720	0.253	0.361	0.244	0.0650
$B_{4,1/2}(k)$	0.0625	0.250	0.375	0.250	0.0625

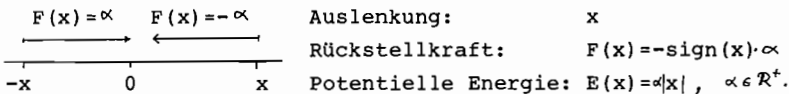
6 Brownsche Molekularbewegung im konstanten Zentralkraftfeld

Dieses Kapitel ist eine Ergänzung zu Kapitel 5. Es soll skizziert werden, wie sich das Experiment und die Theorie zur Normalverteilung auf die Exponentialverteilung übertragen lassen.

6.1 Simulation der Brownschen Bewegung im konstanten Zentralkraftfeld

6.1.1 Physikalischer Hintergrund

Wir betrachten die eindimensionale Brownsche Molekularbewegung, diesmal nicht in einem elastischen, sondern in einem konstanten Zentralkraftfeld gemäß der folgenden Skizze:



Nach dem Boltzmannschen Verteilungssatz ist die Aufenthaltswahrscheinlichkeit eines Teilchens wiederum gegeben durch die Dichte

$$\psi(x) = \psi_0 e^{-\frac{E(x)}{kT}} = \psi_0 e^{-\frac{\alpha}{kT}|x|}, \quad \psi_0 = \frac{\alpha}{2kT},$$

die eine zweiseitige Exponentialverteilung mit $\mu=0$ und $\sigma=1/(\sqrt{2}\psi_0)=\sqrt{2}kT/\alpha$ darstellt. Man kann sie auch schreiben in der Form

$$\psi(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{|x|}{\sigma}}$$

Anmerkung: Läßt man Bewegungen mit $x < 0$ nicht zu, dann hat man stattdessen eine einseitige Exponentialverteilung

$$\psi(x) = \begin{cases} \psi_0 \cdot e^{-\frac{\alpha}{kT}x}, & x \geq 0 \\ 0, & x < 0 \end{cases}; \quad \psi_0 = \frac{\alpha}{kT}$$

mit $\mu = \sigma = KT/\alpha$. Man kann sie auch schreiben als

$$\psi(x) = \frac{1}{\sigma} e^{-\frac{x}{\sigma}} \text{ für } x \geq 0, \quad \psi(x) = 0 \text{ für } x < 0.$$

Wir werden uns auf die zweiseitige Exponentialverteilung konzentrieren; ergänzende Anmerkungen zum einseitigen Fall finden sich in 6.4.

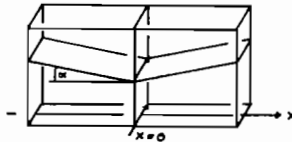
6.1.2 Experiment

Zur Simulation der Brownschen Bewegung im konstanten Zentralkraftfeld ersetzen wir den kreisförmigen Boden unserer Ratterkiste durch einen ebenen, der zur Mitte hin geneigt ist.

Skizze:

Länge der Kiste: $l=32 \text{ cm}$

Neigungswinkel: $\alpha = 5^\circ$



6.1.3 Quantitative Ergebnisse

Die Ergebnisse der Auswertung, diesmal von 14 Momentaufnahmen mit je 50 Kugeln, sind der Tabelle auf der folgenden Seite zu entnehmen. Die relative Häufigkeitsverteilung r_{700} der insgesamt 700 Kugeln liefert ein Stichprobenmittel $\mu=0,008$ und eine Stichprobenstandardabweichung $\sigma=4,413$. Wie ein Vergleich mit der zugehörigen Exponentialverteilung zeigt, stimmen die experimentellen Ergebnisse mit den theoretisch zu erwartenden wieder gut überein, von der "zugehörigen" Normalverteilung weichen sie dagegen beträchtlich ab.

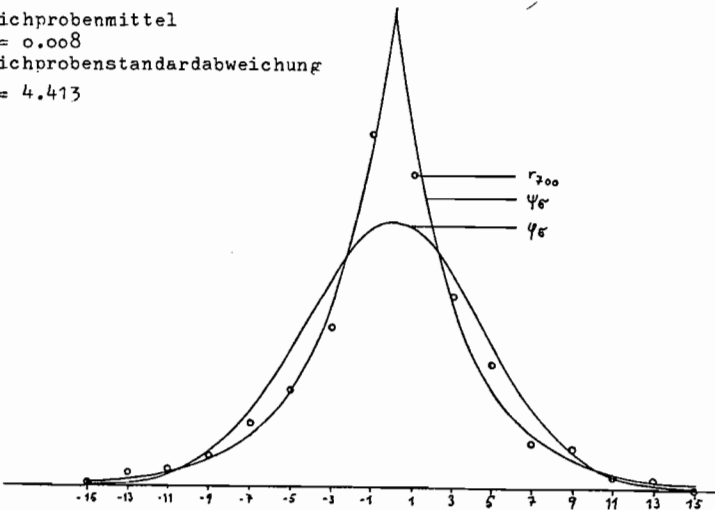
Auswertung von 14 Aufnahmen mit je 50 Kugeln

Nr. der Aufnahme	Mittelpunkte der Teilintervalle															
	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11	13	15
1	1	1	2	2		2	3	10	10	7	6	1	4		1	
2		1	1	1	2	3	8	12	9	7	3	2		1		
3			1			3	5	12	14	7	5	3				
4			1	1	2	2	5	13	10	5	8	1	1		1	
5				2	6	3	2	15	6	8	3		3	1	1	
6			1	2	1	3	10	9	7	8	5	1	1	2		
7				2	3	2	2	14	10	6	8	2	1			
8		1		2	3	5	7	11	8	5	4	2	1	1		
9	1	1		1	3	1	6	8	16	6		4	1		2	
10					6	5	5	13	7	6	5		2		1	
11					1	6	5	13	16	4	2	1	2			
12					1	6	4	13	17	4	2	1	2			
13		1		1	1	2	9	14	8	9	4	1				
14		1	2		1	4	4	11	11	9	4	2	1			
\bar{a}_{700}	2	6	8	14	30	47	75	168	149	91	59	21	19	6	5	
r_{700}	.003	.009	.011	.020	.043	.067	.107	.240	.213	.130	.084	.030	.027	.009	.007	.000
ψ_{700}	.003	.005	.009	.018	.034	.065	.123	.233	.233	.123	.065	.034	.018	.009	.005	.003
ψ_{700}	.001	.002	.008	.023	.051	.095	.143	.176	.176	.143	.095	.051	.023	.008	.002	.001

Stichprobenmittel

 $\mu = 0.008$

Stichprobenstandardabweichung

 $\sigma = 4.413$ 

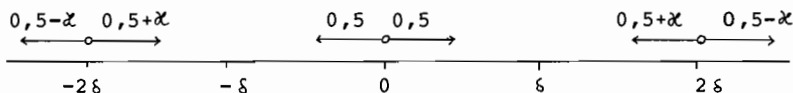
6.2 Diskretes Modell der Brownschen Bewegung

Wir wählen das Modell für die Bewegung einer Kugel unseres "Gases" analog zu 5.3.1:

- (a) Die Kugel soll sich nur an den Stellen $x_k = k \cdot \delta$, $k \in \mathbb{Z}$ aufhalten.
- (b) Je Zeiteinheit soll sie einen Sprung um δ machen, und zwar von der Stelle x_k aus mit der Wahrscheinlichkeit $l(x_k)$ nach links, mit der Wahrscheinlichkeit $r(x_k)$ nach rechts.
- (c) Wegen der für $x_k > 0$ bzw. $x_k < 0$ konstanten Rückstellkraft gelte

$$\left. \begin{array}{l} l(x_k) = 0,5 + \alpha \\ r(x_k) = 0,5 - \alpha \end{array} \right\} \text{für } x_k > 0 \quad \left. \begin{array}{l} l(x_k) = 0,5 - \alpha \\ r(x_k) = 0,5 + \alpha \end{array} \right\} \text{für } x_k < 0. \quad (\alpha > 0)$$

Skizze:



Berechnung der Wahrscheinlichkeitsverteilung:

Die Gleichgewichtsbedingungen für die stationäre Verteilung $s(x_k) := s_k$ lauten:

$$(1) \quad 0,5 \cdot s_0 = (0,5 + \alpha) s_1$$

$$(2) \quad (0,5 - \alpha) s_i = (0,5 + \alpha) s_{i+1}, \quad i > 0 \quad \text{bzw.}$$

$$(3) \quad s_{i+1} = \frac{0,5 - \alpha}{0,5 + \alpha} s_i = \left(\frac{0,5 - \alpha}{0,5 + \alpha} \right)^i \cdot s_1 = q^i s_1,$$

wobei $q := \frac{0,5 - \alpha}{0,5 + \alpha}$ gesetzt wurde. Wegen $0,5 + \alpha = \frac{1}{1+q}$ folgt aus (1)

$$(4) \quad s_0 = \frac{2}{1+q} s_1.$$

Wegen der Symmetrie des Problems hat man weiter $s_i = s_{-i}$, und da die Summe aller Wahrscheinlichkeiten 1 ergeben muß, folgt:

$$\begin{aligned}
 1 &= \sum_{i \in \mathbb{Z}} s_i = s_0 + 2 \sum_{i \in \mathbb{N}} s_i = s_1 \left(\frac{2}{1+q} + 2 \sum_{i \in \mathbb{N}} q^{i-1} \right) \\
 &= s_1 \left(\frac{2}{1+q} + \frac{2}{1-q} \right) = s_1 \frac{4}{1-q^2}, \text{ also}
 \end{aligned}$$

$$(5) \quad s_1 = \frac{1-q^2}{4}.$$

Berechnung der Varianz:

Wegen $\mu=0$ gilt

$$\begin{aligned}
 \sigma^2 &= \sum_{i \in \mathbb{Z}} (\delta \cdot i)^2 s_i = 2\delta^2 \sum_{i \in \mathbb{N}} i^2 s_i = 2\delta^2 \sum_{i \in \mathbb{N}} i^2 q^{i-1} s_1 \\
 &= \frac{2}{q} \delta^2 \left(\sum_{i \in \mathbb{N}} i^2 q^i \right) s_1 = \frac{2}{q} \delta^2 \frac{q(1+q)}{(1-q)^3} s_1 = \frac{2(1+q)}{(1-q)^3} \cdot \frac{(1+q)(1-q)}{4} \delta^2 \\
 &= \frac{1}{2} \left(\frac{1+q}{1-q} \right)^2 \delta^2 = \frac{1}{2} \left(\frac{1}{2\lambda} \right)^2 \delta^2 = \frac{1}{8\lambda^2} \delta^2.
 \end{aligned}$$

Bestimmung von λ :

Um unser Modell den experimentellen Ergebnissen anzupassen, müssen wir λ so wählen, daß die Varianzen gleich werden, also

$$\lambda = \frac{\delta}{\sqrt{8} \sigma}.$$

Die Gleichgewichtsbedingung (2) lautet dann für $x=x_k=k \cdot \delta > 0$:

$$(6) \quad \left(0,5 + \frac{\delta}{\sqrt{8} \sigma}\right) s(x+\delta) = \left(0,5 - \frac{\delta}{\sqrt{8} \sigma}\right) s(x),$$

analog für $x < 0$.

