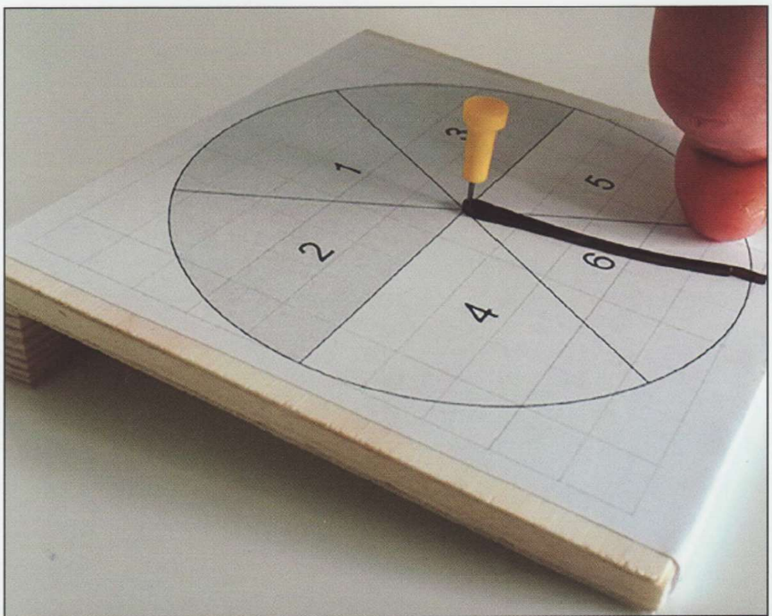
Bestell-Nr. 524210

MU

DER MATHEMATIK-UNTERRICHT



Wie ändern sich die Wahrscheinlichkeiten durch Schiefstellen des Glücksrades?

Stochastik

FRIEDRICH

Jahrgang 65 · Heft 6 · Dezember 2019

Jahrgang 65 · Heft 6–2019

DER MATHEMATIK-UNTERRICHT

Beiträge zu seiner fachlichen und fachdidaktischen Gestaltung

Schwerpunkt **Stochastik**

StD Dr. Wolfgang Riemer

Adressen der Autoren

Daniel Behrens
daniel.behrens87@gmail.com

Georg Berschneider
georg.berschneider@ovgu.de

Norbert Henze
Henze@kit.edu

Henning Körner
hen.koerner@t-online.de

Wolfgang Riemer
w.riemer@arcor.de

Judith Schilling
judith.schilling@kit.edu

René L. Schilling
rene.schilling@tu-dresden.de

Reinhard Schmidt
schmidt@seminargyge.de

Günter Seebach
guenter.seebach@t-online.de

BEITRÄGE

HENNING KÖRNER, WOLFGANG RIEMER

Beurteilende Statistik: ab Klasse 8!

1 Prolog

„Kann man eine Münze als fair durchgehen lassen, obwohl sie in einem oder in mehreren Versuchen häufiger Wappen als Zahl zeigte?“ „Wann gilt man als allgemeingebildet, wenn man an einem zugehörigen Test in MC-Format teilnimmt?“ „Wir haben eine Umfrage zum Mensaessen gemacht, was können wir daraus schließen?“

Bei der Beantwortung solch alltagsnaher Fragen hilft beurteilende Statistik, die, wenn überhaupt, in Kursen auf erhöhtem Niveau am Ende der Schulzeit thematisiert wird. Orientiert man sich an den Vorstellungen vieler Fachmathematiker, dann sollte dieser Themenbereich eher gar nicht an der Schule behandelt werden. Er sei zu schwierig. Eine Beschränkung auf elementare Wahrscheinlichkeitsrechnung, ergänzt um beschreibende Statistik, reiche aus.

Hier wird die genau komplementäre Position dazu eingenommen. Warum? Aus Gründen der Allgemeinbildung ist die Beschäftigung mit o. g. Fragen und den Antworten, die beurteilende Statistik bereithält, in Schule hochgradig wichtig, weil Menschen ständig davon betroffen sind. Aus Gründen einer fachwissenschaftlichen Orientierung auf entsprechende Studiengänge mag dieser Themenbereich weniger bedeutsam erscheinen.

Nun ist eine sinnvolle Durchdringung natürlich nur möglich, wenn man didaktisch reduziert und so elementarisiert, dass adäquate Grundvorstellungen, altersgerechte Umsetzungen und begriffliche Festlegungen möglich werden. Es muss „intellektuell redlich“ bleiben.

Ausgehend von einer Unterrichtsszene aus Klasse 8 wird gezeigt, wie eine Beschäftigung mit Fragen beurteilender Statistik im Stochastikunterricht der Sekundarstufe 1 angebahnt werden kann. Dabei ist der Einsatz digitaler Werkzeuge für Simulationen und Veranschaulichungen überaus hilfreich und erhellend.

