Von ersten stochastischen Erfahrungen mit großen Zahlen bis zum $1/\sqrt{n}$-Gesetz – ein didaktisch orientiertes Stufenkonzept

ROLF BIEHLER, PADERBORN UND ANDREAS PRÖMMEL, GOTHA

Zusammenfassung: *In diesem Aufsatz möchten wir Anregungen dazu geben, bei Lernenden adäquate Grundvorstellungen zum Problemkreis des empirischen Gesetzes der großen Zahlen zu entwickeln. Wir sehen dabei das Experimentieren mit Simulationsumgebungen als zentrales didaktisches Mittel an, um Fehlvorstellungen aufzudecken und angemessenere, fachlich intendierte Vorstellungen, sogenannte Sekundärintuitionen, zu verankern. Dafür haben wir ein unterrichtliches Stufenkonzept von der Klasse 6 bis zur Klasse 12 mit unterrichtlich erprobten Lernumgebungen entwickelt, das verschiedene Aspekte des empirischen Gesetzes der großen Zahlen thematisiert.*

1 Vorbemerkung

Das Stufenkonzept, das wir vorschlagen, kann unterschiedlich auf die Klassenstufen 6 bis 12 verteilt werden. Für die sinnvolle Behandlung von Vertrauensintervallen in der Sekundarstufe II ist die vorgängige Behandlung des $1/\sqrt{n}$-Gesetzes notwendig. Für die stochastische Allgemeinbildung ist es aber wichtig, mit Vorbereitungen wesentlich früher zu beginnen. Das Gesetz der großen Zahlen thematisiert gleichsam den Schluss von der Population auf die Stichprobe, von der Wahrscheinlichkeit auf die relativen Häufigkeiten in Experimenten. Die „Umkehrung“, bei vorliegenden relativen Häufigkeiten Aussagen

Stochastik in der Schule 33 (2013) 2, S. 14–25

über die unbekannte Wahrscheinlichkeit zu machen, kann darauf aufbauen. Es ist naheliegend, ein analoges Stufenkonzept für Vertrauensintervalle parallel zu realisieren, dass dann in Stufe 4 mit der formalen Definition von Konfidenzintervallen endet. Dies auszuführen, würde aber den Rahmen dieses Aufsatzes sprengen

2 Eine Aufgabe zum Einstieg

Wir stellen zunächst eine Aufgabe vor, die wir zu Beginn eines Unterrichtsversuches in einer 8. Klasse zur Einführung in die Wahrscheinlichkeitsrechnung eingesetzt haben (Prömmel & Biehler 2008). Dann beschäftigen wir uns mit üblichen Fehlvorstellungen und den zu entwickelnden adäquaten Grundvorstellungen, bevor wir ein Stufenkonzept zum empirischen Gesetz der großen Zahlen vorschlagen.

Mit der Aufgabe (Abb. 1) sollte untersucht werden, ob Lernende den Einfluss der Stichprobengröße bei diesem konkreten Zufallsexperiment beachten und ggf. mit ihren intuitiv vorhandenen Vorstellungen zum empirischen Gesetz der großen Zahlen verknüpfen und wie sich die Vorstellungen der Lernenden durch den Stochastikunterricht entwickeln. In unserem Unterrichtsversuch haben wir intensiv mit Simulationen gearbeitet und das empirische Gesetz der großen Zahlen in verschiedenen Varianten erarbeitet. Schwierig war es, eine Aufgabe zu konstruieren, die Vorwissen ohne Unterricht erfassen kann: Kreisdiagramme bei zwei Ausprägungen erlauben einen intuitiven Vergleich von Anteilen. Die Aufgabe erfordert,

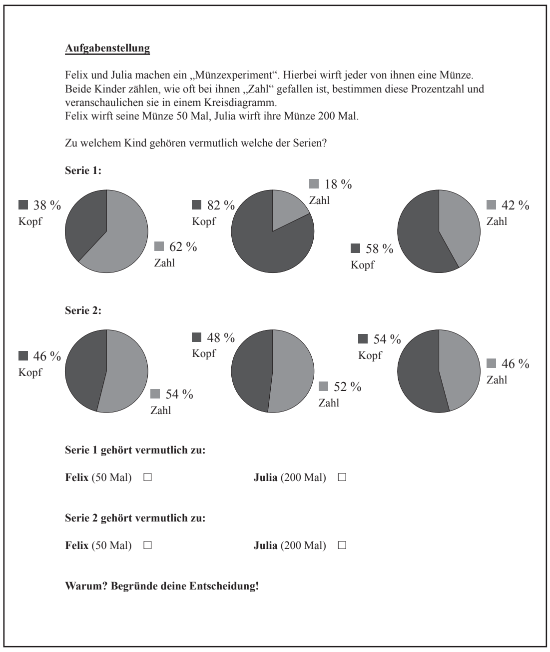


Abb. 1: Testaufgabe Münzexperiment zum Gesetz der großen Zahlen (vgl. Grosch 2009, S. 311)

dass Schüler erkennen, dass die Schwankungen der relativen Häufigkeit bei $n = 200$ geringer sind als bei $n = 50$. Die Ergebnisse zeigen, dass sowohl bei den richtigen Antworten als auch bei den Begründungen deutliche Zunahmen vom Vortest zum Nachtest zu verzeichnen sind (Tab. 1). (Aus den gepaarten Daten entnehmen wir, dass sich 6 % der Schüler gegenüber dem Vortest verschlechtern.)

	Vortest	**Nachtest**
Anteil korrekter Antworten (in %)	54	78
Anteil angemessener Begründungen (in %)	9	30
Anteil teilweise richtiger Begründungen (in %)	6	28

Tab. 1: Ergebnisse Münzexperiment zum Gesetz der großen Zahlen, $n = 67$ Schüler (Grosch 2009, S. 315 ff.)

Die Schülerbegründungen von Vor- und Nachtest wurden für weitere Detailanalysen kategorisiert. Als Ergebnis zeigt sich, dass die Schülerbegründungen über die Stabilisierung relativer Häufigkeiten (in den Kategorien *Begründung mittels Stabilisierung* und *ungefähre Begründung mittels Stabilisierung* zusammen) von 15 % im Vortest auf 58 % im Nachtest zugenommen hatten, wobei die gepaarten Daten zeigen, dass die Schüler, die bereits im Vortest (teilweise) angemessene Begründungen hatten, sich nicht verschlechtern.

Der hohe Anteil an nicht angemessenen Begründungen im Vortest korrespondiert mit dem geringen Anteil von nur 54 % richtigen Antworten, denn 50 % richtig entspräche der Quote bei reinem Raten. Es liegt daher nahe, den Zuwachs bei den richtigen Antworten und in der Begründungsqualität auf den unterrichtlichen Einsatz von Simulationen zur Veranschaulichung der Stabilisierung relativer Häufigkeiten für große Versuchsanzahlen bei Verteilungen (empirisches Gesetz der großen Zahlen) zurückzuführen. Allerdings haben wir keine Vergleichsgruppe mit „normalem" Stochastikunterricht untersucht (Prömmel & Biehler 2008).

Wir haben in den Schülerbegründungen beim **Vortest** auch typische, nicht tragfähige Primärintuitionen wiedergefunden, wie die Negierung des Einflusses des Stichprobenumfanges und den Bezug zu absoluten Häufigkeiten, wie die nachfolgenden Beispiele zeigen:

„Eigentlich ist es Zufall, man kann das nicht ausrechnen!!"

