Grundvorstellungen und Grundverständnisse für den Stochastikunterricht

von Peter Bender, Paderborn

Zusammenfassung: Im Stochastikunterricht sollen die Lernenden geeignete Grundvorstellungen und -verständnisse (GVV) für die mathematische Begrifflichkeit ausbilden und dabei ihr Denken, die Mathematik und die Anwendungen zum Passen bringen. Der in den Sekundarstufen fundamentale Begriff der Funktion (als Zuordnung zwischen zwei Mengen) liefert dafür eine Basis, allerdings nicht in einer abstrakten Form, sondern mental in der Lebenswelt verankert. — Für einige zentrale Inhalte der elementaren Stochastik werden solche GVV diskutiert.

1. Zur Unzulänglichkeit des realen Stochastikunterrichts

"Ein (guter) Torwart halte im Mittel 40% aller Elfmeter. Es werden 10 Elfmeter geschossen. Wie groß ist die Wahrscheinlichkeit, daß er genau einen hält?"

Der durchschnittliche Absolvent eines Stochastikleistungskurses in der Sekundarstufe II löst diese Aufgabe etwa folgendermaßen: "Es handelt sich um eine B(n,p)-verteilte Zufallsgröße mit n=10 und p=0,4, für die also $\mu=n\cdot p=4$ und $\sigma=\sqrt{n\cdot p\cdot q}$ =1,55 ist. Es gilt: $p(X=k) = \binom{n}{k}\cdot p^k\cdot q^{n-k}$, und das nimmt für k=1 den Wert 4,03% an." Sind n, k und n-k in einer solchen Aufgabe groß und/oder ist nach einer Wahrscheinlichkeit für ein Intervall gefragt, kann der Proband womöglich noch zur Normalverteilung übergehen, diese standardisieren und mit oder ohne Stetigkeitskorrektur unter schematischer Ausnutzung von Symmetrie-Argumenten den korrekten Wert in einer (aus Platzgründen verkürzten) Tabelle ablesen. Mit diesem Wert wird dann die Frage aus der Aufgabe durchaus sinnvoll beantwortet. Die meisten Anwender von Stochastik ermitteln auf prinzipiell ähnliche Weise bestimmte Werte, mit deren Hilfe sie dann gewisse Aussagen machen können, z.B. "dieser Wert ist hochsignifikant, und deswegen verwerfe ich jene Hypothese."

Ein Hebel, mit dem man bei der o.a. (typisiert aufgeschriebenen) Aufgabenlösung überprüfen kann, wie tief denn das stochastische Verständnis wirklich geht, ist die Frage nach der Bedeutung des (großgeschriebenen) 'X'. — "Na ja, das sind halt die Werte, die vorkommen können." — Auch wenn das Wort 'Zufallsgröße' (oder 'Zufallsvariable') verwendet wird, stößt man rasch an Verständnisgrenzen, besonders wenn 'k' durch ein (kleingeschriebenes) 'x' ersetzt wird (was ist denn

Stochastik in der Schule 17(1997)1, S.8-33

der begriffliche Unterschied zwischen 'X' und 'x'?), wenn (jetzt in anderen Aufgaben) die möglichen Werte mit $x_1, x_2, \ldots, x_m$ bezeichnet werden, andere Zufallsgrößen vorkommen, insbesondere solche, die sich auf Teile eines Experiments beziehen: $X_1, X_2, \ldots, X_r$ (wieso sind diese bis r numeriert, die Werte dagegen bis m?); oder der Mittelwert $\bar{X}$ (ist dasselbe wie der Erwartungswert E(X), der ja wiederum gern als $\mu = \mu_x$ bezeichnet wird und in der beschreibenden Statistik sehr wohl Mittelwert bedeutet?).

Auch wenn die o.a. Aufgabe äußerlich sehr souverän gelöst wird, kann man meiner Meinung nach mit dem mangelhaften Verständnis nicht zufrieden sein: Aus Prinzip nicht, weil man doch wohl das grundlegende didaktische Ziel hat, daß die Lernenden den Stoff *verstehen*; aber auch aus ganz praktischen Gründen nicht: Es sind ja nur als solche erkennbare Standardaufgaben, die da so souverän gelöst werden, und es fehlt die Flexibilität, auch andere Probleme erfolgreich zu bearbeiten.

Nach meiner Einschätzung gründen die o.a. Unzulänglichkeiten in einer nicht genügend stabilen Ausbildung stochastischer Grundvorstellungen und -verständnisse (=GVV; s. Bender 1991a). In vielen Lehrgängen werden zwar anfangs sehr wohl Bezüge zu Alltagsvorstellungen hergestellt, und es wird mit dem gesunden Menschenverstand argumentiert. Aber wie leider im gesamten Mathematikunterricht der Sekundarstufen wird diese Verankerung in der Lebenswelt im Laufe der Zeit gelöst, und auf den mehr oder weniger schematischen Umgang mit Formeln und Vorgehensweisen abgestellt. — Gewiß, eine zentrale Idee der Stochastik, in deren Dienste eigentlich ihre sämtlichen Begriffe stehen, ist die *Bändigung der Variabalität, des Zufalls.* Aber diese Bändigung verkommt zum Rechnen mit Parametern u.ä., weil aus dem Blick gerät, was da gebändigt wird. Stochastisches Denken reduziert sich auf algebraisch-arithmetisches Manipulieren, auf Technik.

Dieser Reduktion muß mit der gezielten Ausbildung geeigneter GVV begegnet werden. Das entsprechende didaktische Konzept (s. Bender 1991a, vom Hofe 1995), das den folgenden Ausführungen unterliegt, stellt auf das Zusammenspiel fachlicher Strukturen (in Mathematik *und* Anwendungen) und mentaler Strukturen ab.

2. Das GVV-Konzept angewandt auf den Stochastikunterricht

Im folgenden soll nicht zum Ausdruck gebracht werden, daß ein Stochastiklehrgang den hier dargestellten Überlegungen gehorchen muß. Selbstredend gibt es auch andere interessante fachliche und didaktische Ansätze. Allerdings halte ich mich im großen und ganzen an den begrifflichen Hintergrund üblicher Stochastiklehrgänge. Desweiteren möchte ich betonen, daß es mir hier nicht um ein Curriculum geht, sondern um Ideen quer zu einem curricularen Aufbau.

2.1 Kombinatorik als ein Paradebeispiel für das GVV-Konzept

Die Kombinatorik wird immer wieder diffamiert, weil sie angeblich Zeit kostet, die dann für die Behandlung echter Anwendungen fehlt. — Andererseits sind aber für die Begriffsbildung zunächst simple lebensweltbezogene Situationen mit Fragestellungen erforderlich, die wohl anwendungsorientiert, aber durchaus auch unrealistisch sein können. Hierzu kann die Kombinatorik einen gewichtigen Beitrag leisten; sie ist geradezu der Prototyp für die Ausbildung von GVV.

Für den kombinatorisch Gebildeten löst sich so manches stochastische Problem in Nichts auf, z.B.: "Welche Konstellation wird bei Familien mit 6 Kindern eher auftreten (vorausgesetzt, daß die Wahrscheinlichkeit für Jungen- (j) und Mädchengeburten (m) gleich sind): mjjmjm oder jjjjjj?" Zwar geht es hier um Wahrscheinlichkeiten, aber die Frage ist eine kombinatorische, und es ist lediglich zu klären, ob (bzw. zu durchschauen, daß) es hier um angeordnete Auswahlen geht.

Kombinatorisches Wissen ist hilfreich für ein Verstehen der gängigen diskreten Verteilungen (zu deren Bedeutung s. 2.4) und der Prinzipien der Beurteilenden Statistik (s.a. 2.7). Das Ziehen von Stichproben läßt sich nun einmal gut als Aufbau komplexer Experimente und deren Wahrscheinlichkeitsverteilungen mit kombinatorischen Mitteln aus einfachen Experimenten und einfachen Verteilungen verstehen.

Dieses Wissen sollten die Lernenden mit einer gewissen Muße, unbelastet von all den formalen, philosophischen und inhaltlichen Problemen des Wahrscheinlichkeitsbegriffs, ausbilden können, zumal der Computer heutzutage die Legitimation der Kombinatorik verstärkt, indem er beliebige Berechnungen mit Binomialkoeffizienten ermöglicht und uns praktisch der Notwendigkeit der Approximation diskreter Sachverhalte durch kontinuierliche Modelle enthebt.

Das Lösen kombinatorischer Aufgaben darf nur nicht auf die Angabe von Aufgaben-Grundtypen mit festgelegten Buchstabenbedeutungen reduziert werden. — Dennoch sollte am Ende eines Kombinatorikkurses eine Übersicht in Form eines Diagramms stehen, als — nicht zeitlicher, sondern — sachlicher *Abschluß*, der noch einmal die Zusammenhänge aufzeigt.