6.3 Verfeinerung der Diskretisierung, Übergang zur Exponentialverteilung

Wir verfeinern die Schrittweite δ in unserem diskreten Modell ($\delta \rightarrow 0$). Wenn $\psi_\delta(x)$ die Wahrscheinlichkeitsdichte für den Aufenthaltsort eines Teilchens ist, dann gilt $s(x_k) \approx \psi_\delta(x_k) \delta$ und die Gleichgewichtsbedingung (6) (für $x > 0$) geht über in

$$\left(0,5 + \frac{\delta}{\sqrt{8} \sigma}\right) \psi_\delta(x+\delta) \delta = \left(0,5 - \frac{\delta}{\sqrt{8} \sigma}\right) \psi_\delta(x) \delta.$$

Division durch δ , Entwicklung von ψ_δ und Vernachlässigung von Termen höherer Ordnung in δ liefert:

$$(0,5 + \frac{\delta}{\sqrt{8}\sigma}) (\psi_\delta(x) + \delta \psi_\delta'(x)) = (0,5 - \frac{\delta}{\sqrt{8}\sigma}) \psi_\delta(x)$$

$$\delta (\frac{1}{\sqrt{8}\sigma} \psi_\delta(x) + 0,5 \psi_\delta'(x)) = -\delta \frac{1}{\sqrt{8}\sigma} \psi_\delta(x)$$

$$0,5 \psi_\delta'(x) = -\frac{2}{\sqrt{8}\sigma} \psi_\delta(x)$$

$$\psi_\delta'(x) = -\frac{\sqrt{2}}{\sigma} \psi_\delta(x) \quad \text{für } x > 0,$$

analog $\psi_\delta'(x) = \frac{\sqrt{2}}{\sigma} \psi_\delta(x) \quad \text{für } x < 0,$

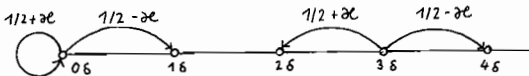
also: $\psi_\delta(x) = c e^{-\frac{\sqrt{2}}{\sigma}|x|}$, $c = \frac{1}{\sqrt{2}\sigma}$ wegen $\int \psi_\delta(x) = 1$.

Dies ist die zweiseitige Exponentialverteilung mit $\mu=0$ und Standardabweichung σ .

6.4 Anmerkung zur einseitigen Exponentialverteilung

Läßt man einen Aufenthaltsort $x < 0$ nicht zu, so hat man in unserem diskreten Modell eine Reflexionsbedingung für $x=0$ zu formulieren. Drei Möglichkeiten scheinen plausibel, sie sind den folgenden Skizzen zu entnehmen.

Das erste Modell

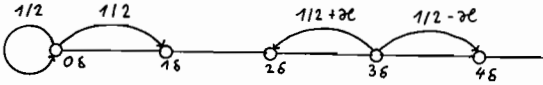


führt zu einer diskreten Verteilung mit

$$\mu = \delta \left(\frac{1}{4\delta\epsilon} - 0,5 \right) \approx \frac{\delta}{4\delta\epsilon} \quad \text{für kleines } \delta\epsilon$$

$$\sigma^2 = \delta^2 \left(\frac{1}{16\delta\epsilon^2} - 0,25 \right) \approx \frac{\delta^2}{16\delta\epsilon^2} \quad \text{für kleines } \delta\epsilon.$$

Das zweite Modell

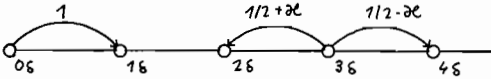


führt zu einer diskreten Verteilung mit

$$\mu = \delta \frac{1}{4\alpha\epsilon} \frac{1+2\alpha}{1+4\alpha} \approx \frac{\delta}{4\alpha\epsilon} \text{ für kleines } \alpha$$

$$\sigma^2 = \delta^2 \frac{1}{16\alpha^2} \left(2 \frac{1+2\alpha}{1-2\alpha} - \left(\frac{1+2\alpha}{1-2\alpha} \right)^2 \right) \approx \frac{\delta^2}{16\alpha^2} \text{ für kleines } \alpha.$$

Das dritte Modell



führt zu einer diskreten Verteilung mit

$$\mu = \delta \frac{1}{4\alpha\epsilon}$$

$$\sigma^2 = \delta^2 \frac{1}{16\alpha^2}.$$

Sind μ und σ vorgegeben, so gilt bei feiner Diskretisierung (δ , also auch α sehr klein) stets:

$$(*) \quad \alpha \approx \frac{\sigma}{4\mu} \quad \text{bzw.} \quad \alpha \approx \frac{\delta}{4\sigma}.$$

Deswegen liefern beim Grenzübergang $\delta \rightarrow 0$ alle drei Modelle die gleiche Differentialgleichung, diesmal

$$\psi_\sigma'(x) = -\frac{1}{\sigma} \psi_\sigma(x)$$

mit der einseitigen Exponentialverteilung

$$\psi_\sigma(x) = \frac{1}{\sigma} e^{-\frac{1}{\sigma}x}$$

als Lösung. Das dritte Modell ist jedoch für diese Herleitung am bequemsten, weil die Näherungsbeziehung (*) exakt gilt.

7 Heuristische Begründung des zentralen Grenzwertsatzes

7.1 Vorbemerkung

Unbestreitbar gehört der zentrale Grenzwertsatz zu den faszinierendsten Bestandteilen der Stochastik. Leider sind die bekannten exakten Beweise schwierig und sie machen nur "indirekt" sichtbar, warum gerade die Normalverteilung als Grenzverteilung auftritt ¹⁾.

In diesem Kapitel soll gezeigt werden, wie man das rekursive Faltungsprinzip zu einer heuristischen Begründung des zentralen Grenzwertsatzes ausbauen kann, die dem Niveau eines Leistungskurses durchaus angemessen ist: Grob gesprochen leiten wir aus dem Faltungsprinzip eine Rekursionsformel für die normierten Wahrscheinlichkeitsfunktionen zu beliebigen Summenexperimenten ab. Sie liefert nach Erhöhung der Summandenzahl eine Bedingung an die Form der Grenzverteilung, welche nicht von der Form der einzelnen Summanden abhängt und wieder als Differentialgleichung $\varphi''(x) + x\varphi'(x) + \varphi(x) = 0$ formuliert werden kann. Die Normalverteilung löst diese Differentialgleichung.

Wie die Herleitung im Spezialfall der Binomialverteilung mit dem Gedankengang aus Kapitel 5 zusammenhängt, entnimmt man der Gegenüberstellung auf der folgenden Seite.

Der Unterschied zu Kapitel 5 besteht im Grunde genommen nur darin, daß dort die Eigenschaft der Binomialverteilung, stationäre Verteilung einer Markoffkette zu sein, hier aber ihre Eigenschaft, Summenverteilung zu sein, für die Aufstellung von Rekursionsformeln herangezogen wird. Diese Rekursionsformeln gehen nach Normierung in verschiedene Differentialgleichungen der Normalverteilung über.

1) Sie ist die einzige Verteilung, die stabil ist und endliche Varianz hat. Die analytische Form der Normalverteilung geht in die Beweise überhaupt nicht ein.

<p><u>Kap. 5:</u> Binomialverteilung als stationäre Verteilung einer Markoffkette</p>	<p><u>Kap. 7:</u> Binomialverteilung als spezielle Summenverteilung</p>
<p>Rekursionsformel, die die Binomialverteilung als stationäre Verteilung der Markoffkette erfüllt:</p> $B_{N,p}(k+1) = \frac{N-k}{k+1} \frac{p}{q} B_{N,p}(k)$	<p>Rekursionsformel, die die Binomialverteilung als Summenverteilung erfüllt:</p> $B_{N+1,p}(k) = pB_{N,p}(k-1) + qB_{N,p}(k)$
<p>Dahinter steht die Rekursionsformel</p> $\binom{N}{k+1} = \frac{N-k}{k+1} \binom{N}{k}$	<p>Dahinter steht die Rekursionsformel</p> $\binom{N+1}{k} = \binom{N}{k-1} + \binom{N}{k}$
<p>Sie geht durch Normierung über in</p> $s_N(x_{k+1}) = \frac{N-k}{k+1} \frac{p}{q} s_N(x_k)$ <p>Mit $s_N(x) \approx \varphi(x)$ und Taylorentwicklung um $x=x_k = \frac{k-Np}{\sqrt{Np}}$ für großes N folgt:</p> $\varphi(x) + \frac{1}{\sqrt{N}} (\varphi'(x) + q\varphi(x)) =$ $\varphi(x) - \frac{1}{\sqrt{N}} p\varphi(x)$ <p>also</p> $\varphi'(x) + x\varphi(x) = 0.$	<p>Sie geht durch Normierung über in</p> $\varphi_{N+1}\left(\frac{k-(N+1)p}{\sqrt{(N+1)p}}\right) = p\varphi_N(x_k) + q\varphi_N(x_{k+1}),$ <p>mit $\varphi_N(x) \approx \varphi_{N+1}(x) \approx \varphi(x)$ und Taylorentwicklung um $x=x_k = \frac{k-Np}{\sqrt{Np}}$ für großes N folgt:</p> $\varphi(x) - \frac{p}{\sqrt{N}} \varphi'(x) - \frac{1}{2N} x \varphi'(x) =$ $\varphi(x) - \frac{p}{\sqrt{N}} \varphi'(x) + \frac{1}{2N} \varphi(x) + \frac{1}{2N} \varphi'(x)$ <p>also</p> $\varphi''(x) + x\varphi'(x) + \varphi(x) = 0 \quad \text{d.h.}$ $(\varphi'(x) + x\varphi(x))' = 0 \quad \text{oder}$ $\varphi'(x) + x\varphi(x) = C.$

Während sich die in der Gegenüberstellung linksstehenden Überlegungen (Kapitel 5) auf andere Markoffketten (vgl. etwa Kapitel 6) übertragen lassen, besitzt der folgende Zugang (rechte Spalte) für beliebige Summenverteilungen Gültigkeit, liefert also den zentralen Grenzwertsatz.