„Weil Julia mehr Versuche hat und deswegen ist die Wahrscheinlichkeit auch größer, dass sie mehrmals die gleiche Seite wirft."

„Die Wahrscheinlichkeit, öfter „Zahl" zu werfen, ist höher, wenn man öfter wirft."

Aber auch tragfähige Vorstellungen, wie der Bezug auf größere Unterschiede bei geringem Stichprobenumfang und der Bezug auf den Ausgleich bei großem Stichprobenumfang, lassen sich vereinzelt bereits im Vortest bei den Schülerbegründungen finden, wie die nachfolgenden Beispiele zeigen.

„Bei mehr Würfen ist die Wahrscheinlichkeit größer, dass es ausgeglichener ist als bei nur 50 Würfen."

„Je öfter man eine Münze wirft desto ausgeglichener wird es!"

Einige der als angemessen gewerteten Schülerbegründungen aus dem **Nachtest** geben wir zur Illustration an (Grosch 2009, S. 321 ff.):

„(Weil) umso öfter man wirft (mehr Fälle), umso weniger Schwankungen gibt es (also umso stabiler ist das Ergebnis)."

„Je öfter man wirft, desto weniger verändert sich das Diagramm."

„Wenn Felix 50-mal wirft, kann es sein, dass sich die Zahlen noch nicht ausgeglichen haben."

„Weil, wenn man 200-mal wirft, man ein gleichmäßigeres Ergebnis bekommt. Wenn man nur 50-mal wirft, ist das Ergebnis noch ganz durcheinander und es kann sich mit jedem Mal wieder etwas ändern."

„Die Wahrscheinlichkeit ist gleich hoch, Kopf und Zahl zu werfen. Also denke ich, dass Serie 2 Julia ist, weil sie viel ausgeglichener ist (Gesetz der großen Zahlen)."

„Wir hatten so etwas Ähnliches ja mal im Unterricht nur mit Würfeln. Da wurde das Ergebnis am Ende immer deutlicher. Deswegen habe ich bei Serie 2 auf Julia getippt, weil sie öfter geworfen hat."

3 Problemlage

Das empirische Gesetz der großen Zahlen, wie es in Schulbüchern oft formuliert wird, besagt, dass sich die empirische relative Häufigkeit eines Ergebnisses bei einem Zufallsversuch bei großen Versuchsanzahlen „immer mehr" der Wahrscheinlichkeit dieses Ergebnisses annähert. Diese Einsicht scheint zum „gesunden Menschenverstand" zu gehören. Davon ging bereits Jakob Bernoulli aus, dem wir den ersten Beweis des mathematischen Gesetzes der großen Zahlen verdanken. Selbst „recht einfältigen" Menschen sei es klar – so Bernoulli –, dass man eine große Anzahl von Beobachtungen braucht, um verlässlich zu urteilen, und dass „je mehr diesbezügliche Beobachtungen man macht, um so weniger Gefahr läuft, von der Wahrheit abzuirren" (Gigerenzer et al., 1999, S. 29). Bernoulli startete historisch ein Forschungsprogramm, in dem die Gesetze der großen Zahlen immer besser modelliert und untersucht wurden (vgl. Biehler & Steinbring, 1982). Unterricht sollte ebenso nicht beim gesunden Menschenverstand stehen bleiben. Dazu zeigen wir Wege auf, die die Simulation elementarisierend nutzen.

Das Gesetz der kleinen und der großen Zahlen

Unser Vortest hat gezeigt, dass viele Schüler schwache Intuitionen zum Gesetz der großen Zahlen haben, auch nach dem Unterrichtsversuch geben nur ca. 80 % die richtige Antwort. Empirische Untersuchungen mit Erwachsenen zeigen bestärkend, dass die entwickelten Vorstellungen häufig unzureichend sind. Kahneman & Tversky (1971, 1982) haben in ihren Untersuchungen herausgefunden, dass viele Menschen fälschlicherweise an ein „Gesetz der kleinen Zahlen" glauben. Das heißt, sie haben völlig unzureichende Vorstellungen davon, bei welchem Stichprobenumfang n die relativen Häufigkeiten in der Nähe der Wahrscheinlichkeit zu erwarten sind und was überhaupt „nah" heißt. Freudenthal (1972) hat schon vor langer Zeit darauf hingewiesen, dass diese falschen Vorstellungen zumindest teilweise durch unzureichenden Schulunterricht entstehen. Diagramme und Aufgaben in Schulbüchern suggerieren, dass oft schon bei relativ geringem Stichprobenumfang von wenigen Hundert Versuchen die Wahrscheinlichkeit durch die relative Häufigkeit praktisch erreicht wird. Ein Blick in neuere Schulbücher der Sekundarstufe I bestätigt die Nachhaltigkeit dieses Problems. Schüler experimentieren mit Zufallsgeräten und sollen aus recht wenigen Versuchen verlässliche Schlüsse über Wahrscheinlichkeiten ziehen. Größere Stichprobenumfänge werden vermieden, vermutlich weil dies den Wert eigenen Experimentierens schmälern würde.

Variation in Abhängigkeit vom Stichprobenumfang *n* – sample size effect

Für die stochastische Allgemeinbildung ist die Vorstellung, dass sich „auf lange Sicht" die relativen Häufigkeiten der Wahrscheinlichkeit nähern, allein nicht ausreichend. Man hat es immer mit einem endlichen Stichprobenumfang zu tun, auf dem man

Schlüsse aufbauen muss. Man muss also Grundvorstellungen zu folgenden zwei Aspekten entwickeln:

- Wie stark variieren die relativen Häufigkeiten um die Wahrscheinlichkeit bei einem festen Stichprobenumfang n?
- Wie nimmt diese Variation mit wachsendem Stichprobenumfang n ab?

Die Vorstellungen, die hierzu entwickelt werden müssen, sind bereits komplexer: zu jedem festen Stichprobenumfang n muss man sich eine Verteilung der relativen Häufigkeiten vorstellen, deren „Streuung" dann mit wachsendem n abnimmt. Hierzu sind Intuitionen noch schwächer entwickelt. Ein Startpunkt für die Entwicklung solcher Intuitionen können Diagramme wie die aus unserer Eingangsaufgabe sein. Dort werden zwei Stichprobenumfänge ($n = 50$ und $n = 200$) bei einer Wiederholungszahl $N = 3$ verglichen. Die Schwächen in der intuitiven Beurteilung des Einflusses des Stichprobenumfanges auf die Gestalt der Verteilung wurden in verschiedenen psychologischen Studien nachgewiesen (Sedlmeier & Gigerenzer 1997). Dabei wurden in verschiedenen Varianten Aufgaben folgenden Typs verwendet (siehe Kasten).

Aufgabe (maternity ward problem)

In einem großen Krankenhaus werden durchschnittlich jede Woche etwa 90 Kinder geboren. In einem kleinen Krankenhaus werden durchschnittlich jede Woche etwa 40 Kinder geboren. An welchem Krankenhaus ist es wahrscheinlicher, dass in einer Woche mehr als 65 % der geborenen Kinder Jungen sind?