2.2 Wahrscheinlichkeitsraum als Basis für stochastisches Denken

Man verschafft sich einen Überblick über alle möglichen Ausfälle eines Zufallsexperiments und faßt diese zu einer Menge zusammen, die dann einen abgeschlossenen (erneut: das Motiv des Abschlusses) Raum für die weiteren Betrachtungen bildet. Im Zusammenspiel von sachlicher und mentaler Struktur hat das explizite 'Aufstellen' des Wahrscheinlichkeitsraums, das ich auch von fortgeschrittenen Stochastik-Lernenden zu deren (kurzsichtigen) Unmut immer wieder fordere, eine mehrfache Funktion:

Das Aufstellen des Wahrscheinlichkeitsraums ist ein Scharnier zwischen einem realen (oder auch fiktiven, aber in der Realität vorgestellten, auf die Realität bezogenen, aus der Realität idealisierten) Kontext und der mathematischen Begrifflichkeit. Die möglichen Ausfälle des Experiments werden als mathematische Objekte modelliert. Dies geschieht nicht nur, damit dann der Apparat der mathematischen Stochastik angewandt werden kann, sondern es geht um die (durchaus auch alltags-) begriffliche Schärfung des Experiments, um die Herausschälung der relevanten Merkmale, sowie nicht zuletzt um die Bereitstellung eines abgeschlossenen, vertrauten Raums für das Denken. — Dieselben Argumente gelten auch für andere Begriffe der Stochastik (z.B. 'Zufallsgröße'). Es stellt regelmäßig keine Erleichterung für den Lernenden dar, wenn ihm, etwa im Zuge einer durchaus löblichen Konzentration auf die Anwendungen, diese Begriffe vorenthalten oder verwaschen und damit letztlich unverständlich dargeboten werden (s. die Kritik von Ziezold 1982). Jedoch sollen sie von bedeutungsvollen GVV getragen werden, und zwar nicht nur als Aufhänger, sondern auch bei einem elaborierten Umgang.

Natürlich bilden auch die Mengen $\{0, 1, ..., 10\}$ oder $\mathbf{R}$ passende Wahrscheinlichkeitsräume für das Experiment "Halten von 10 Elfmetern". Aber es geht mir um Räume, die das zugrundeliegende Experiment möglichst treu wiedergeben, hier etwa die Menge aller 0-1-Folgen der Länge 10 (vom Umfang 2^{10}). Natürlich muß man nicht bei jeder solchen Aufgabe auf einen solchen experimentähnlichen

Wahrscheinlichkeitsraum zurückgehen, aber man muß jedesmal dazu in der Lage sein. Dabei ist die gezielt konkretisierte, und damit im Vergleich zu einer Handlung verfremdete Vorstellung nützlich, daß für jeden der (hier: 1024) Ausfälle des Experiments entsprechend seiner Wahrscheinlichkeit ein Anteil des Vollwinkels auf einem Glücksrad (o.ä.) vorgesehen ist.

Eigentlich ist das Torwart-Beispiel gar nicht so gut geeignet, weil man es ja lediglich bei hochidealisierter Sichtweise als *Zufallsexperiment* ansehen kann. Diesen Mangel teilt es mit vielen Beispielen aus vielen Lehrgängen, zumal Schulbüchern. Man kann es in dieser Beziehung retten, indem man davon spricht, daß man von den tausenden von Elfmetern (insgesamt N Stück), die der Torwart im Laufe eines Jahres trainiert hat, immer wieder einmal zufällig einen beobachtet hat, insgesamt 10. D.h.: Ein Glücksrad G_N ist in N gleichbreite Abschnitte eingeteilt, davon 40% weiß und 60% schwarz, und es wird 10mal gedreht. Mit seinen N Feldern liefert das Glücksrad aber noch nicht direkt den Wahrscheinlichkeitsraum; es muß vielmehr noch das *mehr*malige Drehen einbezogen werden. Diese Aktionen sind in statische Formen zu gießen, die dann zu Elementen eines Wahrscheinlichkeitsraums werden.

Diese Überlegungen führen zu einem sehr umfangreichen *UrWahrscheinlichkeitsraum* W_N10, bestehend aus allen N^{10} Wörtern der Länge 10 über einem Alfabet mit N Buchstaben mit Wiederholung, die alle die *gleiche Wahrscheinlichkeit haben; — das Ideal einer Stichprobenauswahl.* — Das zugehörige Glücksrad G_N10 ist in N^{10} gleichbreite Ausschnitte eingeteilt, für jeden Ausfall einen. Und hierfür, d.h. für den Fall der Gleichverteilung, erscheint mir das Bild einer *Urne* mit N^{10} *Zetteln* für die N^{10} Wörter besonders passend.

Für i=1,2,...,10 ordnet die Zufallsgröße X_i jedem Ausfall eine 1 zu, falls die i-te Drehung von G_N weiß ergibt, und eine 0 sonst, bzw. — viel direkter — den Wert des Wortes an der i-ten Stelle, und die Zufallsgröße $X = \sum_{i=1}^{10} X_i$ schließlich ordnet jedem Ausfall die Häufigkeit von 'weiß' bei seinen 10 Drehungen, bzw. die Anzahl der Einsen im Wort, zu.

Man merkt, daß man gar nicht das Glücksrad G_N mit seinen N Bereichen braucht, sondern ein Glücksrad G_2 mit nur *einem* weißen Bereich, der 40% und *einem* schwarzen, der 60% einnimmt. Indem man nur noch die beiden Klassen betrachtet, kommt man zur Beschreibung des Experiments mit 10 Elfmetern auf einen Wahrscheinlichkeitsraum $W_2 10$ mit 2^{10} Ausfällen, in denen die Ausfälle aus dem

$W_N 10$ in unterschiedlichen Mächtigkeiten zusammengefaßt sind und die daher nicht mehr gleichwahrscheinlich sind. Dazu gehört statt des Glücksrads G_N10 ein $G_2 10$ mit 2^{10} Bereichen, und wegen der Abwesenheit von Gleichwahrscheinlichkeit paßt das Urnenmodell mit (gleichwahrscheinlichen) Zetteln nicht mehr. — $W_2 10$ gehört eher zu der Situation, daß 10 Elfmeter hintereinander geschossen werden, während $W_N 10$ treuer die Situation beschreibt, daß aus vielen tausenden Elfmetern 10 ausgewählt werden.

Sinnvollerweise behandelt man beide Glücksräder mit ihren Wahrscheinlichkeitsräumen und setzt sie in Beziehung zueinander. — Dabei kann es sich auch anbieten, den von mir so bezeichneten Ur-Wahrscheinlichkeitsraum $W_N 10$ erst später zu betrachten, u.a. damit die Gleichwahrscheinlichkeit nicht immer den gedanklichen Primat hat.

Selbstredend kann man auch direkt von dem durch die Zufallsgröße X vermittelten Wahrscheinlichkeitsraum W_{11} mit den natürlichen Zahlen von 0 bis 10 als Ausfällen und einem Glücksrad G_{11} ausgehen. Allerdings benötigt man zum Verstehen und zum Berechnen von dessen Wahrscheinlichkeitsverteilung doch einen der o.a. Ur-Wahrscheinlichkeitsräume. Vor allem aber gerät bei der Betrachtung von W_{11} der Stufen- und Stichprobencharakter des Experiments in den Hintergrund.

Immerhin hat W_{11} noch den Vorzug, daß das Experiment mit einer einzigen Drehung des Glücksrads vollzogen ist. — Weniger günstig erscheint mir dagegen ein Stehenbleiben bei G_N bzw. G_2 und 10maligem Drehen. Da wird nämlich der Akzent auf die Wahrscheinlichkeit von 0,4 für das Halten eines einzigen Elfmeters gelegt, deren Konkretisierung an dieser Stelle des Stochastiklehrgangs bereits trivial sein sollte, während es jetzt darauf ankäme, die N^{10} bzw. 2^{10} Ausfälle sichtbar zu machen. Dies wird bei G_N bzw. G_2 nicht genügend geleistet, weil sich hier die Ausfälle in der (als solche unsichtbaren) Wiederholung verbergen.

Nicht erst das Rechnen, sondern schon der Begriff des Wahrscheinlichkeitsraums, die Vorstellung von einem Glücksrad, die Zeichnung eines Baumdiagramms usw. nehmen einem Zufallsexperiment eigentlich schon einen Großteil seiner Dynamik, — das ist geradezu die Aufgabe solcher Modelle hier und anderswo inner- und außerhalb der Mathematik. Es geht darum, flüchtige Abläufe in statische Formen zu bannen und einen Ablauf damit als Ganzes und im Vergleich mit anderen (die vielleicht real gar nicht mit ihm vereinbar wären) dauernd vor dem geistigen Auge zu haben.

Einen dynamischen Wesenszug hat aber der Wahrscheinlichkeitsraum dennoch: Er modelliert den Sachverhalt nur *vor* Durchführung des Experiments. — 'Das Experiment durchführen' bedeutet nun, den Wahrscheinlichkeitsraum (= Zusammenfassung der Möglichkeiten) zu realisieren, ihn aufzulösen; denn dann steht ja fest, welcher Ausfall eingetreten ist, und alle anderen Möglichkeiten bestehen als solche nicht mehr. Die Situation wird *nun* durch den *einen* realisierten Ausfall beschrieben, der eine Wahrscheinlichkeit von 100% hat. — Möchte man das Experiment (eventuell nur in Gedanken) wiederholen, braucht man einen neuen Wahrscheinlichkeitsraum, und erst wenn man sich vergewissert hat, daß die Situation dieselbe ist wie vor der ersten Durchführung (die gezogene Kugel ist wieder zurückgelegt und gut untergemischt; am Würfel ist nichts abgeplatzt usw.), kann man den alten *wieder verwenden*.

Hier haben wir einen Prototyp für die von Wiesemann (1995) auf den Punkt gebrachte Unterscheidung zwischen mathematischer und metamathematischer Aktion und die dabei auftretenden scheinbaren Paradoxien zwischen der gleichzeitig vorliegenden Konstanz und Veränderung von mathematischen Gegenständen dabei.

Die Vorstellung der Auflösung des Wahrscheinlichkeitsraums nach Durchführung des Experiments zwingt dazu sich klar zu machen, wann das Experiment als durchgeführt gilt. Bei vielen Anwendungen wird etwas, das schon für sich Experiment-Charakter und einen eigenen Wahrscheinlichkeitsraum hat bzw. haben könnte, n-mal durchgeführt. Diese mehrmalige Durchführung ist eigentlich das Experiment, und dieses besteht aus der n-maligen Hintereinanderausführung dessen, was zunächst 'Experiment' genannt wurde und was man jetzt 'Experiment*stufe*' nennen könnte.