Wir wollen seine heuristische Begründung gleich allgemein, d. h. für Verteilungen durchführen, die von der Binomialverteilung verschieden sind. Dabei soll durch Anschluß an konkrete Beispiele (vgl. 2.4, "Der umgekehrte Wahrscheinlichkeitsbaum") bezweckt werden, daß der Gedankengang von Schülern in beträchtlichem Umfang selbst erarbeitet werden kann.

Der fachdidaktische Wert der heuristischen Herleitung wird erkennbar, wenn man bedenkt, daß der einzige Zugang zu diesem Satz auf Schulniveau (und in einführenden Hochschulvorlesungen) bisher rein "optisch-experimenteller" Natur war ¹⁾ und eine Begründung weder für die Unabhängigkeit der Grenzverteilung von den Ausgangsverteilungen noch ihre analytische Form gegeben wurde. Zunächst sei jedoch an die genaue Formulierung des zentralen Grenzwertsatzes erinnert.

7.2 Der zentrale Grenzwertsatz

Es sei p_i , $i \in \mathbb{N}$ eine Folge von Wahrscheinlichkeitsverteilungen über der reellen Achse mit Erwartungswerten $\mu(p_i) = \mu_i$ und Varianzen $\sigma^2(p_i) = \sigma_i^2$. Die zu den Summenexperimenten der p_i gehörigen Verteilungen $P_N = \sum_{i=1}^N p_i$, $N \in \mathbb{N}$ haben die Varianzen $\sigma_N^2 = \sum_{i=1}^N \sigma_i^2$ und den Erwartungswert $\sum_{i=1}^N \mu_i$. Ihre Standardisierungen, die wir mit $f_N(x) := (x - \sum_{i=1}^N \mu_i) / \sigma_N$ als $f_N \circ (\sum_{i=1}^N p_i)$ bezeichnen wollen, haben den Erwartungswert 0 und die Varianz 1.

1) Vgl. etwa [Ineichen, S. 70 f], [Winter, S. 30 f] oder die Schulbücher der Sekundarstufe II.

Der zentrale Grenzwertsatz besagt, daß bei Erfülltsein der Lindebergbedingung ¹⁾

$$\forall \varepsilon \in \mathbb{R}^+ \quad \frac{1}{\sigma_N^2} \sum_{i=1}^N \int_{|x-\mu_i| > \varepsilon \sigma_N} (x-\mu_i)^2 p_i(dx) \rightarrow 0$$

die Folge der standardisierten Summenverteilungen schwach gegen die Normalverteilung $N(0,1)$ konvergiert. Äquivalent dazu ist die gleichmäßige Konvergenz der zugehörigen Verteilungsfunktionen gegen die Verteilungsfunktion der Normalverteilung.

Unter Benutzung des zentralen Grenzwertsatzes beweist man den lokalen Grenzwertsatz von Gnedenko ²⁾:

Es sei p_i , $i \in \mathbb{N}$ eine Folge identischer Wahrscheinlichkeitsverteilungen über Z mit Schrittweite 1, Erwartungswert μ , Varianz $\sigma^2 \in \mathbb{R}^+$. Dann hat die zum Summenexperiment der p_i gehörige Verteilung $P_N = \prod_{i=1}^N p_i$ den Erwartungswert $N\mu$ und die Standardabweichung $\sigma\sqrt{N}$. Normiert man die zugehörigen Wahrscheinlichkeitsfunktionen

$$P_N: k \mapsto \left(\prod_{i=1}^N p_i \right)(k) = P_N(k)$$

durch Verschiebung um $-N\mu$ sowie Stauchung mit Faktor $\sigma\sqrt{N}$ in Richtung der Längsachse und Streckung mit dem gleichen Faktor in Richtung der Hochachse zu

$$\varphi_N: \left(\frac{k - N\mu}{\sigma\sqrt{N}} \right) \mapsto \sigma\sqrt{N} \cdot P_N(k),$$

so konvergiert die Folge der φ_N gleichmäßig gegen die Dichte der Normalverteilung. In Formeln:

$$\lim_{N \rightarrow \infty} \left| \sigma\sqrt{N} \cdot P_N(k) - \frac{1}{\sqrt{2\pi}} e^{-0,5 \left(\frac{k - N\mu}{\sigma\sqrt{N}} \right)^2} \right| = 0.$$

1) etwa im einfachsten Fall, daß alle p_i miteinander identisch sind und endliche Varianz haben.

2) [Fisz, S. 252].

Der lokale Grenzwertsatz von de Moivre-Laplace ist mit $P_N(k) := B_{N,p}(k)$ ein Spezialfall hiervon.

Bemerkungen zum Beweis des zentralen Grenzwertsatzes:

In der Literatur findet man im wesentlichen zwei Wege, den ursprünglichen Beweis von [Lindeberg (1921)], dem [Trotter (1959)] eine moderne und von der Struktur her durchsichtige Form gegeben hat ("Operatorenmethode") und den Beweis über Fouriertransformationen, der auf [Levy (1925)] zurückgeht.

Zur Operatorenmethode:

Jedem Wahrscheinlichkeitsmaß q wird mittels Faltung ein linearer Operator $\bar{q} : C^0(\mathbb{R}) \rightarrow C^0(\mathbb{R})$, $f \mapsto q * f$ zugeordnet, wobei $(q * f)(x) := \int_{\mathbb{R}} f(x-t) q(dt)$ gilt ¹⁾. Die schwache Konvergenz von Wahrscheinlichkeitsmaßen ist dann äquivalent zur punktweisen Konvergenz der zugehörigen Operatoren. Zum Beweis der im zentralen Grenzwertsatz behaupteten Konvergenzaussage benutzt man wesentlich die Eigenschaft der Operatoren

$$(1) \quad \left\| \overline{q_1 * q_2 * \dots * q_n}(f) - \overline{r_1 * r_2 * \dots * r_n}(f) \right\|_{\infty} \leq \left\| \overline{q_1}(f) - \overline{r_1}(f) \right\|_{\infty} + \left\| \overline{q_2}(f) - \overline{r_2}(f) \right\|_{\infty} + \dots + \left\| \overline{q_n}(f) - \overline{r_n}(f) \right\|_{\infty},$$

die Zerlegbarkeit der Normalverteilung:

$$(2) \quad N(0,1) = N(0, \sigma_1^2) * N(0, \sigma_2^2) * \dots * N(0, \sigma_n^2), \quad \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 = 1$$

und die Tatsache, daß für $f \in C^2(\mathbb{R})$ $\left\| \overline{q_i}(f) - \overline{r_i}(f) \right\|_{\infty}$ sehr klein wird, falls q_i und r_i ²⁾ sehr kleine Varianz haben. Zur Rolle, die die Normalverteilung als Grenzverteilung spielt, schreibt Trotter: "The proof has the interesting feature that the only facts about the normal distribution used in the argument are that it has finite variance, and that the sum of independent normal varieties is again normal. In

1) $C^n(\mathbb{R})$ bezeichnet den Raum der Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ mit beschränkten und gleichmäßig stetigen Ableitungen bis zur Ordnung n , $\|f\|_n := \sup_{x \in \mathbb{R}} |f(x)|$.

particular , the explicit analytic form of the distribution plays no part." [Trotter, S. 226].

Beim Beweis über Fouriertransformationen

ordnet man jedem Wahrscheinlichkeitsmaß q seine Fouriertransformierte \hat{q} mit $\hat{q}(x) = \int_{\mathbb{R}} e^{ixy} q(dy)$ zu und führt die schwache Konvergenz von Wahrscheinlichkeitsmaßen zurück auf die punktweise bzw. über kompakten Intervallen gleichmäßige Konvergenz der Fouriertransformierten, für die statt der Operatoreigenschaft (1) die Beziehung

$$(3) \quad \widehat{q * r} = \hat{q} \cdot \hat{r}$$

gilt. Zum Konvergenzbeweis führt man dann Taylorentwicklungen der Fouriertransformierten durch, benutzt die Tatsache, daß Maße mit Momenten bis zur Ordnung k k -mal stetig differenzierbare Fouriertransformierte haben - und natürlich wieder die Zerlegbarkeit (2) der Normalverteilung.

7.3 Herleitung der Differentialgleichung für die Grenzverteilung im Spezialfall gleicher Summanden

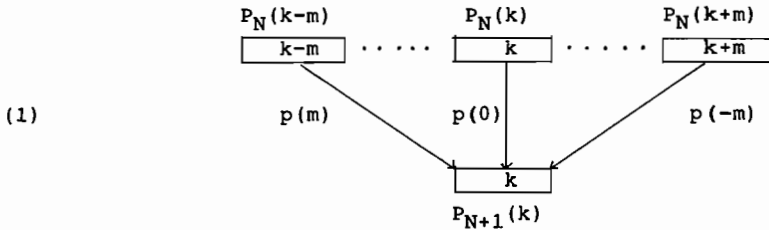
Es soll nun das Auftreten der Normalverteilung φ als Dichte der Grenzverteilung im zentralen Grenzwertsatz (bzw. als Grenzwert der normierten Wahrscheinlichkeitsfunktionen im lokalen Grenzwertsatz von Gnedenko) heuristisch begründet werden. Dabei betrachten wir in diesem Abschnitt nur diskrete beschränkte Summandenverteilungen

$$p_i: \{-m, \dots, m\} \rightarrow [0; 1],$$

die voneinander nicht verschieden sind ($p_i = p$ für alle $i \in \mathbb{N}$).