- im großen Krankenhaus
- im kleinen Krankenhaus
- in beiden gleichwahrscheinlich

Begründung:

Solch eine Aufgabe kann man mit der Vorstellung beantworten, dass die relative Häufigkeit bei größerem n „wahrscheinlich näher" an der theoretischen Wahrscheinlichkeit liegt als bei geringerem n. Tragfähiger wäre eine Lösung mittels Vorstellungen zur Verteilung der relativen Häufigkeit um die theoretische Wahrscheinlichkeit bei den Stichprobenumfängen 40 und 90.

In der GESIM-Studie (Biehler & Prömmel 2010 und Prömmel 2012) haben wir Schüler der Klasse 12 zu Beginn des Stochastikunterrichts mit Hilfe von Simulationen beim Aufbau adäquater Verteilungsvorstellungen unterstützt. So wurden u. a. Aufgaben zum sample size effect im Unterricht in einem prozessorientierten, mehrschrittigen Arbeitsblatt thematisiert: Nach einer offline-Arbeitsphase mit einer intuitiven Schätzung und dem Anfertigen von Verteilungsskizzen haben die Schüler dann am Computer durch ihre eigenen Simulationsergebnisse ihre eingangs getroffenen Annahmen selbständig reflektieren und ggf. revidieren können.

Die Aufgabe aus dem Kasten haben wir in einem Vortest und in einem Nachtest eingesetzt. Eine Lösungsrate von etwa 28 % im Vortest und etwa 77 % im Nachtest für die Antworten und von 18 % im Vortest und von 59 % im Nachtest für die angemessenen Begründungen zeigen, dass geeignete Interventionen im Unterricht demnach dazu beitragen können, fachlich tragfähige Sekundärintuitionen zum Einfluss der Stichprobengröße aufzubauen (vgl. Prömmel, 2012, S. 487 ff.).

In der Mathematikdidaktik wurden Vorschläge für eine unterrichtliche Behandlung dieses Problemkreises vor allem von Riemer (1991) unter dem Stichwort „$1/\sqrt{n}$-Gesetz" entwickelt, auch optional für die Sekundarstufe I. Wir greifen diese Ideen auf, sind aber überzeugt, dass die quantitative Präzisierung in Vorstufen vorbereitet werden muss und dass noch stärker als bei Riemer Verteilungsvorstellungen entwickelt werden müssen.

Ein Gesetz der großen Zahlen für absolute Häufigkeiten?

Eine weitere verbreitete Fehlvorstellung ist, dass ein Gesetz der großen Zahlen auch für absolute Häufigkeiten gilt, etwa in der folgenden Form: Bei einem Münzwurf, in dem die Wahrscheinlichkeit für Kopf $p = 0{,}5$ ist, erwarten wir bei n Würfen etwa $n/2$-mal Kopf. Je größer n, desto näher kann man die absoluten Häufigkeiten bei $n/2$ erwarten. Diese Aussage ist falsch, es variieren die absoluten Häufigkeiten immer stärker um den Erwartungswert. Nur die relativen Abweichungen werden geringer. Diese Problematik ist in didaktischen Überlegungen kaum aufgegriffen worden. Allerdings widmen sich bereits von Harten und Steinbring (1984, S. 49 ff.) diesen Fehlvorstellungen, indem sie einen fiktiven Dialog in Anlehnung an Freedman, Pisani, & Purves (1978) als didaktisches Mittel einsetzen.

4 Ein unterrichtliches Stufenkonzept

Überblick über das Stufenkonzept

Tab. 2 zeigt unseren Vorschlag für eine stufenweise unterrichtliche Entwicklung von tragfähigen Sekundärintuitionen zum empirischen Gesetz der großen Zahlen. Die in diesem Artikel genannten Lernumge-

Stufe	Thema	Relative Häufigkeiten	Absolute Häufigkeiten	Lern-umgebungen
0	Schwankungsvergleiche für zwei verschiedene n	Banddiagramme	Direkter Vergleich der Abweichungen	LU 0
1	Entwicklung für wachsendes n	Demonstration der Stabilisierung	Demonstration der zunehmenden Abweichung vom Erwartungswert	LU 1 LU 2 LU 3
2	Variation in Abhängigkeit vom Stichprobenumfang bei zwei verschiedenen n's; von Einzelversuchen bis zur Verteilung (Wiederholungsanzahl $N = 5, 25, 1000$)	Punkt- und Verteilungsgraphiken	Punkt- und Verteilungsgraphiken	LU 4 LU 5 LU 6
3	Quantitative Präzisierung der Veränderung der Streuung in Abhängigkeit von n	$1/\sqrt{n}$-Gesetz Faustregeln	$\sqrt{n}$-Gesetz Faustregeln	LU 7
4	Theoretische Herleitung der Streuungsveränderung mit der Binomialverteilung	Anteil der Erfolge Y: $\sigma(Y) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$ $P(Y \in [p - 1{,}96\sigma(Y), p + 1{,}96\sigma(Y)]) \approx 95\,\%$	Anzahl der Erfolge X: $\sigma(X) = \sqrt{np(1-p)}$ $P(X \in [np - 1{,}96\sigma(X), np + 1{,}96\sigma(X)]) \approx 95\,\%$	

Tab. 2: Stufenkonzept für den Unterricht zum Gesetz der großen Zahlen

bungen LU 0 bis LU 7 finden Sie als Zusatzmaterial zum Aufsatz im Online-Archiv der Zeitschrift auf www.stochastik-in-der-schule.de. Anregungen für konkrete problemorientierte unterrichtliche Umsetzungen finden Sie ferner in Kapitel 9.2 und Kapitel 11.2 von Biehler et al. (2011). Alle Lernumgebungen sind mit der Software FATHOM erstellt und daher auch ohne spezifische Programmierkenntnisse für den Nutzer leicht nachvollziehbar bzw. direkt erweiterbar. Die Software Fathom zeichnet sich insbesondere durch die einfache Realisierung von Modellen und Simulationen stochastischer Sachverhalte mittels interaktiver und dynamischer Objekte aus. Sie bietet daher viele Vorteile für den Stochastikunterricht in den Sekundarstufen 1 und 2. Unter einer Lernumgebung verstehen wir eine softwaregestützte interaktive Arbeitsumgebung zur Erschließung neuer Sachverhalte. Solche Lernumgebungen können sowohl als Lehrerdemonstration im Plenum als auch für Schülerarbeitsphasen am Computer genutzt werden.

In den Stufen 0 bis 3 werden Gesetzmäßigkeiten durch Beobachtungen gewonnen und haben damit den Charakter empirischer Gesetze. Diese werden auf der Modellebende – Stufe 4 – durch mathematische Gesetze erklärt, verallgemeinert und abgesichert.

Als Rechtfertigung für die Einbeziehung von Simulationen in den Stochastikunterricht lassen sich drei Perspektiven angeben (Biehler & Maxara 2007):

- Die Durchführung realer Experimente ist zeitintensiv und meist nur in geringer Versuchsanzahl durchzuführen. Verlässliche Aussagen sind dadurch kaum möglich. Eine computergestützte Simulation kann helfen, stochastische Konzepte durch die gegenständliche Repräsentation von Zufallsexperimenten erfahrbar zu machen.
- Im Wechselspiel mit theoretischen Zugängen können Ergebnisse sowohl gestützt als auch überprüft werden. So kann eine Simulation Ergebnisse liefern, „die einen heuristischen Wert für die Suche nach analytischen Lösungen haben können“ (Biehler & Maxara 2007, S. 46).
- Ist die analytische Lösung eines Problems zu komplex oder nicht bekannt, dann werden in der Praxis häufig Simulationen zur Problemlösung eingesetzt. Mit dem Einsatz von Computern ist die Simulation damit auch eine anerkannte Methode zur Forschung und Problemlösung geworden.