Der Wahrscheinlichkeitsraum ist erst aufgelöst, wenn alle n Experimentstufen durchlaufen sind. Allerdings kann man ihn nach jeder Experimentstufe verkleinern; denn bis dahin ist ja bereits ein bestimmtes Ereignis eingetreten (100%), und die Alternativen bei dieser Stufe nicht (0%). — Hier schließen sich die Begriffe der bedingten Wahrscheinlichkeit, Pfadwahrscheinlichkeit usw. an.

Dagegen ist die Modellierung einer Stichprobe vom Umfang n durch eine Abfolge von n (identischen) Wahrscheinlichkeitsräumen weniger gut geeignet. Ein stimmiger Gebrauch von Zufallsgrößen (als Funktionen in meinem Sinn) setzt nämlich voraus, daß diese, wenn sie miteinander verknüpft werden sollen, ein und denselben Definitionsbereich haben, z.B. die X_i, $i=1,2,\ldots,n$, die der Experimentstufe Nr. i einen Wert zuordnen, deren Summe X oder deren arithmetisches Mittel $\overline{X}$.

Beispiel (nach Engel 1987, S.194ff): In 6 Städten S_1 bis S_6 werden der Fluorgehalt des Trinkwassers f_1 bis f_6 und die Kariesrate bei Kindern k_1 bis k_6 erhoben und in Rangplätze x_1 bis x_6 bzw. y_1 bis y_6 übersetzt. Für diese 6 Städte hat man damit einen Zusammenhang zwischen diesen beiden Größen ermittelt, der sich mit Hilfe des Rangkorrelationskoeffizienten r formalisieren läßt. — Möchte man dieses Ergebnis mit stochastischen Mitteln interpretieren, so braucht man 'irgendwie' ein Zufallsexperiment. Woher nimmt man dieses? Man betrachtet (sämtliche) (deutsche) Städte, schreibt für sämtliche (angeordnete) Folgen von solchen Städten (ohne Wiederholung) der Länge 6 je einen Zettel, legt die Zettel gut gemischt in eine Lostrommel und zieht dann einen. Die Rangplätze und der Rangkorrelationskoeffizient werden dadurch zu Zufallsgrößen X_i, Y_i und R. Für unabhängige X_i und Y_i ist $E(R)=0$ und $Var(R)=0{,}2$ (in $\frac{1}{n-1}$ ist hier n=6 gesetzt). Eine starke Abweichung des in der realen Untersuchung ermittelten Werts r von 0 spricht also für eine Abhängigkeit von Fluorgehalt und Kariesrate.

Für das Frühstadium eines Stochastiklehrgangs möchte ich als Prototyp eines Zufallsexperiments den gleichzeitigen Wurf eines grünen und eines roten Würfels empfehlen, wobei ich die Assoziation der "Würfelbuden-Mathematik" für unschädlich halte. Der zugehörige Wahrscheinlichkeitsraum mit 36 Würfelzahlenpaaren ist gut zugänglich und leicht visualisierbar, und zugleich komplex genug, damit die stochastische Begrifflichkeit nicht trivial wird. Dieser Prototyp kann auch leicht variiert werden (einen Würfel zweimal werfen, mit oder ohne Beachtung der Reihenfolge usw.).

2.3 Die Rolle des erkennenden Subjekts bei der Generierung einer Wahrscheinlichkeitsverteilung

Nach meinem Dafürhalten wird in den gängigen Vorschlägen zu wenig die Beteiligung des erkennenden Subjekts bei der Generierung einer Wahrscheinlichkeitsverteilung betont. Indem *ich* von einem idealen Würfel rede, lege ich fest, daß jeder Ausfall die Wahrscheinlichkeit $\frac{1}{6}$ erhält. Entsprechendes gilt für die Unterstellung einer Binomial-, einer Normalverteilung, der Unabhängigkeit von Zufallsgrößen usw. Auch der (mehr anwendungsorientierte) Weg über relative Häufigkeiten (bei den beiden Lagen des Reißbrettstifts:  , beim gezinkten

Würfel, beim Torwart, bei der Wahlprognose, kurz: bei jeder Stichprobe) ist nicht objektiv: *Ich* erhalte dabei lediglich Anhaltspunkte für die Verteilung der Wahrscheinlichkeit auf die Ausfälle. In der Beurteilenden Statistik ist der Grad *meiner* Überzeugung von einer bestimmten Hypothese eine subjektive Sache. — Dieser subjektivistische Zug bedeutet keineswegs Willkür, sondern es existiert dafür eine ausgearbeitete Theorie (s. Wickmann 1990, Riemer 1989), die allerdings von der Schulstochastik i.w. ignoriert wird.

Auch in einem 'objektivistisch' verstandenen Konzept von Wahrscheinlichkeit erscheint mir folgendes Bild hilfreich: Ich habe eine Wahrscheinlichkeits'masse' (die ich mir als elastischen, beliebig teilbaren und gut formbaren Stoff vorstelle) vom Umfang 1 in der Hand und verteile diese Masse auf die Menge der Ausfälle des Experiments, den Wahrscheinlichkeitsraum. Jeder Ausfall bekommt den Anteil zugewiesen, den ich aufgrund theoretischer Überlegungen oder empirischer Untersuchungen für angemessen halte.

Man kann dieses Bild noch konkretisieren, indem man sich die Ausfälle als Zellen eines quadratischen Gitters vorstellt und auf diesen Zellen Säulen errichtet mit der zugewiesenen Wahrscheinlichkeit als Höhe. Die 'Glücksfee', die nachher das Zufallsexperiment durchführt, wird dann eher bei einem Ausfall mit höherer Säule hängen bleiben. — Allerdings muß man hier aufpassen, daß die Lernenden reif genug sind und all diese Metaphern nur als solche verstehen.

Hat man den Wahrscheinlichkeitsraum mit den Spitzen eines Baumdiagramms identifiziert, das für einen Zufallsprozeß steht, dann kann man die Wahrscheinlichkeitsmasse durch den ganzen Baum 'drücken': An jedem Verzweigungspunkt kommt ein bestimmter Teil an, und dieser wird auf die von da abgehenden Wege entsprechend deren (sich zu 1 aufsummierenden) bedingten Wahrscheinlichkeiten verteilt und weitergeschoben. An jeder Baumspitze kommt schließlich der ihr zustehende Anteil von Wahrscheinlichkeit an.

Selbstverständlich darf der Zusammenhang mit relativen Häufigkeiten nicht unterschlagen werden. Das Wissen um sie hat ja schon die Verteilung der Wahrscheinlichkeitsmasse beeinflußt. Umgekehrt liefert der Begriff der relativen Häufigkeit für viele stochastische Größen ein operatives Verständnis: Hat ein Ereignis E die Wahrscheinlichkeit 0,4 oder eine Zufallsgröße X den Erwartungswert μ und die Standardabweichung σ, dann kann man bei 1 Million unabhängigen Durchführungen des Experiments etwa 400.000mal mit E rechnen, und die Werte, die X annimmt, streuen um den Mittelwert μ, und zwar umso stärker, je größer σ ist. —

Die Genauigkeitseinschränkungen 'etwa' und 'streuen' haben nichts damit zu tun, daß die erhaltenen Werte ungenau seien; diese Einschränkungen sind auch bei einem idealen Wahrscheinlichkeitsraum mit genauen Werten erforderlich. Und: Die 1.000.000malige Durchführung des Experiments kann (ideal!) sogar erhebliche Abweichungen der realisierten Werte von den idealen Werten ergeben. — Hier hat man eine operative Grundlage für den Begriff der *stochastischen* Konvergenz, wenn man denn diesen behandeln will.

Ist bei einer *bedingten Wahrscheinlichkeit* die Bedingung durch das Ereignis B gegeben, so bedeutet dies in dem Bild, daß die komplette Wahrscheinlichkeitsmasse auf B konzentriert wird (außerhalb B ist die Wahrscheinlichkeit 0, bzw. der Wahrscheinlichkeitsraum ist auf B reduziert) und jeder Ausfall entsprechend seinem Anteil an B neu mit Wahrscheinlichkeitsmasse bedacht wird.

Indem schließlich das Experiment durchgeführt wird und die Glücksfee sich für einen Ausfall entscheidet, werden von allen anderen Ausfällen die Wahrscheinlichkeiten ab- und auf diesen einen zusammengezogen. Besteht das Experiment aus mehreren Stufen, so kann man sich diese Konzentration von Wahrscheinlichkeit auch stufenweise vorstellen: Auf jeder im Experiment erreichten Stufe wird die Wahrscheinlichkeit von allen Alternativen auf die eine übertragen, die da realisiert wird. Alle (bis auf eine) Alternativen mit den Teilbäumen, die aus ihnen sprießen, werden quasi ausgetrocknet, und die realisierte Alternative erhält den ganzen Segen (wobei hier auf jeder Stufe jede Alternative einschließlich des Pfades von der Wurzel bis zu ihr zu sehen ist). Dies führt dazu, daß die (unbedingten) Wahrscheinlichkeiten der jetzt noch erreichbaren Verzweigungspunkte proportional wachsen. — Dies geht so lange, bis auf der letzten Stufe die komplette Wahrscheinlichkeit auf demjenigen Ausfall konzentriert ist, mit dem schließlich das gesamte Experiment realisiert wurde. — Die vor dem Experiment mühevoll durchgeführte Verteilung der Wahrscheinlichkeitsmasse nacheinander auf sämtlichen Stufen bis zu den Spitzen des Baumdiagramms ist wieder vollständig zurückgenommen.