2) Die Rolle der q_i wird übernommen durch $N(0, \sigma_i^2 / \mathcal{T}_N)$, die Rolle der r_i wird übernommen durch die Verteilungen $(1/\mathcal{T}_N) \circ p_i$.

Gilt $\sigma(p) = \sigma > 0$ und o.B.d.A. $\mu(p) = 0$, so hat man für die Wahrscheinlichkeitsverteilung $P_N = \sum_{i=1}^N p$ des zugehörigen N-fachen Summenexperimentes $\mu(P_N) = 0$ und $\sigma(P_N) = \sigma\sqrt{N}$. Mit dem umgekehrten Wahrscheinlichkeitsbaum



erhält man die Rekursionsformel

(2)
$$P_{N+1}(k) = \sum_{i=-m}^m p(i) P_N(k-i).$$

Für die zugehörigen normierten Wahrscheinlichkeitsfunktionen φ_N , definiert durch

$$\varphi_N\left(\frac{k}{\sigma\sqrt{N}}\right) = \sigma\sqrt{N} P_N(k) \quad \text{bzw.} \quad P_N(k) = \frac{1}{\sigma\sqrt{N}} \varphi_N\left(\frac{k}{\sigma\sqrt{N}}\right)$$

folgt daraus die Rekursionsformel

$$\frac{1}{\sigma\sqrt{N+1}} \varphi_{N+1}\left(\frac{k}{\sigma\sqrt{N+1}}\right) = \sum_{i=-m}^m p(i) \frac{1}{\sigma\sqrt{N}} \varphi_N\left(\frac{k-i}{\sigma\sqrt{N}}\right), \text{ also}$$

(3)
$$\varphi_{N+1}\left(\frac{k}{\sigma\sqrt{N+1}}\right) = \sqrt{\frac{N+1}{N}} \sum_{i=-m}^m p(i) \varphi_N\left(\frac{k-i}{\sigma\sqrt{N}}\right).$$

Zur Weiterentwicklung veranschaulichen wir uns den bisherigen Gedankengang am Beispiel des Glücksrades aus 3.6:

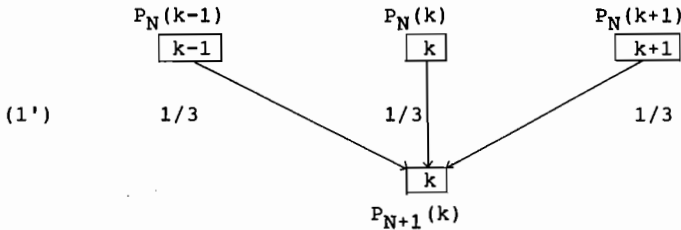


$$p(0) = p(1) = p(-1) = 1/3$$

$$\mu = 0 \quad \sigma = \sqrt{2/3}$$

Es wird N mal hintereinander gedreht, wir summieren die Ergebnisse und fragen nach der Wahrscheinlichkeit $P_N(k) = \sum_{i=1}^N p(k)$, daß sich dabei insgesamt k Punkte ergeben. Die

rekursive Berechnung mit dem umgekehrten Wahrscheinlichkeitsbaum



(2')

$$P_{N+1}(k) = \frac{P_N(k-1) + P_N(k) + P_N(k+1)}{3}$$

liefert die Tabelle auf der folgenden Seite, deren Anfang schon in 3.6 durch den Faltungsabakus enaktiv bestimmt worden war. Die zugehörige Skizze auf der übernächsten Seite zeigt die zu P_9 und P_{10} gehörigen normierten Wahrscheinlichkeitspolygone φ_9 und φ_{10} mit Stützstellen bei $x_k = \frac{k}{\sqrt{48/3}}$ bzw. $x_k = \frac{k}{\sqrt{20/3}}$. Die Rekursionsformel (3), die für unser Glücksrad die Form

(3')

$$\varphi_{N+1}\left(\frac{k}{\sqrt{20/3}}\right) = \sqrt{\frac{N+1}{N}} \frac{1}{3} \left[\varphi_N\left(\frac{k-1}{\sqrt{N}}\right) + \varphi_N\left(\frac{k}{\sqrt{N}}\right) + \varphi_N\left(\frac{k+1}{\sqrt{N}}\right) \right]$$

hat, liefert für $N=9$ und $k=4$

(3'')

$$\varphi_{10}\left(\frac{4}{\sqrt{10}}\right) = \sqrt{\frac{10}{9}} \frac{1}{3} \left[\varphi_9\left(\frac{3}{\sqrt{9}}\right) + \varphi_9\left(\frac{4}{\sqrt{9}}\right) + \varphi_9\left(\frac{5}{\sqrt{9}}\right) \right]$$

Ihre Interpretation wurde in der Skizze durch Pfeile angedeutet, die den umgekehrten Wahrscheinlichkeitsbaum wiedererkennen lassen: Man bildet das (allgemein durch p gewichtete) arithmetische Mittel aus 3 $(2m+1)$ nebeneinander liegenden Stützstäben zu φ_9 (φ_N), verlängert den entstehenden Stab um $\sqrt{10/9}$ ($\sqrt{(N+1)/N}$), schiebt ihn von der Stelle $\frac{4}{\sqrt{9}}$ ($\frac{k}{\sqrt{N}}$) auf die Stelle $\frac{4}{\sqrt{10}}$ ($\frac{k}{\sqrt{N+1}}$) nach links und erhält den zugehörigen Stützstab von φ_{10} (φ_{N+1}). Er liegt wie alle anderen Stützstäbe sehr nahe am Polygonzug φ_9 (φ_N).

Wenn die φ_N gegen eine Grenzfunktion φ konvergieren (und

Rekursive Berechnung der Verteilung $P_{10}^* = \sum_{i=1}^{10} p_i$ als zehnfache Faltung

	Nenner	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	
$p_1 = p$	3																						
$*p$																							
$p_2 = p_1 * p$	9																						
$*p$																							
$p_3 = p_2 * p$	27																						
$*p$																							
$p_4 = p_3 * p$	81																						
$*p$																							
$p_5 = p_4 * p$	243																						
$*p$																							
$p_6 = p_5 * p$	729																						
$*p$																							
$p_7 = p_6 * p$	2187																						
$*p$																							
$p_8 = p_7 * p$	6561																						
$*p$																							
$p_9 = p_8 * p$	19683																						
$*p$																							
$p_{10} = p_9 * p$	59049	1	10	55	210	615	1452	2850	4740	6765	8350	8953	8350	6765	4740	2850	1452	615	210	55	10	1	
$p_{10} = * p$.000	.000	.001	.004	.010	.025	.048	.080	.115	.141	.152	.141	.115	.080	.048	.025	.010	.004	.001	.000	.000	
$\sqrt[4]{0,120/3}$.000	.000	.001	.004	.010	.024	.047	.079	.114	.143	.155	.143	.114	.079	.047	.024	.010	.004	.001	.000	.000	

Zum zentralen Grenzwertsatz für identische Summanden

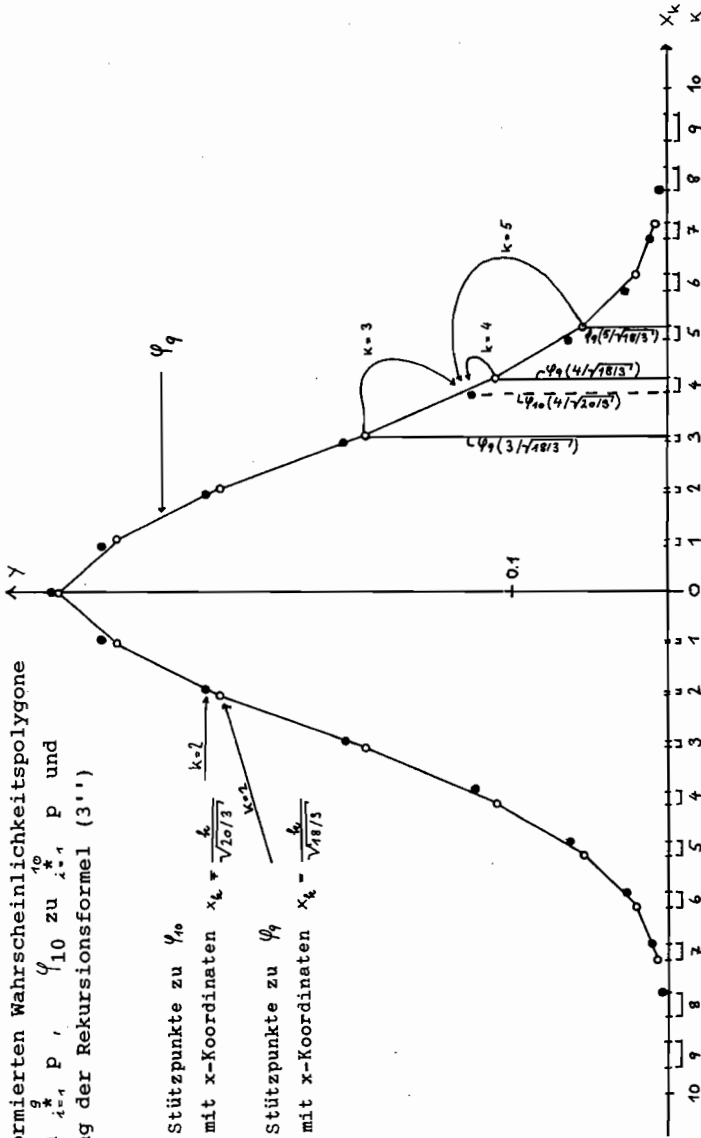
Die normierten Wahrscheinlichkeitspolygone φ_9 zu ξ_1, \dots, ξ_{10} zu $\xi_1^*, \dots, \xi_{10}^*$, p und Deutung der Rekursionsformel (3'')