Stufe 0 – Schwankungsvergleiche für zwei verschiedene *n*

In der Lernumgebung **LU 0** (Abb. 2) sind in einem Banddiagramm die prozentualen Anteile von Wappen (W) und Zahl (Z) bei verschieden langen Münzwurfserien in Anlehnung an die Einstiegsaufgabe

(Abb. 1) dargestellt. Bei mehrfacher Wiederholung des Zufallsversuches können Schüler die größeren Schwankungen der prozentualen Anteile bei kleinerem Stichprobenumfang (Felix) direkt in der Graphik erkennen.

Hußmann & Prediger (2009) und Schnell (2011) haben mit dem Wettkönig eine Lernumgebung mit Spielkontext entwickelt, mit der ähnliche Erfahrungen vermittelt werden sollen. In unserem Stufenkonzept gehen wir offensiver und expliziter von Beginn an mit den möglichen Fehlvorstellungen zu den Schwankungen der absoluten Häufigkeiten um. In der Lernumgebung **LU 0** kann simultan das geringer werdende Schwanken der relativen Häufigkeiten und das stärker werdende Schwanken der absoluten Häufigkeiten für einen größeren Stichprobenumfang demonstriert werden, letzteres indem man Abweichungen von dem Erwartungswert $n/2$ mit Zeigerdiagrammen visualisiert.

Wiederholt man mit **Strg+Y** die Simulation der Münzwurfserien, kann man als Lernender leicht erkennen, dass die Anteile von Wappen und Zahl bei Julia ausgeglichener sind als bei Felix. Zugleich kann deutlich werden, dass die absoluten Abweichungen bei Julia stärker schwanken als bei Felix.

Stufe 1 – Entwicklung für wachsendes *n*

Eine Standardumgebung sind Graphiken mit Bahnkurven (Trajektorien) zur Entwicklung der relativen und absoluten Häufigkeiten beim einfachen Münz-

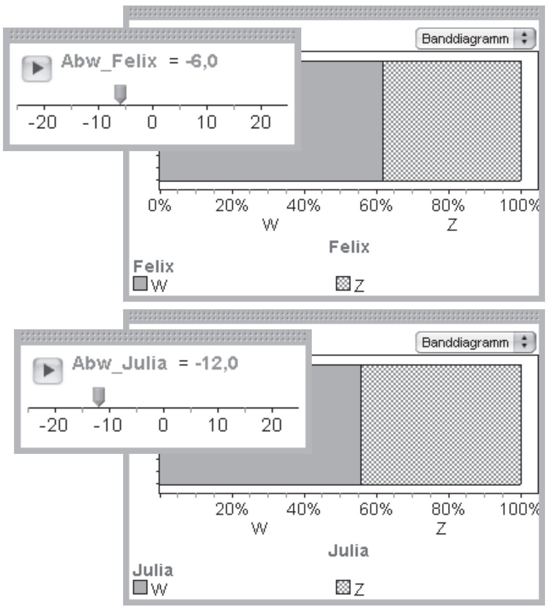


Abb. 2: Lernumgebung LU 0 – simultane Visualisierung der Schwankungen der relativen Häufigkeit sowie der absoluten Abweichungen vom Erwartungswert $n/2$ bei $n = 50$ (Felix, oben) und $n = 200$ (Julia, unten).

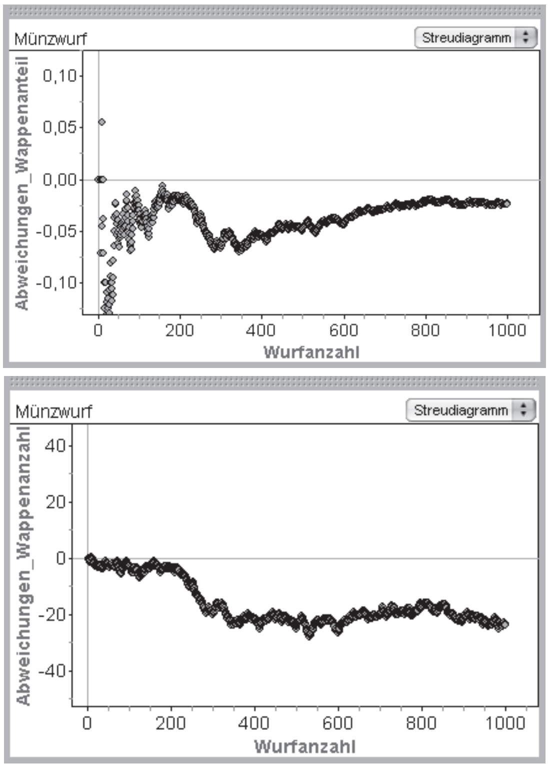


Abb. 3: Lernumgebung LU 2 – relative (oben) und absolute (unten) Abweichungen vom Erwartungswert

wurf. Während in **LU 1** (ohne Bild) der Näherungsprozess der relativen Häufigkeiten in verschiedenen Zoom-Stufen beobachtet werden kann, können in **LU 2** die relativen und absoluten Abweichungen vom Erwartungswert simultan erforscht werden (Abb. 3). Gegenüber einem einzigen statischen Bild hat die Lernumgebung den prinzipiellen Vorteil, dass bei Erneuerung des Zufalls „immer wieder“ die Stabilisierung zu beobachten ist und zugleich die Unterschiedlichkeit der Stabilisierungen bei Wiederholung sichtbar wird.

In **LU 2** wird die Stabilisierung für einzelne Ergebnisse visualisiert. Um von Vorneherein den Verteilungsbegriff in den Fokus zu nehmen, ist es sinnvoll, die Stabilisierung der Häufigkeitsverteilung insgesamt zu demonstrieren. Dazu kann man Histogramme der empirischen Häufigkeitsverteilung nehmen und deren Veränderung und Stabilisierung bei wachsendem *n* beobachten. Diese Idee haben wir in LU 3 umgesetzt, indem wir die Schwankungen des Histogramms für 3 feste *n* vergleichend beobachten.

In **LU 3** zum Spiel „Differenz trifft“ besteht die Möglichkeit, simultan anhand von drei Stichprobenumfängen ($n = 50$, 200, 1000) durch mehrfache Wiederholung der Simulation zu erleben, wie sich bei größerem *n* die Verteilung weniger ändert als bei kleinem *n* (vgl. Abb. 4).