Eine andere luzide (allerdings nur qualitative) Veranschaulichung der Wahrscheinlichkeitsverteilung (statt Säulen) ist Farbintensität: Es steht rote Farbe (= Wahrscheinlichkeit) zur Verfügung, und jeder Ausfall wird gefärbt, und zwar wird die Farbe entsprechend der Wahrscheinlichkeit dieses Ausfalls mehr oder weniger verdünnt aufgetragen. — Das Merkmal 'Farbintensität' hat zwei Vorzüge: Mit ihm können auch noch Funktionen mit dreidimensionalem Definitionsbereich (wenigstens in der Vorstellung) veranschaulicht werden; und: Bei einem kontinuierlichen Definitionsbereich entspricht (mental!) Farbintensität besser der dort relevanten Wahrscheinlichkeitsdichte (s. 2.4). — Farbintensität, wie auch die Säulen, sollten eigentlich sowieso aus der Funktionenlehre der Sekundarstufe I bekannt sein als Darstellungsmöglichkeiten für Funktionsgraphen.

2.4 Kontinuierliche Verteilungen

Auch eine kontinuierliche Verteilung kann man sich gut mit Wahrscheinlichkeits'masse' vorstellen. Hat man den (ein- oder zweidimensionalen) Wahrscheinlichkeitsraum als reelles (eventuell unendliches) Intervall oder als (eventuell unendlich große) Kreisscheibe modelliert, so verteilt man die Masse darauf, und es entsteht ein Funktionsgebirge, oft ein einziger Berg. — Bemerkenswert ist, daß mit einer endlichen Masse ein unendlicher Bereich komplett bedeckt werden kann. Das geht natürlich nicht, wenn die Masse überall gleich dick aufgestrichen werden soll; aber es geht, wenn man sie zu den Außenbereichen hin sehr dünn und immer dünner aufträgt. Bei **R** (eventuell $\mathbf{R}^{>0}$) als Definitionsbereich ist z.B. $\frac{1}{|x|}$ zu dick aufgetragen, aber $\frac{1}{x^2}$ oder $\exp(-x^2/2)$, mit einer passenden Konstanten multipliziert, würden funktionieren (was man in der Analysis untersucht). In jedem Punkt (Ausfall) besteht das Funktionsgebirge aus einer Säule (zwar der Grundfläche 0, aber doch) endlicher Höhe. Deren Interpretation führt zu einem mehrfachen mentalen Konflikt:

Einerseits leuchtet ein, daß bei einer *stetigen* Verteilung jeder der unendlich (ja, überabzählbar) vielen Ausfälle allein (formal als Elementarereignis aufgefaßt) die Wahrscheinlichkeit 0 haben muß; denn sonst bräuchte man für jede, noch so kleine, Umgebung um einen Punkt mit Wahrscheinlichkeit >0 eine unendlich große Wahrscheinlichkeitsmasse. Andererseits kommt ja jeder Ausfall für die Realisierung in Frage, ist also nicht unmöglich, seine Wahrscheinlichkeit ist also nur "praktisch", d.h. im Vergleich zum Rest des ganzen Wahrscheinlichkeitsraums, 0. — Ist seine Wahrscheinlichkeit nun =0 oder nur "praktisch" 0? — Hier ist es hilfreich, wenn man Fragen dieser Art bereits in der Elementargeometrie in der Sekundarstufe I diskutiert hat: Wenn man z.B. bei einem Rechteck mit Seiten (-längen) a>0 und b>0 die Seite b (naiv, ohne Grenzwertbegriff, kontinuierlich) gegen 0 gehen läßt, so 'ergibt' sich, daß die Strecke a (gleich dem Rechteck mit Seitenlängen a und 0) den Flächeninhalt =0 hat, obwohl sie kein Nichts ist. — Der epistemologische Konflikt ist in der Elementargeometrie natürlich derselbe, aber dort auf seinen Kern reduziert, von stochastischen Mysterien befreit und daher leichter zu bewältigen oder wenigstens zu akzeptieren.

Der Konflikt geht insofern noch weiter, als man das ganze Rechteck (den ganzen Wahrscheinlichkeitsraum) wieder zusammensetzen kann aus lauter Strecken mit Länge a und Flächeninhalt 0 (aus lauter Elementarereignissen mit Wahrscheinlichkeit 0), und man als Gesamtflächeninhalt (als Summe von unendlich vielen Nullen!) eine Zahl a·b>0 (bzw. eine Gesamtwahrscheinlichkeit von 1) erhält. — Gewiß, in der Integralrechnung kann man lernen, wie man mit Hilfe des Grenzwertbegriffs diesen Konflikt auflöst. Ich meine zwar, daß der Grenzwertbegriff mit dem 'gesunden Menschenverstand' erfaßbar ist (s. Bender 1991b), muß aber zur Kenntnis nehmen, daß die wenigsten Abiturienten ihn erfaßt haben. Außerdem tritt der Konflikt schon, wie gesagt, in der Elementargeometrie sowieso, aber auch in der Stochastik oft vor und unabhängig von der Integralrechnung auf.

Der Hauptkonflikt ist aber weniger fachlich verursacht, sondern rührt aus der (natürlichen!) Vorgehensweise: Wir sind doch froh, wenn die Lernenden bei einem ein- oder zweidimensionalen Funktionsgebirge in jedem Punkt des Definitionsbereichs eine Säule sehen, die den Funktionswert in diesem Punkt visualisiert. Nach allem, was man vorher über diskrete Wahrscheinlichkeitsverteilungen gelernt hat, drängt sich die Interpretation dieser Werte als Wahrscheinlichkeiten geradezu auf.

Zu dieser Fehlinterpretation trägt auch bei, daß kontinuierliche Verteilungen ja gern, spätestens im Zusammenhang mit dem zentralen Grenzwertsatz, pauschal als 'Grenzverteilungen' diskreter Verteilungen identifiziert werden. Mit moderner Computergraphik oder auch mit einigen Zeichnungen und der Vorstellung läßt sich der Übergang z.B. zwischen Binomial- und Normalverteilung *optisch* sehr gut veranschaulichen. — Hier werden aber FVV von Folgen und ihren Grenzwerten gepflegt: Die Normalverteilung stünde am 'Ende' eines Grenzprozesses, bei dem das n von B(n,p)-Verteilungen immer größer gemacht wird (wie der Kreis als 'Grenzwert' seiner einbeschriebenen regelmäßigen n-Ecke bei der Berechnung von π), wodurch mehr Gemeinsamkeiten eines solchen 'Grenzwerts' mit den Folgengliedern suggeriert werden, als vorhanden sind.

Zur Lösung des Hauptkonflikts könnte man einmal die Zufallsgrößen X_n, $n \in \mathbf{N}$, die dem n-maligen Wurf einer fairen Münze die Anzahl der Wappen zuordnen, und ihre Säulendiagramme betrachten. Die Gipfel dieser Graphen befinden sich an den Stellen x= n/2 und verlagern sich mit wachsendem n nach rechts. — Nehme ich mir eine beliebige Stelle x vor und gebe mir eine, noch so kleine, Schranke $\varepsilon>0$ vor, dann kann ich n so groß wählen, daß $p(X_n=x)<\varepsilon$ ist. Dies liegt aber nicht nur daran, daß ich die Gipfel weit genug nach rechts wandern lassen kann.

Ich kann nämlich auch die Zufallsgrößen $Y_n = X_n - \frac{n}{2}$ betrachten, die ihre Gipfel alle in x=0 haben. Lasse ich n wachsen, so werden die Gipfelhöhen beliebig klein (>0), weil die Standardabweichungen σ_n der X_n (und damit der Y_n) $\frac{1}{2} \cdot \sqrt{n}$ betragen, also mit wachsendem n beliebig groß werden, d.h. die Verteilungen zunehmend nach rechts und links auseinanderfließen. Ich kann n so groß machen, daß sich die Verteilung beliebig wenig von der x-Achse abhebt.

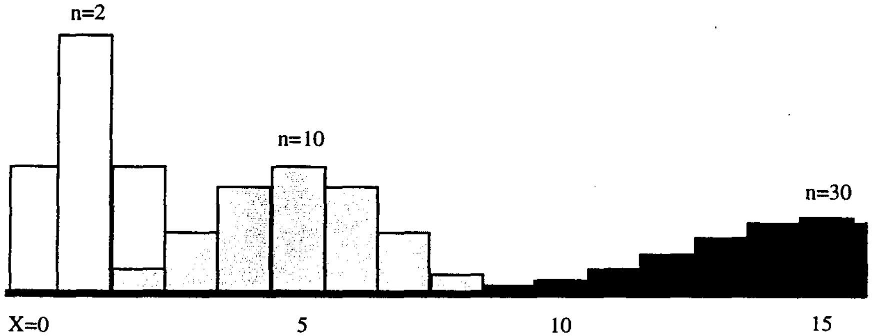

Abb. 1: (nicht-standardisierte) Wahrscheinlichkeitsverteilung (im Originalmaßstab) für die Anzahl X der Wappen bei n Münzwürfen

Erst wenn ich die standardisierten Zufallsgrößen $Z_n = \dfrac{X_n - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$ betrachte, erhalte ich Graphen, die zwar nach wie vor Treppenfunktionen darstellen, aber grob dem Verlauf der Glockenkurve folgen, und zwar umso enger, je größer n ist, für $n \cdot p \cdot q > 9$ schon recht gut, und sich jedenfalls nicht der x-Achse nähern. Es ist dieses Zusammendrücken der für alle X_n einheitlichen Abszisseneinheit e=1 durch die Division durch σ_n zu einer neuen Einheit e_n für jedes Z_n, wodurch in jedem Intervall des Definitionsbereichs bei beliebig wachsendem n immer genügend Wahrscheinlichkeitsmasse erhalten bleibt, so daß diese dort nicht gegen Null tendiert, sondern sich auf einem gewissen Niveau stabilisiert. Dabei liegt es von vorneherein nahe, gerade durch die Standardabweichungen zu dividieren, damit die Zufallsgrößen Z_n die Standardabweichungen 1 haben.