● : Stützpunkte zu φ_{10}

mit x-Koordinaten $x_k = \frac{k}{\sqrt{20/3}}$, $k=2$

○ : Stützpunkte zu φ_9

mit x-Koordinaten $x_k = \frac{k}{\sqrt{10/3}}$



das setzen wir voraus), dann muß auch sie eine Form haben, für die das entsprechend "verlängerte" und "verschobene" arithmetische Mittel aus Funktionswerten von φ wieder ein Funktionswert von φ ist. Nun liegen bei hohem N alle Stellen, die zur Bildung des Mittelwertes benötigt werden, sehr nahe beieinander, und wir können ihn auch durch eine Taylorentwicklung von φ um die Stelle $x = \frac{k}{6\sqrt{N}}$ berechnen. In unserem Spezialfall ergibt sich auf der rechten Seite von (3'):

$$\begin{aligned} & \frac{1}{3} \left[\varphi\left(\frac{k-1}{6\sqrt{N}}\right) + \varphi\left(\frac{k}{6\sqrt{N}}\right) + \varphi\left(\frac{k+1}{6\sqrt{N}}\right) \right] = \\ & \frac{1}{3} \left[\varphi\left(x - \frac{1}{6\sqrt{N}}\right) + \varphi(x) + \varphi\left(x + \frac{1}{6\sqrt{N}}\right) \right] \approx \\ & \frac{1}{3} \left[\varphi(x) - \varphi'(x) \frac{1}{6\sqrt{N}} + \varphi''(x) \frac{1}{2} \left(\frac{1}{6\sqrt{N}}\right)^2 + \varphi(x) + \varphi'(x) + \varphi''(x) \frac{1}{6\sqrt{N}} + \varphi''(x) \frac{1}{2} \left(\frac{1}{6\sqrt{N}}\right)^2 \right] = \end{aligned}$$

$$(4) \quad \varphi(x) + \frac{1}{3} \frac{1}{6\sqrt{N}} \varphi''(x) = \varphi(x) + \frac{1}{2N} \varphi''(x).$$

Wegen $\sqrt{\frac{N+1}{N}} = \sqrt{1 + \frac{1}{N}} \approx 1 + \frac{1}{2N}$ erhalten wir hiermit für die rechte Seite von (3'), also das "verlängerte arithmetische Mittel":

$$(5) \quad \left(1 + \frac{1}{2N}\right) \left(\varphi(x) + \frac{1}{2N} \varphi''(x)\right) \approx \varphi(x) + \frac{1}{2N} \varphi(x) + \frac{1}{2N} \varphi''(x).$$

Die linke Seite von (3'), also der Funktionswert von φ an der "verschobenen Stelle" $\frac{k}{6\sqrt{N+1}} = \frac{k}{6\sqrt{N}} - \delta = x - \delta$ ergibt sich wegen $\frac{k}{6\sqrt{N+1}} \approx \frac{k}{6\sqrt{N}} - \frac{k}{6\sqrt{N}} \frac{1}{2N} = x - x \frac{1}{2N}$ zu

$$(6) \quad \varphi\left(\frac{k}{6\sqrt{N+1}}\right) \approx \varphi(x) - \frac{1}{2N} x \varphi'(x)$$

und als Bedingung an die Grenzfunktion erhält man durch Vergleich von (5) und (6) die Differentialgleichung

$$(7) \quad \varphi(x) + x \varphi'(x) + \varphi''(x) = 0.$$

Da die linke Seite dieser Differentialgleichung eine Ableitung darstellt, können wir sie durch Integration sofort in die lineare Differentialgleichung

$$(8) \quad x \varphi(x) + \varphi'(x) = C, \quad C \in \mathbb{R}$$

erster Ordnung überführen. Weil φ symmetrisch zur y -Achse verlaufen, also $\varphi'(0)=0$ erfüllen wird, folgt durch Einsetzen in (8)

$$C = 0 \cdot \varphi(0) + \varphi'(0) = 0$$

und (8) vereinfacht sich zu $\varphi'(x)/\varphi(x) = -x$. Wie in Kapitel 5 erhalten wir damit durch Integration $\ln(\varphi(x)) = -x^2/2 + k$ und $\varphi(x) = c e^{-\frac{1}{2}x^2}$.

Das Faszinierende an dieser Herleitung - und gleichzeitig das heuristische Argument für die Universalität der normalen Grenzverteilung - besteht darin, daß sie sich auf beliebige Verteilungen p übertragen läßt. Die Mittelwertbildung (auf der rechten Seite von (3)) liefert nämlich mit

$$\begin{aligned} \sum_{i=-m}^m p(i) \cdot \varphi\left(\frac{k-i}{\sigma\sqrt{N}}\right) &= \sum_{i=-m}^m p(i) \cdot \varphi\left(x - \frac{i}{\sigma\sqrt{N}}\right) \approx \\ \varphi(x) \cdot \underbrace{\sum_{i=-m}^m p(i)}_1 - \frac{1}{\sigma\sqrt{N}} \varphi'(x) \underbrace{\sum_{i=-m}^m p(i) \cdot i}_0 + \frac{1}{2\sigma^2 N} \varphi''(x) \underbrace{\sum_{i=-m}^m p(i) \cdot i^2}_{\sigma^2} &= \end{aligned}$$

$$(4') \quad \varphi(x) + \frac{1}{2N} \varphi''(x)$$

ein Ergebnis, das wegen $\sum_{i=-m}^m p(i) = 1$, $\sum_{i=-m}^m p(i) \cdot i = 0$, $\sum_{i=-m}^m p(i) \cdot i^2 = \sigma^2$ von der speziellen Form der Summandenverteilung p unabhängig ist und mit demjenigen für unser Glücksrad (4) übereinstimmt. (5) und (6) liefern dann auch für allgemeines p die Differentialgleichung $\varphi(x) + x\varphi'(x) + \varphi''(x) = 0$ als notwendige Bedingung an die Form der Grenzverteilung.

7.4 Herleitung der Differentialgleichung im allgemeinen Fall nicht identischer Summanden

Die obige Begründung des zentralen Grenzwertsatzes ist bei Veranschaulichung durch den Spezialfall des Glücksrades in Leistungskursen ohne Schwierigkeiten problemorientiert zu erarbeiten. Damit dürfte man allen Intentionen, die mit einer Behandlung der Normalverteilung als einer universellen Grenzverteilung in der Schule angestrebt werden können, gerecht werden.

In diesem Abschnitt soll ergänzend dargelegt werden, wie sich der Gedankengang auf nicht identische, nicht notwendig beschränkte "Summandenverteilungen" übertragen läßt. Dabei wird auch die Stellung der Lindebergbedingung einsichtig.

Es sei also p_i , $i \in \mathbb{N}$ eine Folge von Wahrscheinlichkeitsverteilungen über \mathbb{Z} mit dem Erwartungswert $\mu(p_i) = \mu_i$ und der Varianz $\sigma^2(p_i) = \sigma_i^2$. Die zugehörigen Summenverteilungen $P_N = \prod_{i=1}^N p_i$ haben den Erwartungswert $\nu_N = \sum_{i=1}^N \mu_i$ und die Varianz $\tau_N^2 = \sum_{i=1}^N \sigma_i^2$. Aus der Rekursionsformel (2) von 7.3

$$(1) \quad P_{N+1}(k) = \sum_{i \in \mathbb{Z}} p_{N+1}(i) P_N(k-i)$$

folgt jetzt für die zugehörigen normierten Wahrscheinlichkeitsfunktionen φ_N definiert durch

$$\varphi_N\left(\frac{k - \nu_N}{\tau_N}\right) = \tau_N P_N(k) \quad \text{bzw.} \quad P_N(k) = \frac{1}{\tau_N} \varphi_N\left(\frac{k - \nu_N}{\tau_N}\right)$$

die Rekursionsformel

$$(2) \quad \varphi_{N+1}\left(\frac{k - \nu_{N+1}}{\tau_{N+1}}\right) = \frac{\tau_{N+1}}{\tau_N} \sum_{i \in \mathbb{Z}} p_{N+1}(i) \varphi_N\left(\frac{k - \nu_N - i}{\tau_N}\right)$$

bzw. mit $x = \frac{k - \nu_{N+1}}{\tau_{N+1}}$

$$(3) \quad \frac{\tau_N}{\tau_{N+1}} \varphi_{N+1}(x) = \sum_{i \in \mathbb{Z}} p_{N+1}(i) \varphi_N\left[\left[x - \frac{i - \mu_{N+1}}{\tau_{N+1}}\right] \frac{\tau_{N+1}}{\tau_N}\right].$$

Die Skizze auf der folgenden Seite zeigt die anschauliche Deutung von (2) für $N=7$, $k=9$, $x = \frac{g-\mu_8}{\sigma_8}$ anhand des Beispiels aus 2.4. (Mit den dort eingeführten Bezeichnungen gilt $P_7 = p * q * r * s * p * q * r$, $P_8 = p * q * r * s * p * q * r * s$ und $p_8 = s$.)