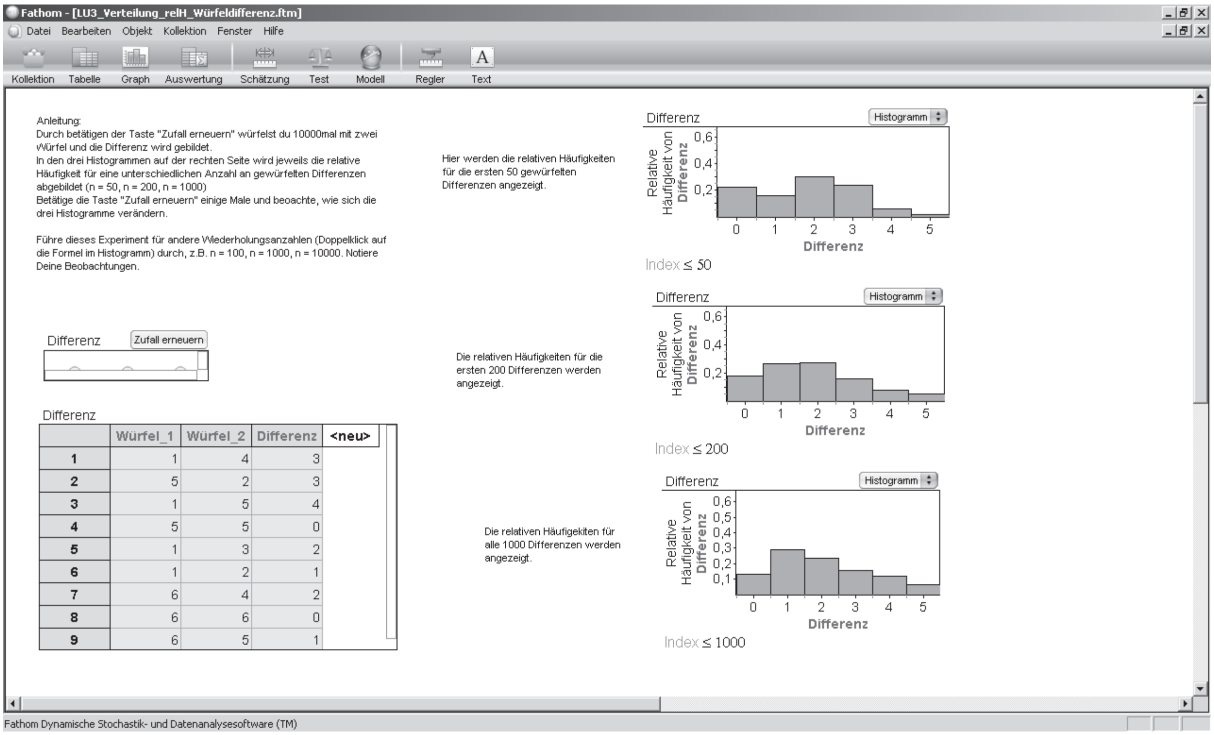


Abb. 4: Lernumgebung LU 3 – Gesetz der großen Zahlen, Stabilisierung der Verteilung für große *n*

Stufe 2 – Variation bei wachsendem *n*

Eine Übergangslernumgebung **LU 4**, die sowohl Trajektorien für wachsendes *n*, wie auch die Variation bei festen *n* enthält, erhalten wir dadurch, dass wir mehrere Trajektorien ($N = 5$) einzeichnen. In Abb. 5 sieht man die tendenziell abnehmende Variation für wachsendes *n*, die sich bei Erneuerung des Zufalls immer wieder so ähnlich ergibt.

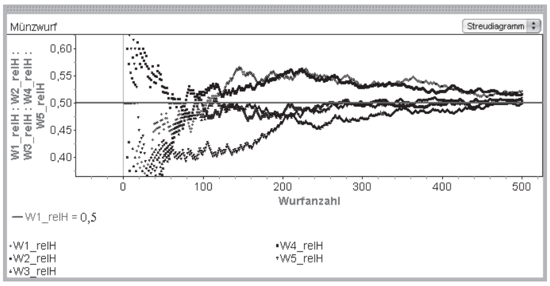


Abb. 5: Lernumgebung LU 4 – Multiple Trajektorien, $N = 5$

Mit der Lernumgebung **LU 4** kann man auch für kleinere Zahlen, z. B. für $n = 50$ und $n = 200$, die jeweils 5 relativen Häufigkeiten aus der zugehörigen Tabelle oder der Grafik ablesen. Man kann die Simulation 5-mal wiederholen und die dann erhaltenen 25 Werte in einer neuen Datentabelle notieren und vergleichen. Eine solche Datentabelle könnte wie in Abb. 6 oben aussehen.

In Abb. 6 unten wurden diese Daten visualisiert. Man sieht qualitativ, dass sich die Werte bei der höheren Stichprobengröße enger um die Wahrscheinlichkeit von $p = 0{,}5$ gruppieren.

Welche Ergebnisse sind eigentlich für die relativen Häufigkeiten von Wappen theoretisch möglich? Für den 50fachen Münzwurf sind es alle Brüche von

Münzexperiment

	Anteil_W_50	Anteil_W_200
1	0,54	0,475
2	0,46	0,525
3	0,5	0,465
4	0,52	0,56
5	0,4	0,57
6	0,52	0,54
7	0,6	0,495
8	0,48	0,48
9	0,6	0,525
10	0,64	0,57

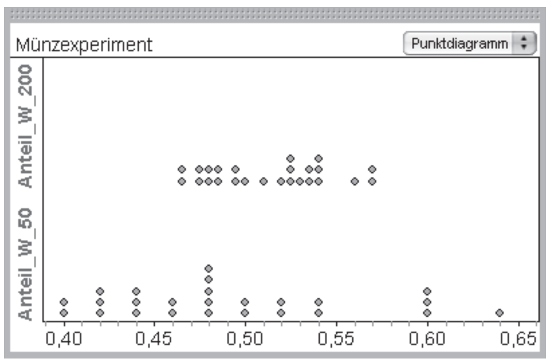
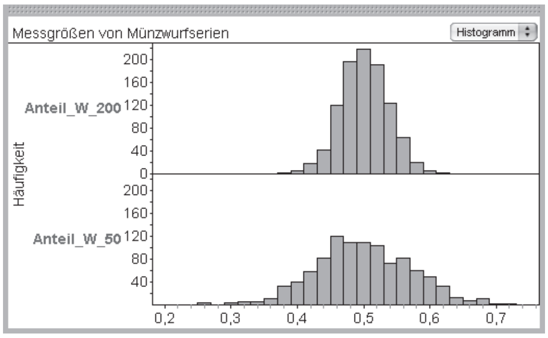


Abb. 6: Datentabelle (oben) und Punktdiagramm (unten) zu Anteil an Wappen für $n = 50$ bzw. $n = 200$ (Wiederholungsanzahl $N = 25$)

0/50, 1/50, ..., 50/50, beim 200fachen Münzwurf alle Brüche von *0/200, 1/200, ..., 200/200*. Aber die Wahrscheinlichkeit, beim 50-fachen Münzwurf genau 50-mal Wappen zu erhalten, ist $(1/2)^{50}$. Es ist eine wesentliche Einsicht dieser Stufe, dass man mit absoluter Sicherheit nur die Vorhersage machen kann, dass die relativen Häufigkeiten zwischen 0 und 1 liegen.

Für praktische Zwecke ist es nun interessant, sich auf Bereiche um ½ zu konzentrieren, in welchen sich die relative Häufigkeit mit einer Wahrscheinlichkeit von 95 % oder 99 % aufhält. Dazu sind nicht nur $N = 25$, sondern sehr viele Durchgänge des 50-fachen und des 200-fachen Münzwurfs zu realisieren und dann zu ermitteln, wo sich die mittleren 95 % (99 %) der Ergebnisse befinden. Grundgedanke der Lernumgebung **LU 5** ist, dass die in Abb. 6 in einer Tabelle eingetragenen Werte nun automatisch gesammelt werden, indem man den 50-fachen Münzwurf wie den 200-fachen Münzwurf jeweils $N = 1000$-mal wiederholen lässt (Abb. 7). Wir sehen wieder die höhere Konzentration um die 0,5 beim 200-fachen Münzwurf. Es ist wichtig, die beiden Teilgrafiken in Abb. 7 identisch zu skalieren, um einen direkten visuellen Vergleich zu ermöglichen.