Die Glockenkurve (Graph der standardisierten Normalverteilung) beschreibt dann an jeder Stelle z, wieviel Wahrscheinlichkeitsmasse an dieser Stelle ungefähr einem Intervall der Länge $\frac{1}{\sigma_n}$ zukommt. — Diese Aussage bezieht sich zunächst auf sämtliche (diskrete) Zufallsgrößen Z_n (für $n \cdot p \cdot q > 9$) und erstreckt sich auf die zusammengefaßte Wahrscheinlichkeitsmasse von sämtlichen Werten k, für die $\dfrac{k - \frac{n}{2}}{\sigma_n}$ im Intervall der Länge $\frac{1}{\sigma_n}$ um z liegt. Sie kann aber auch auf eine (kontinuierliche) N(0,1)-verteilte Zufallsgröße bezogen werden.

Im realen Unterricht der Sekundarstufe II muß immer wieder hinterfragt werden, ob solche Überlegungen für die Mehrheit der Schüler durchschaubar sind, und damit, ob sie wirklich geeignet sind, den oben genannten Hauptkonflikt aufzulösen. Es ist eher zu erwarten, daß die Funktionswerte etwa der Glockenkurve nach wie vor als Wahrscheinlichkeiten interpretiert werden und daß der Widerspruch dazu, daß diese 0 sein müßten, bestehen bleibt. Beim schematischen Rechnen stört dieser ja nicht.

Immerhin ist aber wenigstens das Faktum deutlich geworden, daß nicht nur die Säulenhöhen, sondern auch ihre Breiten (bzw. Grundflächen) eine Rolle spielen. Weist man diesen das Maß $\frac{1}{\sigma_n}$ zu, dann hat die Gesamtfläche (entsprechend das Gesamtvolumen) unter dem Graph das Maß 1 (die gesamte Wahrscheinlichkeitsmasse). — Hierzu sollte man schon in der Beschreibenden Statistik bei der Darstellung relativer Häufigkeiten Erfahrungen gemacht haben: Bildet man z.B. Klassen, d.h. faßt man einzelne Merkmalsausprägungen zusammen, so muß man die Höhe jeder neuen Säule so einrichten, daß ihr Inhalt dem Gesamtinhalt der durch sie zusammengefaßten Säulen gleich ist.

Ebenso sollte man dort schon aufsummierte (relative) Häufigkeiten kennengelernt haben. Als Zugang zur Normalverteilung über Binomialverteilungen bieten sich m.E. *aufsummierte* Wahrscheinlichkeiten besser denn Wahrscheinlichkeiten selbst an. Bei größeren n interessiert man sich doch kaum noch für die Wahrscheinlichkeiten einzelner Werte, sondern für die Wahrscheinlichkeiten ganzer Intervalle, und zwar häufig solcher, die ganz links oder ganz rechts liegen. —

Nimmt also eine Zufallsgröße X die Werte $x_1 < x_2 < ... < x_n$ mit den Wahrscheinlichkeiten w_i, $i=1,2,...,n$, an, dann lauten die (von links) aufsummierten Wahrscheinlichkeiten $s_0=0$, $s_i=s_{i-1}+w_i$, $i=1,2,...,n$, woraus $s_n=1$ folgt. Diese Funktion ist immer monoton steigend. Für eine binomialverteilte Zufallsgröße erhält man einen S-förmigen Verlauf zwischen den Geraden $y=0$ und $y=1$.

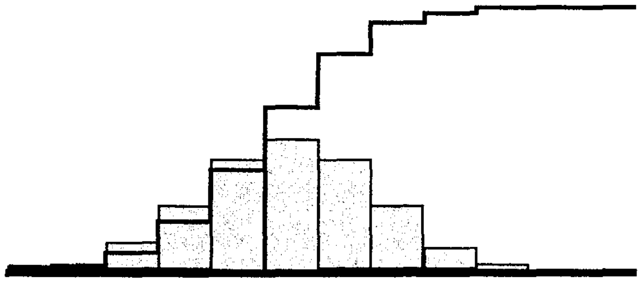

Abb. 2: Eine diskrete Wahrscheinlichkeitsverteilung und ihre (auf die Hälfte gestauchte) Summenfunktion

Man hat hier ein sinnfälliges Beispiel im Diskreten für den Zusammenhang zwischen Summen- und Differenzenfunktion. Denn aus der Funktion der aufsummierten Wahrscheinlichkeiten erhält man durch $w_i=s_i-s_{i-1}$, $i=1,2,...,n$ die Funktion der Wahrscheinlichkeiten selbst. Jede der beiden Funktionen allein bestimmt also, wie gesagt: im Diskreten, die Wahrscheinlichkeitsverteilung vollständig.

Wenn man hier nun, unvermeidlich mit ähnlich schwierigen Überlegungen wie oben, zum kontinuierlichen Fall übergeht, so hat man jedenfalls kein Problem mit dem Verständnis der Funktionswerte der sog. Verteilungsfunktion: Für jede Stelle x gibt die Funktion den Wert $p(X \leq x)$ an, genau wie im diskreten Fall. — Bei der Normalverteilung ergibt sich die Glockenkurve als Ableitungsfunktion, analog zum diskreten Fall. Deren Funktionswerte sagen also etwas aus über die Stärke der Veränderung bei der Verteilungsfunktion; und diese wiederum ist die Flächeninhaltsfunktion für die Glockenkurve. Man kann ihre Werte als Wahrscheinlichkeitsdichte, und sie selbst als Dichtefunktion interpretieren: Wo die Werte am größten sind, ist die Wahrscheinlichkeit am dichtesten und der Anstieg der Verteilungsfunktion am stärksten. — Nun sieht man, warum die in 2.3 erwähnte Farbintensität eine gut geeignete Veranschaulichung für die Wahrscheinlichkeitsverteilung (-dichte) ist.

Der Zusammenhang zwischen Ableitung und Integral, sowie seine Entsprechung im diskreten Fall sind wieder nichts spezifisch Stochastisches, sondern sind ein Teil der Analysis (s. dazu Bender 1990). Verteilungs- und Dichtefunktion liefern ein weiteres Beispiel für funktionale und 'differentiale' Zusammenhänge zwischen Größen unterschiedlichster Art (Blum & Törner 1983, S.92). Solche Beispiele gehören m.E. zu den GVV, die jeder Lernende der Analysis besitzen sollte. Ich würde kontinuierliche Wahrscheinlichkeitsverteilungen gern sowieso erst nach dem Hauptsatz der Integral- und Differentialrechnung behandeln, und zwar als eine komplexe Anwendung der Analysis, i.w. beschränkt auf die Normalverteilung. Der Flächeninhalt unter dem Graph von $\exp(-x^2/2)$ muß zwar zur Motivierung der multiplikativen Konstanten genannt, aber nicht hergeleitet werden, usw.

Mit Recht weist Engel (1987, S.7) indirekt darauf hin, daß die Normalverteilung (wie andere kontinuierliche Verteilungen) im Zeitalter des Computers eine früher entscheidende Funktion verloren hat: Da die Ausrechnung von aufsummierten Wahrscheinlichkeiten bei einer Binomialverteilung wegen der unhandlichen Binomialkoeffizienten praktisch nicht geleistet werden konnte, standardisierte man diese Verteilungen, approximierte sie durch die standardisierte Normalverteilung und las deren Werte in einer einzigen Tabelle ab, und zwar für die praktischen Erfordernisse allemal genau genug. Heute kann man die gewünschten Rechnungen direkt vom Computer durchführen lassen und braucht den Umweg über die Tabelle der Normalverteilung nicht mehr.

Offensichtlich entsprechen diskrete Verteilungen realen Problemen regelmäßig direkter als kontinuierliche Verteilungen, einmal weil es sowieso um endliche Stichproben geht, zum anderen, weil Messungen nur bis zu einer gewissen Genauigkeit möglich und auch nur von Interesse sind, so daß ein Verzicht auf die Behandlung kontinuierlicher Verteilungen durchaus sachangemessen ist.

Natürlich wird man den zentralen Ideen der Stochastik besser gerecht, wenn man auch wenigstens die Normalverteilung behandelt: Als Grenzverteilung von zahlreichen Folgen diskreter Verteilungen bildet sie einen inhaltlichen *Abschluß*, als wesentliches Element des zentralen Grenzwertsatzes hat sie für die Stochastik in Theorie und Praxis einen universellen Charakter, und als Anwendung der Analysis ermöglicht sie Erkenntnisse in beiden Bereichen.

Kontinuierliche Verteilungen haben darüber hinaus den (gewiß sekundären) Vorzug, daß kontinuierliche Wanderungen im Definitionsbereich ebensolche Änderungen der Werte nach sich ziehen. Man kann für vorgegebene Werte i.a. scharfe

Intervallgrenzen angeben, und auch beim Zeichnen tut man sich leichter mit einem schwungvollen Graph als mit einer Treppenfunktion mit lauter gleichlangen waagrechten Strecken, bei denen benachbarte Enden und Anfänge genau lotrecht übereinander liegen müssen.