Wir nehmen in (3) mit $\varphi_N \approx \varphi_{N+1} \approx \varphi$ Taylorentwicklungen um die Stelle $x = \frac{k-\mu_{N+1}}{\sigma_{N+1}}$ vor und zwar bis zur ersten Ordnung in $\delta = \frac{\sigma_{N+1}^2}{\sigma_N^2}$ (also bis zur zweiten Ordnung in $\frac{i-\mu_{N+1}}{\sigma_{N+1}} \sim \sqrt{\delta}$). Mit

$$\frac{\sigma_{N+1}}{\sigma_N} = \sqrt{\frac{\sigma_{N+1}^2}{\sigma_{N+1}^2 - \sigma_{N+1}^2}} = \sqrt{\frac{1}{1 - \frac{\sigma_{N+1}^2}{\sigma_{N+1}^2}}} \approx 1 + \frac{1}{2} \frac{\sigma_{N+1}^2}{\sigma_{N+1}^2}$$

erhalten wir

$$\begin{aligned} \varphi\left(\left[x - \frac{i-\mu_{N+1}}{\sigma_{N+1}}\right] \frac{\sigma_{N+1}}{\sigma_N}\right) &\approx \varphi\left(x + x \frac{1}{2} \frac{\sigma_{N+1}^2}{\sigma_{N+1}^2} - \frac{i-\mu_{N+1}}{\sigma_{N+1}}\right) \approx \\ \varphi(x) + x \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi'(x) - \frac{i-\mu_{N+1}}{\sigma_{N+1}} \varphi(x) + \frac{1}{2} \left(\frac{i-\mu_{N+1}}{\sigma_{N+1}}\right)^2 \varphi''(x). \end{aligned}$$

Durch Summation folgt für die rechte Seite von (3)

$$\begin{aligned} (4) \quad \sum_{i \in \mathbb{Z}} P_{N+1}(i) \varphi\left(\left[x - \frac{i-\mu_{N+1}}{\sigma_{N+1}}\right] \frac{\sigma_{N+1}}{\sigma_N}\right) &\approx \\ \left[\varphi(x) + x \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi'(x)\right] \underbrace{\sum_{i \in \mathbb{Z}} P_{N+1}(i)}_1 - \frac{1}{\sigma_{N+1}} \varphi'(x) \underbrace{\sum_{i \in \mathbb{Z}} P_{N+1}(i) (i-\mu_{N+1})}_0 + \\ &+ \frac{1}{2\sigma_{N+1}^2} \varphi''(x) \underbrace{\sum_{i \in \mathbb{Z}} P_{N+1}(i) (i-\mu_{N+1})^2}_{\sigma_{N+1}^2} = \end{aligned}$$

$$(5) \quad \varphi(x) + \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} x \varphi'(x) + \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi''(x).$$

Für die linke Seite von (3) erhalten wir mit

$$\frac{\sigma_N}{\sigma_{N+1}} = \sqrt{\frac{\sigma_{N+1}^2 - \sigma_{N+1}^2}{\sigma_{N+1}^2}} = \sqrt{1 - \frac{\sigma_{N+1}^2}{\sigma_{N+1}^2}} \approx 1 - \frac{1}{2} \frac{\sigma_{N+1}^2}{\sigma_{N+1}^2}$$

$$(6) \quad \frac{\sigma_N}{\sigma_{N+1}} \varphi(x) \approx \varphi(x) - \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi(x)$$

Ein Vergleich von (5) und (6) liefert auch hier wieder die Differentialgleichung $\varphi(x) + x \varphi'(x) + \varphi''(x) = 0$ als notwendige Bedingung an die Grenzverteilung.

Zum zentralen Grenzwertsatz für nicht identische Summanden

Die normierten Wahrscheinlichkeitspolygone

φ_7 zu $P_7 = p_1 q_1 r_1 s_1 p_2 q_2 r_2 s_2 p_3 q_3 r_3 s_3 p_4 q_4 r_4 s_4 p_5 q_5 r_5 s_5 p_6 q_6 r_6 s_6 p_7 q_7 r_7 s_7$ mit $\mu_7 = 4.0$, $\sigma_7 = 2.11$,

φ_8 zu $P_8 = p_1 q_1 r_1 s_1 p_2 q_2 r_2 s_2 p_3 q_3 r_3 s_3 p_4 q_4 r_4 s_4 p_5 q_5 r_5 s_5 p_6 q_6 r_6 s_6 p_7 q_7 r_7 s_7 p_8 q_8 r_8 s_8$ mit $\mu_8 = 5.5$, $\sigma_8 = 2.39$

(dabei ist $p_8 = s$ mit $s(0) = s(1) = s(2) = s(3) = 1/4$)

und geometrische Deutung der Rekursionsformel (2)

für $N=7$, $k=9$

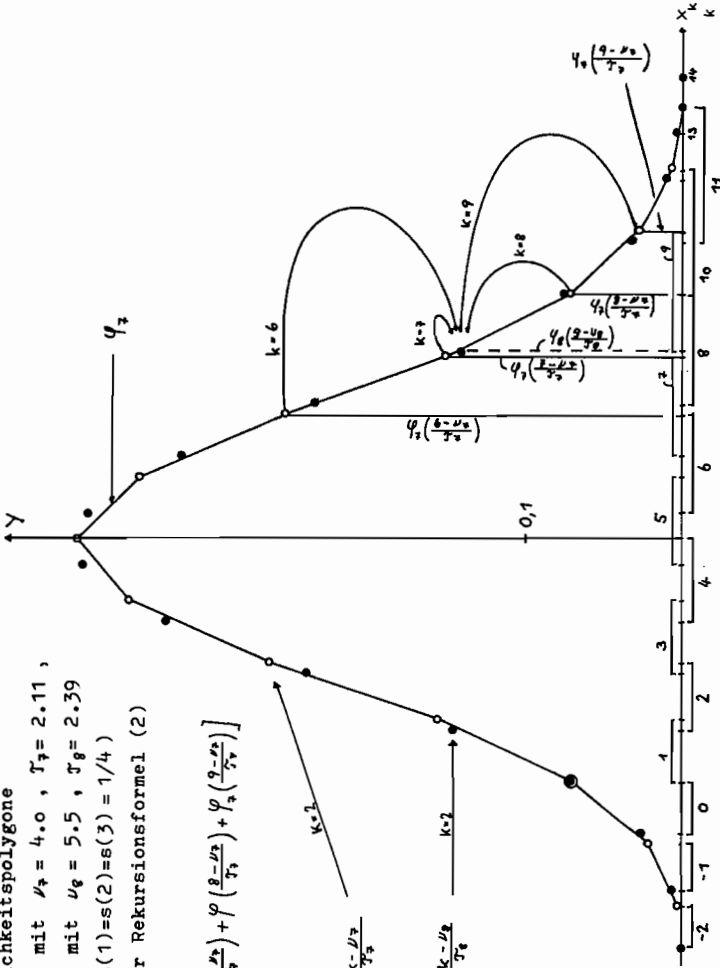
$$\varphi_8 \left(\frac{x - \mu_8}{\sigma_8} \right) = \frac{\sigma_7}{\sigma_8} \frac{1}{4} \left[\varphi_7 \left(\frac{x - \mu_7}{\sigma_7} \right) + \varphi_7 \left(\frac{x - \mu_7}{\sigma_7} \right) + \varphi_7 \left(\frac{x - \mu_7}{\sigma_7} \right) + \varphi_7 \left(\frac{x - \mu_7}{\sigma_7} \right) \right]$$

○ : Stützpunkte zu φ_7

mit x-Koordinaten $x_k = \frac{k - \mu_7}{\sigma_7}$

● : Stützpunkte zu φ_8

mit x-Koordinaten $x_k = \frac{k - \mu_8}{\sigma_8}$



7.5 Die Lindebergbedingung ¹⁾

Die Herleitung 7.4 funktioniert formal für alle Folgen p_i , $i \in \mathbb{N}$, doch sind die Taylorentwicklungen nur sinnvoll, wenn

$$(7) \quad \frac{i - \mu_N}{\sigma_N} \quad \text{und} \quad \frac{\sigma_N}{\sigma_N} = 1$$

sehr klein sind. Da i jedoch alle ganzen Zahlen durchläuft, müssen wir durch eine Voraussetzung dafür sorgen, daß in (3) die Summe aller Wahrscheinlichkeiten $P_{N+1}(i)$ für solche i , die große Abweichungen $(i - \mu_{N+1})/\sigma_{N+1}$ liefern, mit wachsendem N immer kleiner wird, so daß diese Summanden von vornherein zu vernachlässigen sind. Wir setzen also voraus:

$$(8) \quad \forall \varepsilon > 0 \quad \sum_{\left| \frac{i - \mu_N}{\sigma_N} \right| > \varepsilon} P_N(i) \rightarrow 0 \quad \text{für } N \rightarrow \infty.$$

Damit sich trotz dieser Vernachlässigung (beim Übergang von (4) zu (5)) Erwartungswert und Varianz von p_N nicht verändern, müssen wir zusätzlich fordern

$$(9) \quad \forall \varepsilon > 0 \quad \sum_{\left| \frac{i - \mu_N}{\sigma_N} \right| > \varepsilon} P_N(i) (i - \mu_N) \rightarrow 0 \quad \text{für } N \rightarrow \infty$$

und

$$(10) \quad \forall \varepsilon > 0 \quad \sum_{\left| \frac{i - \mu_N}{\sigma_N} \right| > \varepsilon} P_N(i) (i - \mu_N)^2 \rightarrow 0 \quad \text{für } N \rightarrow \infty.$$

Wir stellen fest, daß (10) (bis auf zusätzliche Summation über N , die zum Beweis der Konvergenz von φ_N benötigt wird) mit der Lindebergbedingung übereinstimmt. Da zwischen den genannten Forderungen die Beziehungen $(7) \Leftarrow (10) \Rightarrow (9) \Rightarrow (8)$ ²⁾ bestehen, genügt es, für die heuristische Begründung alleine

1) Die Numerierung der Formeln schließt an 7.4 an.