Messgrößen von Münzwurfserien

	Anteil_W_50	Anteil_W_200
	0,37	0,43
	0,64	0,5675
	0,27	0,1375

S1 = Perzentil (2,5; ?)
S2 = Perzentil (97,5; ?)
S3 = Perzentil (97,5; ?) – Perzentil (2,5; ?)

Abb. 7: LU 5 – Histogramm und Auswertungstabelle zu Anteil an Wappen für n = 50 bzw. n = 200 (Wiederholungsanzahl N = 1000)

Uns interessiert nun, in welchem Bereich um 0,5 herum sich die mittleren 95 % der Wappenanteile bei 1000 Stichproben der beiden Münzwurfserien verteilen. Dazu könnten wir jeweils die 25 kleinsten und die 25 größten der (1000) Werte „abteilen" und das resultierende Intervall der mittleren 95 % der Daten betrachten. Wir halten es für sehr wichtig, dass dieses Intervall im Kontext der gesamten Verteilungsgraphik gesehen wird, welche auch qualitativ die abnehmende Streuung demonstriert. Hier sollten die graphischen Möglichkeiten moderner Software ausgenutzt werden, die bei der Entwicklung der Vorschläge von Riemer (1991) so noch nicht zur Verfügung standen.

Das Intervall kann man mittels Software mithilfe des 2,5 %- und des 97,5 %-Perzentils der Häufigkeitsverteilung bestimmen. Da man im Allgemeinen nicht exakt 2,5 % abteilen kann, muss man das p %-Perzentil $q(p)$ geeignet definieren, so dass höchstens p % der Daten kleiner als $q(p)$ sind, aber mindestens p % größer gleich $q(p)$ sind (für Details der Definition siehe Biehler et al. 2006, S. 62 f.).

Es ist zweckmäßig, für die ermittelten Intervalle einen Terminus einzuführen. Üblich ist es, vom 95 %- bzw. 99 %-*Prognoseintervall* zu sprechen. Die Breite des Intervalls kann man als ein Streuungsmaß ansehen und sie z. B. anschaulich mit 95 %-*Streubreite* bezeichnen.

Geht man bei der Bestimmung so wie oben vor, wird noch nicht der voraussetzungsvolle Begriff der σ-Umgebung benötigt, mit dem in der Oberstufe Prognoseintervalle berechnet werden (vgl. Biehler et al. 2011, Kap. 11).

Stufe 3 – $1/\sqrt{n}$-Gesetz empirisch

Eine neue Stufe kann mit der Frage eingeleitet werden, wie die 95 %-Streubreite vom Stichprobenumfang abhängt. Im Beispiel (vgl. Tabelle in Abb. 7 unten) hat sich bei Vervierfachung des Stichprobenumfangs die 95 %-Streubreite von 0,27 auf 0,1375 reduziert, also in etwa halbiert. An weiteren Beispielen kann man überprüfen, ob diese Gesetzmäßigkeit auch von allgemeinerer Gültigkeit ist.

Das haben wir z. B. im Rahmen der GESIM-Studie in Klasse 12 getan (vgl. Biehler & Prömmel, 2010 und Prömmel 2012). In der Lernumgebung **LU 7**, die wir auch dort verwendet haben, wurde der Münzwurf für verschiedene Stichprobenumfänge n simuliert (Wiederholungsanzahl jeweils $N = 2000$). In Abb. 8 links sind die mittleren 95 %-Prognoseintervalle der relativen Häufigkeit für den Wappenanteil (Ant_W) für wachsendes n in einem Punktdiagramm dargestellt. Die jeweiligen Intervallbreiten wurden in einer Tabelle zusammen mit dem Stichprobenumfang n notiert (Tabelle oben rechts) und die Abhängigkeit von n in einem Streudiagramm unten rechts geplottet.

Die Aussage, dass eine Vervierfachung des Stichprobenumfangs die 95 %-Streubreite halbiert, können die