2.5 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Bekanntlich rührt ein Großteil der Schwierigkeiten mit den Begriffen 'bedingte Wahrscheinlichkeit' und 'Unabhängigkeit' von den kausalen Konnotationen der verwendeten Wörter her. — Solche Assoziationen sind aber nicht von der Hand zu weisen; denn eine zentrale Aufgabe der Stochastik lautet ja, stochastische Zusammenhänge bzw. Auffälligkeiten herauszuarbeiten, damit dann in der jeweiligen Wissenschaft und im Alltag die zugrundeliegenden kausalen Zusammenhänge bzw. Auffälligkeiten untersucht werden können. Die irreführende Wortwahl bei den genannten Begriffen kann man durchaus als Chance verstehen, den Kontrast zwischen stochastischer und kausaler (Un-) Abhängigkeit herauszuarbeiten und damit stochastisches Denken zu fördern.

Trotzdem empfiehlt es sich, anfangs nicht nach der Wahrscheinlichkeit für A unter der Bedingung B zu fragen, sondern nach der Wahrscheinlichkeit für A, wenn man weiß, daß B eingetreten ist (hier zeigt sich wieder der subjektive Zug der Stochastik). Paradebeispiel: "Das Experiment besteht daraus, nacheinander zwei Karten von einem verdeckten Skatkartenstapel zu ziehen. Wie groß ist die Wahrscheinlichkeit, daß die zuerst gezogene Karte ein As ist, unter der Bedingung, daß die zweite Karte dann ein Bube ist?" Diese Frage erscheint dem gesunden Menschenverstand, unterstützt durch die Wörter 'zuerst' und 'dann', als sinnlos, da sie eine Umkehrung des natürlichen Zeitablaufs impliziert. Besser geeignet ist folgender Aufbau der Problemstellung: "Jemand zieht nacheinander zwei Karten von einem verdeckten Skatkartenstapel, schaut sie sich an und teilt mir mit, daß die zweite Karte ein Bube ist. Wie groß ist dann, d.h. mit diesem Wissen, die Wahrscheinlichkeit, daß die zuerst gezogene Karte ein As ist?" (Man beachte, wo das Wörtchen 'dann' jetzt steht.) — Zu der gesamten Problematik s. z.B. (Scholz 1981). — Die Wirkung solcher Formulierungsunterschiede (mit allen möglichen Zwischenstufen) auf die Lernenden zu untersuchen, wäre übrigens eine lohnende Aufgabe für die interpretative Unterrichtsanalyse.

Praktische und theoretische Bedeutung hat der (Un-) Abhängigkeitsbegriff in erster Linie bei mehrstufigen (oder gerade im Hinblick auf ihn mehrstufig gedachten) Experimenten: Wird der Torwart im Laufe der 10 Elfmeter besser? Ändern sich beim Ziehen ohne Zurücklegen die Verhältnisse in der Urne? Usw. — Ohne diese Mehrstufigkeit wirkt dieser Begriff für den Lernenden, zumal bei Zufallsgrößen, gekünstelt und unmotiviert. Für seine mathematische Durchdringung und einen souveränen Umgang (z.B. bei Rechenregeln für Erwartungswerte und Varianzen) bietet sich trotzdem, ganz im Sinne Ziezolds (1982), eine kontextarme Veranschaulichung an: Der Wahrscheinlichkeitsraum W als die Menge aller Punkte eines ebenen Rechtecks, die Ereignisse als Teilmengen (ohne zu problematisieren, welche da nur in Frage kommen), die Wahrscheinlichkeit als sog. "geometrische Wahrscheinlichkeit", d.h. als Flächeninhalt, zwei Ereignisse A_0 und B als Kreisscheiben. Die Wahrscheinlichkeit von A_0 unter der Bedingung B heißt nichts anderes als: Als Ausfälle kommen nicht mehr alle Punkte von W, sondern nur die von B in Frage, d.h. statt W wird nur noch der Wahrscheinlichkeitsraum B betrachtet, insbesondere wird die Wahrscheinlichkeitsmasse auf B konzentriert, und der Anteil von A_0 (genauer: von $A_0 \cap B$) daran ist gesucht.

Der springende Punkt nun ist, daß der Anteil von A_0 (d.h. von $A_0 \cap B$) an B größer, gleich oder kleiner als der Anteil von A_0 an W sein kann. — Fehler, die hier leicht auftreten: Der neue Anteil ist immer gleich oder höher (*mehr*), weil man durch das Wissen über das Eintreten von B *mehr* Information hat. Oder: Der neue Anteil ist größer, weil die Bezugsgröße, der Flächeninhalt von B gegenüber dem von W, kleiner geworden ist. Oder: Der neue Anteil ist so groß wie der alte, weil sich die Wahrscheinlichkeit von A_0 ja nicht ändert. Oder: Der neue Anteil ist kleiner, weil $A_0 \cap B$ kleiner als A_0 ist.

Hier greift nun das *Begünstigen*-Konzept (Scholz 1981, S.91, Borovcnik 1985). Daß bei Einschränkung der Möglichkeiten von W auf B die Wahrscheinlichkeit eines Ereignisses sich vergrößern, sich verkleinern oder gleich bleiben kann, schaut man sich am besten in einem einzigen Bild an mit einem Wahrscheinlichkeitsraum W (mit geometrischer Wahrscheinlichkeit), einem Ereignis B mit $p(B)>0$ und einer Schar von Ereignissen A_t, $t \in [0;1]$, wobei $p(A_t)>0$ konstant ist.

Es wäre noch suggestiver, diese Schar als kontinuierliche 'Bewegung' eines einzigen Ereignisses aufzufassen; aber diese Auffassung kollidiert — strukturell ähnlich wie in der Elementargeometrie — mit dem Verständnis der Punkte als Ausfälle: Da ist nichts, was sich bewegen könnte. Lediglich wenn ich mich als Person einbringe, dann kann ich nacheinander verschiedene Ausfälle (oder Ereignisse) betrachten. Was sich da bewegt, ist die (flüchtige) Eigenschaft betrachtet zu werden. Genau dies wird durch den Begriff der *Schar* zum Ausdruck gebracht. (Hier zeigt sich wieder Wiesemanns (1995) Motiv der Differenzierung von mathematischen und metamathematischen Aktionen).

Es erweist sich als Wohltat, daß man, im zugrundeliegenden Bild der geometrischen Wahrscheinlichkeit und dort natürlich in aller Naivität, *kontinuierliche* Übergänge betrachten kann:

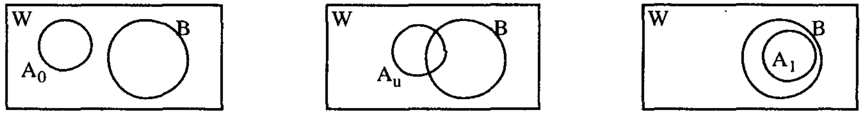

Abb. 3: Die Wahrscheinlichkeit von A_t unter der Bedingung B für $t \in [0;1]$

Man stellt sich also die Schar im stetigen Übergang von A_0 bis A_1 vor. Dabei wächst die bedingte Wahrscheinlichkeit $p_B(A_t)$ kontinuierlich von 0 bis 1, und es gibt eine Stelle u, an der $p_B(A_t)$ gerade gleich $p(A_t)$ ist. Für $t<u$ ist die bedingte Wahrscheinlichkeit kleiner, für $t>u$ ist sie größer. Die Ereignisse A_t werden für $t<u$ durch B benachteiligt, für $t>u$ begünstigt. Der Wert $t=u$ liefert also eine einzelne in einer ganzen Bandbreite von Möglichkeiten.

Es ist vor allem die Additivität der Varianz bei unabhängigen Zufallsgrößen, wo sich Unabhängigkeit nicht nur rechnerisch, sondern auch begrifflich auswirkt.

Es sollte auch die Symmetrie (bezüglich Abhängigkeit) zwischen zwei Ereignissen A und B herausgestellt werden. Qualitativ ergibt sich diese unmittelbar, wenn man in Abb. 3 die Entwicklung von $p_{A_t}(B)$ betrachtet. (Im Unterricht würde man hier A festlassen und eine Schar B_t betrachten, was an der relativen Lage der zu jedem Zeitpunkt t betrachteten beiden Mengen nichts ändern würde, den Lernenden aber, jedenfalls als Fortsetzung der o.a. Überlegungen, zugänglicher wäre.) — Die quantitative Symmetrie entnimmt man folgender Formel, die direkt aus der Definition der bedingten Wahrscheinlichkeit folgt: $\frac{p_B(A)}{p(A)} = \frac{p(A \cap B)}{p(A) \cdot p(B)} = \frac{p_A(B)}{p(B)}$. Zwei Ereignisse begünstigen sich also beide gegenseitig, oder sie benachteiligen sich beide gegenseitig, oder beide sind vom jeweils anderen unabhängig, wobei jedesmal die beiden Grade der Begünstigung bzw. Benachteiligung (der linke und der rechte Bruch in der Formel) gleich sind.