2) (7) \Leftarrow (10) zeigt man wie folgt:

$$\frac{\sigma_N^2}{\sigma_N^2} = \sum_{i \in \mathbb{Z}} \left(\frac{i - \mu_N}{\sigma_N} \right)^2 p_N(i) \leq \underbrace{\sum_{\left| \frac{i - \mu_N}{\sigma_N} \right| \leq \varepsilon} \left(\frac{i - \mu_N}{\sigma_N} \right)^2 p_N(i)}_{\leq \varepsilon^2} + \underbrace{\sum_{\left| \frac{i - \mu_N}{\sigma_N} \right| > \varepsilon} \left(\frac{i - \mu_N}{\sigma_N} \right)^2 p_N(i)}_{\rightarrow 0 \text{ wegen (10)}}$$

die "Lindeberg-ähnliche" Bedingung (10) vorauszusetzen. Der Vollständigkeit halber seien die Stellen im einzelnen angegeben, an denen (10) bzw. (7)-(10) zur Entwicklung der rechten Seite von (3) benötigt werden.

$$\sum_{i \in \mathbb{Z}} P_{N+1}(i) \varphi \left(\left[x - \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right] \frac{\sigma_{N+1}}{\sigma_N} \right) =$$

$$\sum_{\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| \leq \varepsilon} P_{N+1}(i) \varphi \left(\left[x - \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right] \frac{\sigma_{N+1}}{\sigma_N} \right) + \sum_{\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| > \varepsilon} P_{N+1}(i) \varphi \left(\left[x - \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right] \frac{\sigma_{N+1}}{\sigma_N} \right) \approx$$

hier ist wegen ≈ 0 nach (8) $< k$, da φ
 $\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| < \varepsilon$ und (7) beschränkt
 die Taylorentwicklung sinnvoll

$$\approx \left[\varphi(x) + x \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi'(x) \right] \sum_{\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| \leq \varepsilon} P_{N+1}(i) -$$

≈ 1 wegen (8)

$$- \varphi'(x) \frac{1}{\sigma_{N+1}} \sum_{\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| \leq \varepsilon} P_{N+1}(i) (i - \mu_{N+1}) + \varphi''(x) \frac{1}{2\sigma_{N+1}^2} \sum_{\left| \frac{i - \mu_{N+1}}{\sigma_{N+1}} \right| \leq \varepsilon} P_{N+1}(i) (i - \mu_{N+1})^2 \approx$$

≈ 0 wegen (9) $\approx \sigma_{N+1}^2$ wegen (10)

$$\approx \varphi(x) + \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} x \varphi'(x) + \frac{\sigma_{N+1}^2}{2\sigma_{N+1}^2} \varphi''(x).$$

In die Taylorentwicklung auf der linken Seite von (3) geht nur (6) ein.

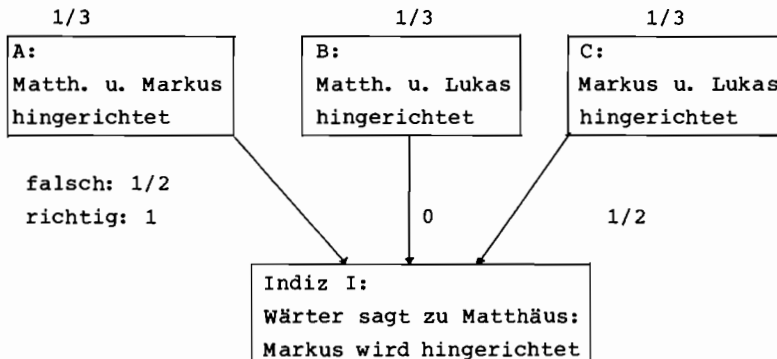
ANHANG

Paradoxien zur Bayesschen Regel

Im folgenden sei an vier bekannte Paradoxien zur Bayesschen Regel erinnert. Sie beruhen darauf, daß man intuitiv entweder a priori Wahrscheinlichkeiten oder die bedingten Wahrscheinlichkeiten falsch einschätzt. Mit dem umgekehrten Baumdiagramm lassen sich diese Fehler prägnant herausarbeiten.

1. Prisoners Paradoxon

[Beckenbach, S. 47], [Engel 1, S. 146]. Wir übernehmen die Formulierung von Engel: Matthäus, Markus und Lukas sind zum Tode verurteilt. Einer der drei wird ausgelost und begnadigt. Der Name des Glücklichen wird streng geheimgehalten. Matthäus sagt sich: Die Wahrscheinlichkeit, daß ich es bin, ist $1/3$. Er sagt dem Wärter: Einer der beiden, Markus oder Lukas, wird sicher hingerichtet. Du wirst mir also nichts verraten, wenn Du mir einen Mann nennst, Markus oder Lukas, der hingerichtet wird. Darauf sagt der Wärter: Markus wird hingerichtet. Diese Antwort hat Matthäus ermutigt. Denn er oder Lukas wurden sicher begnadigt. Daher ist die Wahrscheinlichkeit $1/2$, daß er es ist. Hat Matthäus recht? Der umgekehrte Wahrscheinlichkeitsbaum sieht wie folgt aus:



Es liefert für Matthäus die revidierten Wahrscheinlichkeitsverteilungen:

	A	B	C
Pfalsch	$1/2$	0	$1/2$
Prichtig	$2/3$	0	$1/3$

Matthäus hat nicht recht. Die Wahrscheinlichkeit, daß er hingerichtet wird, ist trotz der Information des Wärters $2/3$ geblieben. Das liegt daran, daß die Alternativen A, B, C das beobachtete Indiz I nicht mit der gleichen Wahrscheinlichkeit nach sich ziehen. Falls nämlich Matthäus und Markus hingerichtet werden, muß der Wärter dem Matthäus den Namen Markus mit Wahrscheinlichkeit 1 nennen. Werden dagegen Markus und Lukas hingerichtet, nennt er den Namen Markus mit Wahrscheinlichkeit $1/2$, er hat jetzt ja Wahlmöglichkeit. Somit ist die Wahrscheinlichkeit, daß der Wärter den Namen Markus nennt, weil Matthäus (er selbst) und Markus sterben, doppelt so groß wie die Wahrscheinlichkeit, daß er ihn nennt, weil Markus und Lukas sterben. Die Tatsache, daß die Information an Matthäus geliefert wird, ist wesentlich.

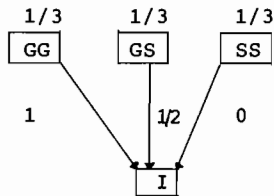
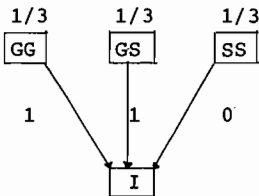
2. Bertrands Schubfächerparadoxon

Gegeben sind drei Kästen mit je zwei Schubladen. In jeder Schublade liegt eine Münze; im ersten Kasten (GG) enthalten beide Schubladen eine Goldmünze, im zweiten Kasten (GS) enthält eine Schublade eine Goldmünze, die andere eine Silbermünze, im dritten Kasten (SS) enthalten beide Schubladen Silbermünzen. Ich wähle zufällig einen Kasten aus, ziehe zufällig eine Schublade, sehe eine goldene Münze und stelle mir jetzt die Frage: Wie groß ist die Wahrscheinlichkeit, daß auch in der zweiten Schublade des gewählten Kastens eine goldene Münze steckt, daß ich also den Kasten GG gewählt hatte? Die Primärintuition diktiert meist die falsche Antwort $1/2$ (denn nach dem Öffnen der Schublade ist der Kasten SS ausgeschlossen, es kann sich also nur noch um GS oder GG handeln). Der Fehler liegt wieder darin, daß wir auch hier übersehen, daß GG und GS mit verschiedenen Wahrscheinlich-

keiten zum beobachteten Indiz führen, GS mit der Wahrscheinlichkeit $1/2$, GG mit der doppelten Wahrscheinlichkeit 1 ¹⁾.

intuitiv akzeptiertes Diagramm

richtiges Diagramm



revidierte Wahrscheinlichkeitsverteilungen:

GG	GS	SS
$1/2$	$1/2$	0

GG	GS	SS
$2/3$	$1/3$	0

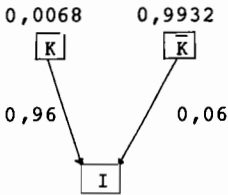
3. "Krebs-Paradoxon"

Das folgende Paradoxon tritt in vielen Varianten auf, wir zitieren [Engel 1, S. 142]: Angenommen, es gibt einen sehr zuverlässigen Test zur Krebsdiagnose: Habe ich Krebs (K), ist der Test positiv (Indiz I) mit Wahrscheinlichkeit 96%, habe ich keinen Krebs (\bar{K}), dann ist der Test nur mit 6%iger Wahrscheinlichkeit positiv. Ich unterziehe mich dem Test, und er ist positiv. Wie groß ist die Wahrscheinlichkeit, daß ich Krebs habe, wenn $1/145=0,68\%$ aller Personen meines Alters Krebs haben ohne es zu wissen a) größer als 70% b) zwischen 30% und 70% c) kleiner als 30%? Sehr wenige Versuchspersonen kreuzen die richtige Lösung c) an. Das liegt daran, daß wir uns intuitiv die geringe a priori Wahrscheinlichkeit für Krebs nicht zueigen machen. Da wir uns normalerweise erst dann einer Untersuchung

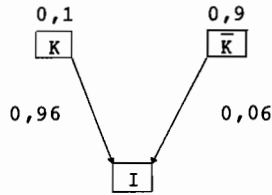
1) In der frequentistischen Interpretation: In der Hälfte aller Fälle, in denen man GS erwischt, wird man die Schublade S ziehen und braucht sich dann die obige Frage gar nicht mehr zu stellen.

unterziehen, wenn wir irgendwelche anderen Anzeichen für die befürchtete Krankheit entdeckt haben, ist die subjektive a priori Wahrscheinlichkeit für K größer als 0,68%, vielleicht 10%:

Diagramm zur
Aufgabenstellung



Intuitiv akzeptiertes
Diagramm



Durch das Indiz revidierte Wahrscheinlichkeitsverteilungen:

K	\bar{K}
0,1	0,9

K	\bar{K}
0,64	0,36

Man kann sich an dieser Stelle trefflich darüber streiten, ob uns hier die Primärintuition einen Streich spielt oder die Mathematiker es sind, die ihre Aufgaben mit unrealistischen bzw. der Primärintuition der Versuchspersonen widersprechenden a priori Wahrscheinlichkeiten konstruieren.