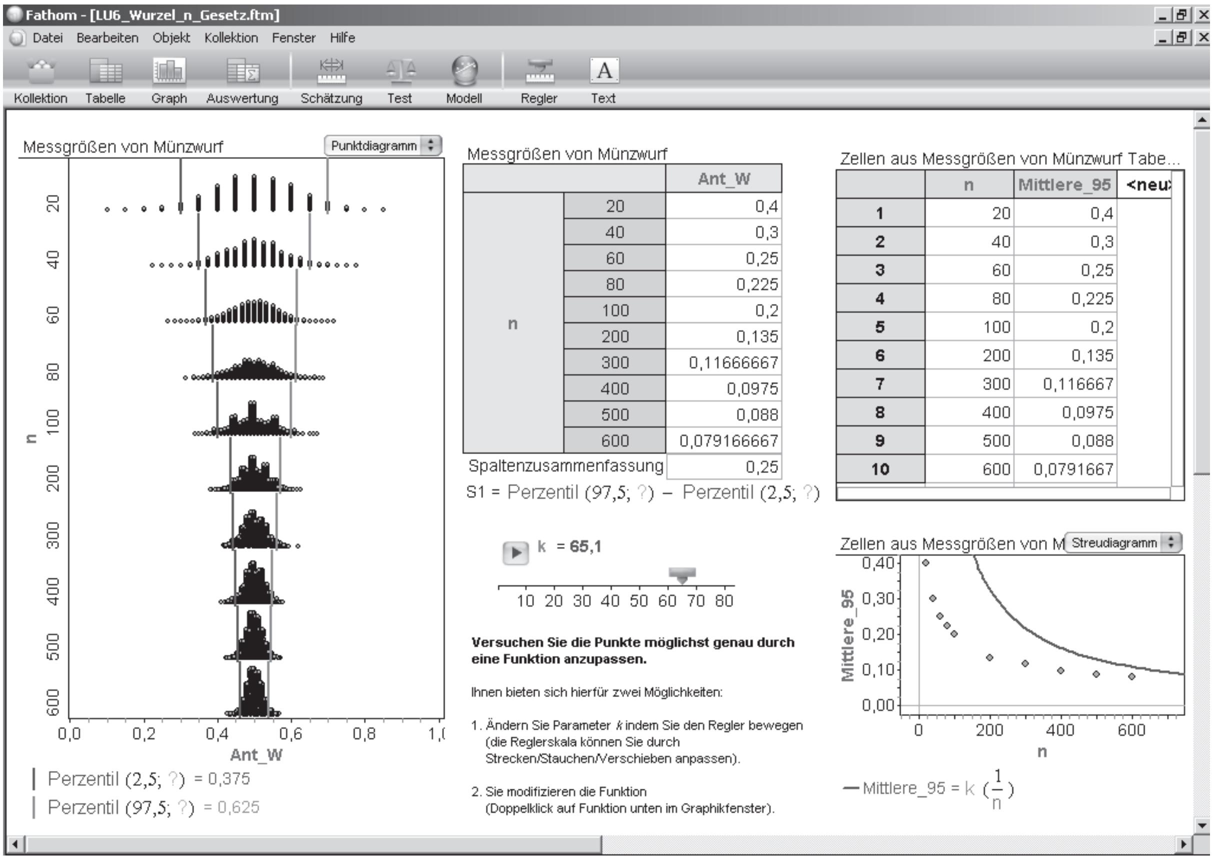


Abb. 8: LU 7 – Lernumgebung zum $1/\sqrt{n}$-Gesetz mit Parameter k = 65,1

Schüler mit dieser Lernumgebung für weitere Stichprobengrößen überprüfen. Die Frage ist, von welcher mathematischen Funktion diese Kovariationseigenschaft (vgl. Leuders & Prediger 2005) erfüllt wird. Ein Hilfsmittel zum Finden einer passenden Funktion ist das Streudiagramm. Viele Schüler tippen zunächst auf den Funktionstyp k/n, wenn sie die Daten im Streudiagramm sehen. Man erkennt schnell, dass bei Variation von k nie eine gute Anpassung erzielt werden kann. Die Funktionen k/n haben die Kovariationseigenschaft, das eine Vervierfachung von n zu einer Viertelung des Funktionswertes führt. Da aber nur eine Halbierung beobachtet wird, liegt es nahe, eine Anpassung mit Funktionen des Typs $f(n) = k/\sqrt{n}$ zu versuchen. Eine zu den Daten passende Funktion findet sich mit $f(n) = 2/\sqrt{n}$. Die Schüler in unserer GESIM-Studie hatten allerdings Schwierigkeiten die Wurzelfunktion völlig selbständig zu finden.

Dass die 95 %-Streubreite ziemlich genau durch $2/\sqrt{n}$ berechnet werden kann, lässt sich mathematisch herleiten. Man nennt dies das $1/\sqrt{n}$-Gesetz. Daraus ergeben sich auch die bekannten Faustregeln für die 95 %-Streubreite (vgl. Kasten unten). Diese Form der Präzisierung des Gesetzes der großen Zahlen ist wesentlich einsichtiger und tragfähiger als vage Konvergenzvorstellungen.

Stichprobenumfang n	**95 %-Streubreite**	**Radius der mittl. 95 %**
50	0,28	0,14
100	0,20	0,10
1000	0,06	0,03
10000	0,02	0,01

Tab. 3: Faustregeln für die Streubreite

Diese Faustregeln kann man mit Vorsicht auch gleichsam als Umkehrung für „intuitive Konfidenzintervalle“ nutzen. Beobachtet man eine relative Häufigkeit von 60 % bei $N = 1000$, so schätzt man die unbekannte Wahrscheinlichkeit als $0{,}60 \pm 0{,}03$. Diese Nutzung kann in diesem Aufsatz nicht detailliert erläutert werden.

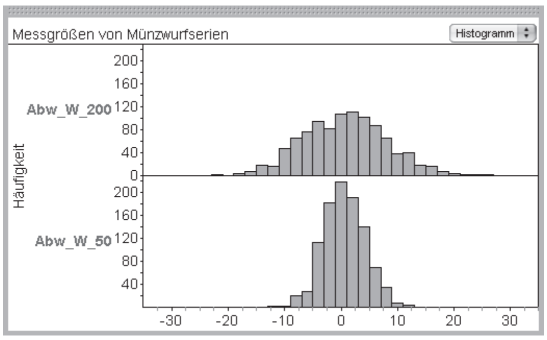
Korrektur von Fehlvorstellungen zur absoluten Häufigkeit

Mit der **LU 2** wurde anhand der Trajektorien demonstriert, dass sich die absoluten Häufigkeiten keineswegs dem Erwartungswert $n/2$ annähern. Dieser Aspekt sollte nun parallel zur Fortentwicklung der Vorstellung zu den Schwankungen der relativen Häufigkeiten mit entsprechenden Diagrammen und Simulationen verfeinert werden.

Wie könnte ein zu Abb. 7 analoges Diagramm aussehen? Während bei den relativen Häufigkeiten der Erwartungswert unabhängig von n bei 0,5 liegt, nimmt er bei den absoluten Häufigkeiten zu. Würde man also direkt die beiden analogen Verteilungen in ein Diagramm eintragen, würden diese nicht direkt untereinander liegen, sondern um 25 bzw. um 100 streuen. Die unterschiedliche Streuung wird durch diese Verschiebung nicht so gut sichtbar. Deshalb ist es besser, die Abweichungen der absoluten Häufigkeit vom jeweiligen Erwartungswert zu visualisieren, da diese beide um den Wert 0 streuen.

In der Lernumgebung **LU 6** (Abb. 9) haben wir dies realisiert. Es ist leicht zu sehen, dass sich die Verhältnisse bei den Streuungen im Vergleich zu Abb. 7 umgekehrt haben.

Um eine solche Häufigkeitsverteilung wie in Abb. 9 zu erzeugen, ermittelt man die Abweichung der Wappenanzahl vom jeweiligen Erwartungswert. Beträgt z. B. in einer simulierten 50er Münzwurfserie die Wappenanzahl 28, dann ist die Abweichung 3. Zeichnet man nun $N = 1000$ Münzwurfserien auf, ergeben sich die entsprechenden Häufigkeitsverteilungen der Abweichungen um $E(X_{50}) = 25$ bzw. $E(X_{200}) = 100$. Betrachtet man wiederum die mittleren 95 % der Häufigkeitsverteilungen (vgl. Tabelle in Abb. 9), dann erkennt man, dass sich die Schwankungsbreite verdoppelt, wenn man den Versuchsumfang vervierfacht. Analog wie bei der relativen Häufigkeit kann man nun zunächst mit der Untersuchung weiterer Beispiele, dann mit Streudiagrammen und Funktionsgraphen erarbeiten, dass die Funktion $k \cdot \sqrt{n}$ diese Eigenschaft hat und mit $k = 2$ gut zu den Daten passt.



Messgrößen von Münzwurfserien

	Abw_W_50	Abw_W_200
	-7	-14
	7,5	15
	14,5	29

S1 = Perzentil (2,5; ?)
S2 = Perzentil (97,5; ?)
S3 = Perzentil (97,5; ?) − Perzentil (2,5; ?)

Abb. 9: LU 6 – Histogramm und Auswertungstabelle für die Abweichungen vom Erwartungswert der Wappenanzahl bei $n = 50$ bzw. $n = 200$ Münzwürfen (Wiederholungsanzahl $N = 1000$)

Stufe 4 – $1/\sqrt{n}$-Gesetz theoretisch und Verbindung zu σ-Regeln

Bis zur Stufe 3 sind die gewonnenen Gesetzmäßigkeiten gleichsam empirische Gesetze, die an Beispielen verifiziert wurden. In der Oberstufe kann man nach dem Behandeln der Binomialverteilung den mehrfachen Münzwurf als Bernoulli-Kette theoretisch modellieren und das $1/\sqrt{n}$-Gesetz theoretisch aus dieser Modellierung herleiten.

Die Wappenanzahl beim Münzwurf ist eine binomialverteilte Zufallsgröße X mit den Parametern n und p. Der Wappen*anteil* wird dann durch die Zufallsgröße $Y = \frac{X}{n}$ beschrieben. Für die Standardabweichungen gilt $\sigma(X) = \sqrt{np(1-p)}$ und daraus folgend $\sigma(Y) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$. Für $n \cdot p \cdot (1-p) > 9$ kann man die Normalapproximation der Binomialverteilung verwenden:

$$P(X \in [np - 1{,}96\sigma(X), np + 1{,}96\sigma(X)]) \approx 95\ \%$$

Die 95 %-Streubreite ist also hier $2 \cdot 1{,}96\sqrt{np(1-p)} \leq 2\sqrt{n}$, da man $p(1-p)$ im Intervall zwischen 0 und 1 durch $\frac{1}{4}$ abschätzen kann. Dies verallgemeinert, quantifiziert und begründet die qualitativen Phänomene aus Abb. 9.