2.6 Zufallsgrößen als Funktionen

In Anbetracht der überragenden Rolle des Funktionsbegriffs im Mathematikunterricht der Sekundarstufen würde man sich wertvoller didaktischer Möglichkeiten begeben, wenn man Zufallsgrößen nicht als Funktionen begreifen würde: Man würde auf die Stärkung des Funktionsbegriffs durch Subsumierung eines eigenständigen Beispiels verzichten, und man würde die begriffliche Klärung des Zufallsgrößenbegriffs mit dem ja bekannten Funktionsbegriff versäumen. Diesen möchte ich hier recht naiv verstehen: Eine Vorschrift, die jedem Element eines Definitionsbereichs einen Wert aus einem Wertebereich zuordnet, veranschaulicht durch Pfeile. Die Begriffe 'Vorschrift', 'zuordnen', 'Definitions-' und 'Wertebereich' werden nicht weiter präzisiert, und nach meinen Erfahrungen kann auf dieser Basis in der Sekundarstufe I (vielleicht nicht bei jedem Lerngruppenniveau) ein tragfähiger Begriff aufgebaut werden, zu dem auch Addition, Subtraktion usw. von Funktionen gehören, wenn diese denn denselben Definitionsbereich und außerdem einen gemeinsamen Wertebereich haben, in dem diese Verknüpfungen vorhanden sind.

Die Schwierigkeiten mit Zufallsgrößen sind weniger solche mit dem Funktionsbegriff, sondern solche mit dem Unterschlagen, Ignorieren, Vergessen des Funktionsbegriffs dabei, insbesondere der notwendigen Existenz eines (bei mehreren arithmetisch zu verknüpfenden Zufallsgrößen: gemeinsamen) Wahrscheinlichkeitsraums als Definitionsbereich.

Sobald eine B(n,p)-verteilte Zufallsgröße X auftritt, so ist dahinter ein n-mal unter gleichen Bedingungen wiederholter Versuch mit zwei Ausgängen mit Wahrscheinlichkeiten p und $q= 1-p$ zu sehen (z.B. n Würfelwürfe mit jedesmal den beiden Ausgängen '6' und 'nicht 6' mit den Wahrscheinlichkeiten 1/6 und 5/6), bzw. es ist *ein* Experiment mit n Stufen zu sehen. Die Ausfälle sind n-tupel aus den Würfelzahlen, und X ordnet jedem Ausfall seine Anzahl von Sechsen zu.

Der zugehörige Erwartungswert ist das arithmetische Mittel der Werte, die die Zufallsgröße annehmen kann, *gewichtet* mit den Wahrscheinlichkeiten ihres Auftretens. — Das gewichtete arithmetische Mittel sollte, spätestens aus der Beschreibenden Statistik bezüglich relativer Häufigkeiten, bekannt sein. Nun geht es zwar um subjektiv (oft: ideal) verteilte Wahrscheinlichkeitsanteile und nicht um objektiv festgestellte Häufigkeitsanteile, aber die gemeinsame Struktur ist offenbar. Allerdings darf man nicht müde werden, den epistemologischen und mathematischen Unterschied zwischen relativer Häufigkeit und Wahrscheinlichkeit zu betonen; andererseits muß man zur Interpretation des Erwartungswertes auf eine Operationalisierung mittels Häufigkeiten zurückgreifen: Wenn ich das Experiment des n-maligen Würfelwurfs 1000mal durchführe, so werden die Anzahlen der Sechsen pro Experiment um $E(X) = \frac{n}{6}$ streuen.

Dieser Wert $E(X) = n/6$ ist so zwar plausibel, aber ihn oder $Var(X)$ direkt über die Definition zu ermitteln, ist etwas umständlich. Die Ermittlung wird sehr einfach, wenn man zusätzlich noch die primitiven Zufallsgrößen X_i, $i=1,\ldots,n$, einführt, die jedem Ausfall die Anzahl der Sechsen beim i-ten seiner n Würfe zuordnen und also nur die Werte 1 (mit Wahrscheinlichkeit 1/6) und 0 (mit Wahrscheinlichkeit 5/6) annehmen können. Alle diese X_i sind von allen anderen unabhängig, sie haben dieselbe Verteilung, nämlich $B(1, \frac{1}{6})$, und ihre Erwartungswerte und Varianzen sind leicht zu errechnen und außerdem plausibel:

$$E(X_i) = p\cdot 1 + q\cdot 0 = p \text{ und } Var(X_i) = p\cdot(1-p)^2 + q\cdot(0-p)^2 = p\cdot q\,.$$

Speziell für $p = 1/2$ ist die mittlere Abweichung $\sqrt{Var}$ vom Erwartungswert auch 1/2 . Wegen $X = \sum_{i=1}^{n} X_i$ ist dann $E(X) = n\cdot p$ und $Var(X) = n\cdot p\cdot q$. — Man kann außerdem die relative Anzahl $\overline{X} = \frac{1}{n}\cdot X$ der Sechsen (bezogen auf die Wurfzahl n), sowie deren Erwartungswert p und Varianz $\frac{p\cdot q}{n}$ betrachten und das Gesetz der großen Zahlen (zunächst exemplarisch) beleuchten.

Wenn nun eine Zufallsgröße X jedem Ausfall einen Wert zuordnet und sie, als Funktion, als Summe (o.ä.) anderer Funktionen dargestellt werden soll, dann muß jede dieser anderen Funktionen auch *jedem Ausfall* einen Wert zuordnen, und bei jedem Ausfall muß die Summe aller dieser Werte den Wert von X ergeben. Diese Forderung ist naheliegend und für Lernende i.a. einsichtig. Erst wenn es um gestufte Experimente geht und Funktionen betrachtet werden, deren Werte bereits durch eine einzige Stufe bestimmt sind, wie die o.a. X_i, die man als Projektionen bezeichnen könnte, geht die Einsicht in die Erfordernis eines gemeinsamen Definitionsbereichs leicht verloren. Großenteils verantwortlich dafür ist der Verzicht auf die mathematische Begrifflichkeit wenigstens in naiver Form, der ja paradoxerweise nicht am Anfang des Lehrgangs, sondern beim Fortschreiten stattfindet.

Der verständige Umgang mit diesen Projektionen X_i ist m.E. ein Prüfstein für den Erfolg eines Stochastiklehrgangs, und zwar unbeschadet der Tatsache, daß sie nicht nur gebraucht werden, um das Rechnen zu erleichtern, sondern daß sie mit ihrem Zusammenspiel zu einer binomialverteilten Zufallsgröße das Gesetz der großen Zahlen verdeutlichen und damit einen wesentlichen Beitrag zur zentralen stochastischen Idee der Bändigung des Zufalls leisten.

Um aber das, um dessen Bändigung es geht, nämlich den Zufall, nicht aus den Augen zu verlieren, müssen Erwartungswert und Varianz auf ihn bezogen werden, etwa: Wenn ich das Experiment des n-fachen Würfelwurfs 1000mal durchführe, dann werden die 1000 Werte, die X dabei animmt, als arithmetisches Mittel etwa den Wert $E(X) = \frac{n}{6}$ haben, und die 1000 Werte, die die Wurzel der quadratischen Abweichung von X von E(X) dabei annimmt, als arithmetisches Mittel etwa den Wert $\sigma(X) = \sqrt{n\cdot\frac{1}{6}\cdot\frac{5}{6}}$. Entsprechend sind die Werte $E(\overline{X}) = \frac{1}{6}$ und $\sigma(\overline{X}) = \sqrt{\frac{1}{6}\cdot\frac{5}{6}/n}$ zu operationalisieren.

Hier ist auf zwei Ebenen jedesmal dasselbe geschehen, nämlich ein Experiment ist oft wiederholt (gedacht) worden. Auf der ersten Ebene wurde die Wiederholung formalisiert und mit in das Experiment und damit in den Wahrscheinlichkeitsraum aufgenommen: Das Experiment besteht nicht aus einem, sondern aus n Würfelwürfen; jeder Wurf ist kein eigenes Experiment, sondern nur eine Experimentstufe. Alle Zufallsgrößen mit ihrer *einmaligen* Realisierung beziehen sich auf das Experiment mit seinen n Stufen. Auf der zweiten Ebene denkt man sich dieses n-stufige Experiment 1000mal durchgeführt, die Zufallsgrößen also 1000mal realisiert. Man könnte auch die zweite Ebene formalisieren und ein Megaexperiment betrachten mit 1000 Stufen, deren jede aus n Teilstufen besteht, mit entsprechenden Zufallsgrößen. Man könnte sich dieses Experiment wieder sehr häufig durchgeführt denken und diese dritte Ebene erneut formalisieren, usw.

Oder man nimmt einmal von der ersten Ebene die Formalisierung weg: Jetzt besteht das Experiment aus einem einzigen Würfelwurf, ein Sachverhalt, der im Curriculum schon vorher analysiert ist. Ist S die Zufallsgröße, die jedem Ausfall die Anzahl der Sechsen zuordnet, also 1 mit Wahrscheinlichkeit $\frac{1}{6}$ und 0 mit

Wahrscheinlichkeit $\frac{5}{6}$, so hat S die Struktur wie oben die X_i. Denkt man sich nun das Experiment sehr oft, nämlich n-mal, wiederholt, so wird das arithmetische Mittel der Werte, die S dabei annimmt, etwa 1/6 sein. Diese Überlegung wird durch die Zufallsgröße $\overline{X}$ in dem oben betrachteten Wahrscheinlichkeitsraum formalisiert, dessen Ausfälle die Folgen von Würfelzahlen der Länge n sind.
Hier wird wieder Wiesemanns (1995) Differenzierung zwischen mathematischen und metamathematischen Aktionen deutlich.

2.7 Anfänge des Hypothesentestens

Die Aufgabe vom Anfang könnte noch folgendermaßen variiert und fortgesetzt werden: "Der Torwart hat einen von 10 Elfmetern gehalten. Es soll die Hypothese überprüft werden, daß seine Haltequote mindestens 0,4 sei." Mit den oben genannten Mitteln käme der durchschnittliche Absolvent eines Stochastikleistungskurses auf die Gleichung $p(X \leq 1) = 4{,}6\%$ und das Ergebnis, die Hypothese auf dem 5%-Signifikanzniveau abzulehnen, eventuell mit der verräterischen Ergänzung, daß die Hypothese mit einer Wahrscheinlichkeit von 95,4% falsch sei.