4. Simpsons Paradoxon

Auch dieses Paradoxon, das eigentlich mit der Bayesschen Regel nichts zu tun hat, kann anhand eines umgekehrten Baumdiagramms prägnant dargestellt werden. Seine mathematische Formulierung lautet wie folgt:

- (1) $p(A|B \cap C) < p(A|\bar{B} \cap C)$ und
- (2) $p(A|B \cap \bar{C}) < p(A|\bar{B} \cap \bar{C})$ } implizieren nicht
- (3) $p(A|B) < p(A|\bar{B})$

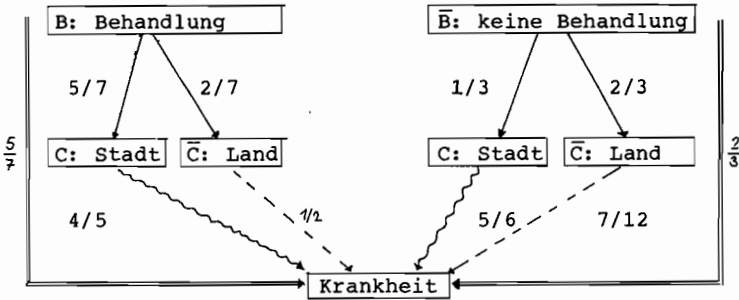
Mit der Interpretation [Morgenstern, S. 48]

A: Erkrankung (Indiz)

B: prophylaktische Behandlung

C: Stadt

erhält es folgende wirklich paradoxe inhaltliche Deutung: Sowohl in der Stadt (C) als auch auf dem Land (\bar{C}) ist die Krankheit bei Behandelten (B) weniger wahrscheinlich als bei Unbehandelten (\bar{B}), während insgesamt die gegenteilige Beziehung gelten kann. Wir übersetzen das von Morgenstern gegebene numerische Beispiel in ein umgekehrtes Baumdiagramm:



Man sieht, daß ein behandelter Landbewohner mit kleinerer Wahrscheinlichkeit ($\frac{1}{2}$) erkrankt als ein nicht behandelter ($\frac{7}{12}$) (----). Ebenso erkrankt ein behandelter Stadtbewohner mit kleinerer Wahrscheinlichkeit ($\frac{4}{5}$) als ein nicht behandelter ($\frac{5}{6}$) (\rightsquigarrow). Insgesamt erkrankt jedoch eine behandelte Person mit größerer Wahrscheinlichkeit ($\frac{5}{7}$) als eine nicht behandelte ($\frac{2}{3}$) (\longrightarrow).

Das Baumdiagramm macht den Grund deutlich: Die bedingten Wahrscheinlichkeiten in (1), (2) (Pfeile $----\rightarrow$ und \rightsquigarrow) müssen noch mit unterschiedlichen Gewichtungsfaktoren multipliziert werden, ehe man aus ihnen die bedingten Wahrscheinlichkeiten (3) (Pfeile \longrightarrow) berechnen kann. Im Beispiel heißt dies: Eine behandelte Person lebt mit sehr großer Wahrscheinlichkeit ($\frac{5}{7}$) in der Stadt, und hier ist die Ansteckungsgefahr trotz Behandlung mit $\frac{4}{5}$ sehr viel größer als die Ansteckungsgefahr ($\frac{7}{12}$) für die überwiegend ($\frac{2}{3}$) auf dem Lande lebenden nicht behandelten Personen.

LITERATUR

Bangen, C.

Wahrscheinlichkeitsrechnung und mathematische Statistik
Frankfurt ⁶ 1974.

Beckenbach, E. F.

Combinatorics for school mathematics curricula. In: The
teaching of probability and statistics. Hrsg. L. Råde.
Stockholm 1970.

Böhme, W.

Erscheinungsformen und Gesetze des Zufalls.
Braunschweig 1964.

Bruner, J. S. (1)

Der Prozeß der Erziehung. Berlin ⁵ 1980.

Bruner, J. S. (2)

Entwurf einer Unterrichtstheorie. Berlin 1974.

Cohen, J. / Hansel, M.

Risk and gambling: The study of subjective probability.
New York 1956.

Davies, C. M.

Development of the probability concept in children.
Child Development 36 (1965): 779-788.

Dinges, H.

Schwierigkeiten mit der Bayesschen Regel.
Mathematisch-Physikalische Semesterberichte 1978/1: 113-156.

Doherty, J.

Levels of four concepts of probability possessed by children
of fourth, fifth and sixth grade before formal instruction.
University of Missouri 1965.

Ehrenfest, P. und T.

Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. Physikalische Zeitschrift 8 (1907): 311-314.

Engel, A. (1,2)

Wahrscheinlichkeitsrechnung und Statistik Bände 1, 2. Stuttgart 1976.

Engel, A. (3)

Bemerkungen zur Statistik. Tagungsbericht der Tagung "Stochastik auf der Schule", Oberwolfach 1978. Zentralblatt für Didaktik der Mathematik 1979/3: 157.

Engel, A. (4)

The probabilistic Abacus. Educational Studies in Mathematics 6 (1975): 1-22.

Engel, A. (5)

Why does the probabilistic abacus work? Educational Studies in Mathematics 7 (1976): 59-69.

Falk, R./ Falk, R./ Levin, I.

A potential for learning probability in young children. Educational Studies in Mathematics 11 (1980): 181-202.

Feller, W. (1, 2)

An introduction to probability theory and its applications. Vol. 1, 2 New York 1950/1966.

Fisz, M.

Wahrscheinlichkeitsrechnung und mathematische Statistik. Berlin 1971.

Fischbein, E. (1)

The intuitive sources of probabilistic thinking in children. Dordrecht 1975.

Fischbein, E. (2)

Image and concept in learning mathematics. Educational studies in Mathematics 8 (1977): 153-165.

Fischbein, E. (3)

Intuition, structure and heuristic methods in the teaching of mathematics. Developments in Mathematical Education: Proceedings of the Second International Congress on Mathematical Education. Hrsg. A. G. Howson. Cambridge 1973: 222-232.

Fischbein, E./ Barbat, I./ Minzat, I.

Intuitions primaires et intuitions secondaires dans l'initiation aux probabilités. Educational Studies in Mathematics 4 (1971): 264-280 (abgedruckt in Fischbein (1)).

Fischbein, E./ Pampu, I./ Minzat, I. (1)

L'intuition probabiliste chez l'enfant. Enfance 2 (1967): 193-208 (abgedruckt in Fischbein (1)).

Fischbein, E./ Pampu, I./ Minzat, I. (2)

Effects of age and instruction on combinatory ability in children. The British Journal of educational Psychology 40 (1970): 261-270 (abgedruckt in Fischbein (1)).

Freudenthal, H. (1)

The empirical law of large numbers or the stability of frequencies. Educational Studies in Mathematics 4 (1972): 448-490.

Freudenthal, H. (2)

The crux of course design in probability. Educational Studies in Mathematics 5 (1974): 261-277.

Goldberg, S.

Probability judgements by preschool children; task conditions and performance. Child Development 37 (1966): 157-167.

Gratch, G.

The development of expectation of the nonindependence of random events in children.

Child development 30 (1959): 217-227.

Heigel, F.

Elementare Ableitung der Laplaceschen Formel der Wahrscheinlichkeitsrechnung. Elemente der Mathematik 1976: 53-57.

Heitele, D. (1)

Über den Wert und die Problematik der vorliegenden Befunde zur Entwicklung stochastischer Ideen beim Kinde. In:

Beiträge zum Mathematikunterricht. Hannover 1975: 49-53.

Heitele, D. (2)

Didaktische Ansätze zum Stochastikunterricht in Grundschule und Förderstufe. Diss. PH Dortmund 1976.

Hirzebruch, F./ Scharlau, W.

Einführung in die Funktionalanalysis. Mannheim 1971.

Hoemann, H. W./ Ross, B. M.

Children's understanding of probability concepts.

Child Development 42 (1971): 221-236.

Hütte

Des Ingenieurs Taschenbuch, Bd. 1. Berlin ²⁸ 1955.

Ineichen, R.

Über die Behandlung der Normalverteilung.

Der Mathematikunterricht 4 (1966): 62-75.

Kac, M.

Random walk and the theory of Brownian motion.

American Mathematical Monthly 54 (1947): 369-390.

Kroll, W.

Ausgleichsrechnung als Anwendung der elementaren Analysis in Grundkursen. Der Mathematische und Naturwissenschaftliche Unterricht 33 (1980): 334-341.

Kütting, H.

Didaktik der Wahrscheinlichkeitsrechnung. Freiburg 1981.

Lambacher/ Schweizer

(Hrsg.) Themenheft Mathematik: Wahrscheinlichkeitsrechnung und Statistik. Stuttgart 1977.

Leake, L.

The status of three concepts of probability in children of the 7th, 8th and 9th grades. University of Wisconsin 1962.

Leffin, W. W.

A study of three concepts of probability possessed by children in the 4th, 5th and 6th grades. Techn. Report 170. University of Wisconsin 1971.

Levy, P.

Calcul des probabilités. Paris 1925.

Lindeberg, J. W.

Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift 15 (1922): 211-225.

Mathai, A. M./ Pederzoli, G.

Characterisations of the normal probability law. New Delhi 1977.

Morgenstern, D.

Didaktische Schwierigkeiten mit der Stochastik. Mathematisch-Physikalische Semesterberichte 1978/1: 31-51.

Müller, G./ Wittmann, E.

Der Mathematikunterricht in der Primarstufe.
Braunschweig 1977.

Pohl, R. W.

Mechanik, Akustik und Wärmelehre. Berlin ¹³ 1955.

Riedl, R.

Biologie der Erkenntnis. Berlin ² 1980.

Scheid, H. (1)

Begriffe der linearen Algebra in der Wahrscheinlichkeitsrechnung. Universität Wuppertal 1979.

Scheid, H. (2)

Didaktik der Stochastik. Universität Wuppertal 1980.

Schrage, G.

Schwierigkeiten mit stochastischer Modellbildung - zwei Beispiele aus der Praxis. Journal für Didaktik der Mathematik 1 (1980): 86-101.

Schröder, H./Uchtmann, H.

(Hrsg.) Einführung in die Mathematik: Angewandte Mathematik. Frankfurt a. M. 1974.

Schrödinger, E./Kohlrausch, F.

Das Ehrenfestsche Modell der H Kurve.
Physikalische Zeitschrift 27 (1926): 306-313.

Shaughnessy, J. M.

Misconceptions of probability. Educational Studies in Mathematics 8 (1978): 295-315.

Strehl, R.

Wahrscheinlichkeitsrechnung und elementare statistische Anwendungen. Freiburg 1974.

Trotter, H. F.

An elementary proof of the central limit theorem.
Archiv der Mathematik 10 (1959): 226-234.

Winter, H.

Erfahrungen zur Stochastik in der Grundschule.
Didaktik der Mathematik 1976/1: 22-37.

Yost, P. A./Siegel, A./Andrews, J.

Nonverbal probability judgements by young children.
Child development 33 (1962): 769-780.