Hieraus folgt wegen $Y = \frac{X}{n}$ auch:

$$P(Y \in [p - 1{,}96\sigma(Y), p + 1{,}96\sigma(Y)]) \approx 95\ \%$$

Letzteres ist äquivalent zu

$$P\left(Y \in \left[p - 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \approx 95\ \%$$

Wegen der obigen Abschätzung für $p(1-p)$ folgt die einfachere Faustformel

$$P\left(Y \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \approx 95\ \%$$

Hieraus folgt das $1/\sqrt{n}$-Gesetz und alle bisher dargestellten Abschätzungen. Leider thematisieren nicht alle Schulbücher bei den σ-Regeln die Bedeutung von $1/\sqrt{n}$.

4 Rückschau

Die Lernenden werden bei unserem Vorschlag aktiv in den Prozess der Erkenntnisgewinnung einbezogen, z. B. über gezielte Arbeitsaufträge zu den einzelnen computergestützten Lernumgebungen. Dafür muss aber genügend Unterrichtszeit zur Verfügung gestellt werden. Mit den verschiedenen Lernumgebungen entwickeln die Schülerinnen und Schüler ein wesentlich facettenreicheres Bild des Gesetzes der großen Zahlen, als es üblicherweise in der Sekundarstufe I und II erreicht wird.

Der von uns vorgeschlagene Stufengang beruht auf Vorschlägen von Freudenthal zur Behandlung des empirischen Gesetzes der großen Zahlen (1972) und auf Vorschlägen von Riemer (1991) zur Behandlung des $1/\sqrt{n}$-Gesetzes in der Sekundarstufe I. Auch Borovcnik (1992) hat ein auf diesen Vorschlägen basierendes Vorgehen bereits vor knapp 20 Jahren diskutiert. Leider haben diese Vorschläge immer noch keinen nachhaltigen Eingang in den Mathematikunterricht der Schule gefunden. Mit unserer Idee für ein gestuftes Vorgehen gehen wir über die bisher gemachten Vorschläge hinaus, da wir den Fokus noch stärker auf Verteilungen richten. Insbesondere der Einsatz von Simulationen, deren Erzeugung und Auswertung durch die Werkzeugsoftware FATHOM in besonders intuitiver Weise unterstützt wird, ermöglicht den Aufbau von tragfähigen und mathematisch korrekten Vorstellungen zum Gesetz der großen Zahlen.

Für einen Einstieg in die Werkzeugsoftware FATHOM ist die e-Learning-Umgebung eFATHOM (http://eFATHOM.math.uni-paderborn.de/) zu empfehlen. Mit dieser videogestützten Lernumgebung kann man in kurzer Zeit die Grundlagen für den Umgang mit FATHOM erlernen. Als weiterführende Literatur mit vielen konkreten Beispielen und Beschreibungen bietet sich das Buch *FATHOM 2. Eine Einführung* von Biehler et al. (2006) an. Konkrete Unterrichtsvorschläge mit Aufgaben und Materialen für die Sekundarstufen I und II finden sich in Biehler et al. (2011).

Software

Fathom: Download einer kostenlosen Testversion, u. a. um mit den Lernumgebungen zu arbeiten, unter http://fathom.math.uni-paderborn.de.

Literatur

Biehler, R., & Steinbring, H. (1982). Bernoullis Theorem: Eine „Erklärung“ für das empirische Gesetz der großen Zahlen? In H.-G. Steiner (Hrsg.), *Mathematik Philosophie Bildung* (S. 296–334). Köln: Aulis.

Biehler, R., Hofmann, T., Maxara, C., & Prömmel, A. (2006). *FATHOM 2. Eine Einführung.* Heidelberg: Springer.

Biehler. R., Hofmann, T., Maxara, C., & Prömmel, A., (2011). *Daten und Zufall mit FATHOM. Unterrichtsideen für die Sekundarstufen 1 und 2.* Hannover: Schroedel.

Biehler, R., & Maxara, C. (2007). Integration von stochastischer Simulation in den Stochastikunterricht mit Hilfe von Werkzeugsoftware. In: *Der Mathematikunterricht, 53*(3), 45–61.

Biehler, R., & Prömmel, A. (2010). Developing students computer-supported simulation and modelling competencies by means of carefully designed working environments. *Proceedings of the 8th International Conference on Teaching Statistics.* Ljubljana (http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8D3_BIEHLER.pdf).

Borovcnik, M. (1992). *Stochastik im Wechselspiel von Intuitionen und Mathematik.* Mannheim: Bibliographisches Institut.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics.* Toronto: McLeod.

Freudenthal, H. (1972). The empirical law of large numbers or „the stability of frequencies“. In: *Educational Studies in Mathematics, 4,* 484–490.

Gigerenzer, G., Swijtink, Z., Porter, T. M., Daston, L., Beatty, J., & Krüger, L. (1999). *Das Reich des Zufalls.* Heidelberg: Spektrum.

Grosch, S. (2009). *Stochastik in komplexen Spielsituationen: Analyse von Lernprozessen und Lernergebnissen von Schülerinnen und Schülern der Klasse 8.* Staatsexamensarbeit. Kassel: Universität Kassel

Harten, G. v., & Steinbring, H. (1984). *Stochastik in der Sekundarstufe I.* Köln: Aulis.

Kahnemann, D., & Tversky, A. (1982). On the study of statistical intuitions. In: *Cognition, 11,* 123–141.

Hußmann, Stephan/Prediger, Susanne (2009): Je größer die Wurfzahl, desto sicherer die Wette – Mit dem Spiel Wettkönig den Zufall auf lange Sicht erkunden. In: *Praxis der Mathematik in der Schule, 51*(25), 24–29.

Leuders, Z. & Prediger, S. (2005) (Hrsg.): Funktioniert’s? Denken in Funktionen. In: *Praxis der Mathematik in der Schule, 47*(2).

Prömmel, A. (2013). *Das GESIM-Konzept – Rekonstruktion von Schülerwissen beim Einstieg in die Stochastik mit Simulationen.* Heidelberg: Springer Spektrum.

Prömmel, A. & Biehler, R. (2008). *Einstieg in die Wahrscheinlichkeitsrechnung mit FATHOM. Mathematikunterricht Klasse 8. Didaktisch kommentierte Materialien für den Unterricht.* Kassel: Universität Kassel

Riemer, W. (1991). Das ‚Eins durch Wurzel aus *n*‘ Gesetz. Einführung in statistisches Denken auf der Sekundarstufe I. In: *Stochastik in der Schule, 11*(3), 24–36.

Schnell, S. (2011). „Je höher die Zahlen, desto weniger Bewegung“. Lernende erkunden das empirische Gesetz der großen Zahlen. In: *Praxis der Mathematik in der Schule, 53*(39), 9–13.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: the empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33–51.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. In: *Psychological Bulletin, 76*(2), 105–110.

Anschrift der Verfasser

Prof. Dr. Rolf Biehler
Universität Paderborn
Institut für Mathematik
Warburger Str. 100
33098 Paderborn
biehler@math.upb.de

OStR Dr. Andreas Prömmel
Gymnasium Ernestinum Gotha
Bergallee 8
99867 Gotha
aproemmel@me.com