Auch an dieser Stelle mache ich für das offenbar fehlende Verständnis die Vernachlässigung des Zufallscharakters und eines Ur-Wahrscheinlichkeitsraums einerseits und die zu frühe Konzentration auf genaue Rechnungen und auf Begriffe, die damit verbunden sind (Signifikanzniveau, Vertrauensintervall, Entscheidungsregel usw.), durch Lehrgangs- (Aufgaben-) Konstrukteure, Lehrende und Lernende andererseits verantwortlich. — Gewiß sind diese Begriffe und Rechnungen für Theorie und Praxis erforderlich, aber sie können die folgende Argumentationsfigur des Hypothesentests (als Signifikanztest) verschütten:

Eine Hypothese in Wissenschaft oder Alltag soll überprüft werden. Die Zahl der Fälle, auf die sie sich bezieht, ist sehr groß oder unendlich (tausende von Elfmetern im Training des Torwarts, 60 Mio Wahlberechtigte, alle Glühbirnen einer Produktion, alle denkbaren Brenndauern von Glühbirnen eines Herstellverfahrens usw.). Es wird ein *Zufallsexperiment* (i.d.R. die *zufällige* Auswahl von n Fällen) mit einer Zufallsgröße konstruiert, so daß die Hypothese in eine Aussage über deren Erwartungswert gefaßt werden kann. Wenn nun bei der Realisierung des Experiments die Zufallsgröße einen Wert annimmt, der von dem der Hypothese gemäßen Erwartungswert (im wahrsten Sinne des Wortes:) *unwahrscheinlich* weit weg liegt, dann wird man an der Hypothese *zweifeln*, und zwar umso stärker, je extremer der eingetretene Wert von diesem Erwartungswert entfernt ist.

Zur Verdeutlichung dieser Argumentationsfigur gehört auch, daß man einmal Grundgesamtheit und Stichproben von so kleinem Umfang wählt, daß man die komplette Situation von Hand (mit dem Taschenrechner) darstellen kann und insbesondere ein Gefühl dafür kriegt, für wie wenig Stichproben die Zufallsgröße (z.B. im Fall der Binomialverteilung um mehr als $3 \cdot \sigma$) weit vom Erwartungswert liegen (etwa im Hemden-Beispiel, das Borovcnik, 1981, S.105ff, ausgearbeitet hat). — Es kommt hier darauf an, daß die Lernenden sich Zeit und Muße nehmen, damit die Eindrücke sich stabilisieren können. Sogar wenn der Computer zur Verfügung steht, sollte es wenigstens einmal eine einfache Situation sein, bei der man in der Rolle des explizit alles Wissenden ist, — eine Rolle, die dann als Metapher für die ganze Beurteilende Statistik dienen kann.

3. Stochastik zwischen Mathematik und Anwendungen

'Anwendbarkeit und Nützlichkeit in vielen Bereichen der Gesellschaft' (viel augenfälliger als bei der Linearen Algebra), sowie 'in wesentlichen Teilen leichte Zugänglichkeit' (viel elementarer als bei der Analysis) und trotzdem 'gehaltvoller Stoff' sollten eigentlich schlagkräftige Argumente für eine ausführliche Behandlung der Stochastik in der allgemeinbildenden Schule sein. Daß sie dort nach wie vor nicht heimisch ist, liegt vor allem daran, daß die Sekundarstufenmathematiklehrer in Stochastik durchweg viel weniger intensiv ausgebildet sind als in Analysis und Linearer Algebra, und zwar schon in ihrer eigenen Schulzeit, so daß sie kaum in der Lage sind, die Kluft zwischen der mathematischen Statistik, wenn sie diese denn kennengelernt haben, einerseits und den Anwendungen und der Schulstochastik andererseits zu überbrücken.

Allerdings halte ich die Forderung, die Stochastikausbildung den Mathematikern aus der Hand zu nehmen (s. die Diskussion im College Mathematics Journal; Moore u.a. 1988), auf jeder Stufe des Bildungssystems für verfehlt. Vielfältige Erfahrungen mit Mathematik- und Statistikkursen in den Wirtschafts- und Sozialwissenschaften, Psychologie, Pädagogik usw. sprechen gegen diese Verlagerung der Verantwortung: Da das Interesse von Dozenten und Studenten dort sich regelmäßig nicht auf die Stochastik, sondern auf die jeweilige Wissenschaft richtet, die die stochastischen Methoden ja nur verwendet, verkommen solche Kurse oft zu Rezeptesammlungen.

Auch wenn, was häufig vorkommt, nicht-mathematische Anwender stochastische Begriffe und Methoden entwickeln, so treiben sie Mathematik. Und wenn solche Begriffe und Methoden zu einem Bestandteil der Allgemeinbildung erhoben werden, dann geht es um ihren (durchaus anwendbaren und angewandten) mathematischen, logischen und philosophischen Kern, der außerhalb der Mathematikausbildung nur unzulänglich freigelegt, analysiert und verstanden werden kann.

Die Anwendungen müssen Teil dieser Ausbildung sein; jedoch kann dabei das folgende Dilemma auftreten: Aus Zeitgründen (weil sonst die Anwendungen zu kurz kommen) und aus Gründen der Gewichtigkeit (weil sonst die Bedeutung der Anwendungen zu kurz kommt) wird der zugrundeliegende mathematische Korpus gern auf das scheinbar Notwendige reduziert bzw. im 'Ideal'fall sogar gemeinsam mit den Anwendungen entwickelt, und zwar nur so weit, wie gerade nötig. Für jemand, der über eine solide Begrifflichkeit verfügt, liegen die Zusammenhänge zwischen Mathematik und Anwendungen auf der Hand, und er stellt diese vielleicht mit Leichtigkeit dar. Für den Lernenden, der diese Begrifflichkeit gerade erst erwirbt, basiert das Durchschauen dieser Zusammenhänge jedoch auf harter Arbeit, und zwar über weite Strecken innerhalb der Mathematik.

Wohlgemerkt: Bei dieser Mathematik kann es sich nicht um formalistisches, sinnfreies Tun handeln. Vielmehr müssen ihre Begriffe in Alltagsvorstellungen und -verständnissen gründen, angesiedelt in mehr oder weniger realistischen Situationen mit genuinen Anwendungen bis hin zu spielerischen, märchenhaften Problemen, eine Bandbreite, die vor allem Engel (1987 u.v.a.) mit seinen vielen Beispielen abgedeckt hat. Zur Formung bzw. Stabilisierung von GVV müßten diese Beispiele aber noch nach den in der vorliegenden Arbeit entwickelten Prinzipien unterfüttert werden und könnten dann, auf der Basis von Borovcniks (1984, 1992 u.v.a.) praktischen und theoretischen Überlegungen, die Ausbildung stochastischen Denkens wesentlich befördern.

Literatur

Bender, Peter (1990): Zwei "Zugänge" zum Integral-Begriff? In: mathematica didactica 14, 102–127

Bender, Peter (1991a): Ausbildung von Grundvorstellungen und Grundverständnissen — ein tragendes didaktisches Konzept für den Mathematikunterricht — erläutert an Beispielen aus den Sekundarstufen. In: Helmut Postel, Arnold Kirsch & Werner Blum (Hrsg.): Mathematik lehren und lernen. Festschrift für Heinz Griesel. Hannover: Schroedel, 48–60

Bender, Peter (1991b): Fehlvorstellungen und Fehlverständnisse bei Folgen und Grenzwerten. In: Der mathematische und naturwissenschaftliche Unterricht 44, 238–243

Blum, Werner & Günter Törner (1983): Didaktik der Analysis. Göttingen: Vandenhoeck & Ruprecht

Borovcnik, Manfred (1981): Statistik im Selbststudium. Klagenfurt: Universität

Borovcnik, Manfred (1984): Was bedeuten statistische Aussagen? Wien: Hölder-Pichler-Tempsky

Borovcnik, Manfred (1985): Begünstigen — eine stochastische Intuition. In: Praxis der Mathematik 27, 327–333

Borovcnik, Manfred (1992): Stochastik im Wechselspiel von Intuitionen und Mathematik. Mannheim u.a.: BI Wissenschaftsverlag

Engel, Arthur (1987): Stochastik. Stuttgart: Klett

Hofe, Rudolf vom (1995): Grundvorstellungen mathematischer Inhalte. Heidelberg u.a.: Spektrum

Moore, David S. u.a. (1988): Should Mathematicians Teach Statistics? In: College Mathematics Journal 19, 3–35

Riemer, Wolfgang (1989): Neue Aspekte in der beurteilenden Statistik mit dem Computer und der Regel von Bayes. In. Der Mathematikunterricht 35, Heft 4, 48–62

Scholz, Roland W. (1981): Stochastische Problemaufgaben — Analysen aus didaktischer und psychologischer Perspektive. IDM Materialien und Studien 23. Bielefeld: Universität

Wickmann, Dieter (1990): Bayes-Statistik: Einsicht Gewinnen und Entscheiden bei Unsicherheit. Mannheim: Bibliographisches Institut

Wiesemann, Horst (1995): Mathematische Aktionen. Potsdam: Manuskript

Ziezold, Herbert (1982): Die formale Beschreibung von Zufallsexperimenten durch mengentheoretische Begriffe als Vorstadium stochastischen Denkens. In: Der mathematische und naturwissenschaftliche Unterricht 35, 201–206

Adresse des Autors:

Prof.Dr.Peter Bender
Lothringer Weg 5c
33102 Paderborn