2 Eine Unterrichtsszene

Situation: Eine Reißzwecke kann genau wie eine Münze auf zwei Seiten landen, wenn man mit ihr „würfelt“. Ein Landen auf der Spitze ist noch „unmöglicher“ als das Landen einer Münze auf der Kante.

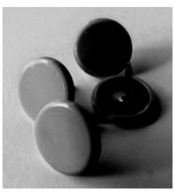


Abb. 1: einmal Rückenlage (R), dreimal Seitenlage (S)

Abb. 2: normierter als Reißzwecken sind flache Legosteine, die nur auf dem Fuß (F) oder auf dem Kopf (K) landen.

Frage: Ist das Werfen einer Reißzwecke genauso fair wie das Werfen einer Münze, sind also die Chancen, mit Rücken (R) zu gewinnen, genauso groß wie die Chance für Seite (S)?

Wir spekulieren und stimmen dann ab (man könnte von einer ersten Stufe im Modellierungskreislauf sprechen):

18 Schülerinnen und Schüler entscheiden sich für „fair“,

2 meinten, dass R günstiger ist, 9, dass S „günstiger" sei.

Es werden Begründungen versucht:

- Der Schwerpunkt liegt mehr beim Rücken, also ist R häufiger
- Der Auflagebereich bei Seitenlage ist größer, also wird S häufiger kommen.

Was beim Wurf einer Münze aufgrund nicht hinterfragter Symmetrie von vornherein klar ist, ist bei der Reißzwecke verzwickter.

Durch reine Überlegungen ist die Sache nicht zu klären. Werfen wir also einfach Reißzwecken, steigen wir in den Modellierungskreislauf ein.

Wenn wir ganz viel werfen, müsste es ja ungefähr „fifty-fifty" ausgehen, wenn das Spiel fair wäre. Wir werfen in 50er-Serien:

R	28	26	19	20	21	23	28	21	32	13	27	29	27	28	19	18
S	22	24	31	30	29	27	22	29	18	37	23	21	23	22	31	32
h(R)	0,56	0,52	0,38	0,40	0,42	0,46	0,56	0,42	0,64	0,26	0,54	0,58	0,54	0,56	0,38	0,36

und werten aus: 8-mal hat R, 8-mal hat S gewonnen.

Absolute Häufigkeiten: R: 379, S: 421; relative Häufigkeiten: R: 0,47; S: 0,53.

Da wir ja wissen, dass beim Münzwurf wegen der Zufallsschwankungen auch „nie" exakt „fifty-fifty" herauskommt und das Ergebnis doch recht nahe daran liegt, außerdem auch R und S gleich häufig gewonnen haben, verfestigt sich die Vermutung, dass es sich beim „Zweckenwürfeln" um ein faires Spiel handelt. Es wird erneut abgestimmt:

Alle bis auf Martin halten „fifty-fifty" für ein *gutes Modell*. Nur Martin zweifelt: S ist günstiger, nicht aufgrund der größeren relativen Häufigkeit, sondern wegen des größeren Auflagebereichs.

Wir schauen das kumulierte Diagramm der relativen Häufigkeit an für Rückenlage:

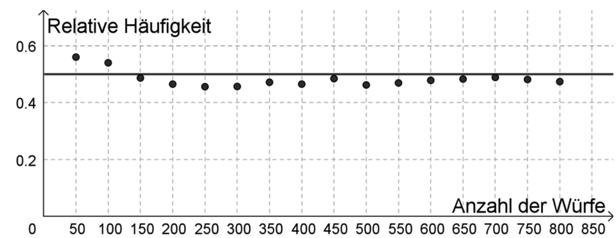


Abb. 3:

Die relativen Häufigkeiten für Rückenlage liegen zwar mehrheitlich etwas unterhalb von 0,5, aber die Lerngruppe bleibt bei ihrem fifty-fifty-Modell („Das ist eben Zufall").

Frage: Bei welcher Abweichung von „fifty-fifty" würdet ihr **bei 800 Würfen** anfangen, die Fairness der Zwecke zu bezweifeln? Wieder gibt es unterschiedliche Positionen:

Katrin: 0,35/0,65 – Marco: 0,4/0,6 – Christian: 0,45/0,55.

Einig ist man sich nur hinsichtlich der „unangefochtenen Gültigkeit" folgender Aussage:

„Je größer die Abweichung, desto stärker der Zweifel."

Um – jenseits aller Spekulationen – ein sicheres Gefühl für „normale" Zufallsschwankungen zu entwickeln, wird „simuliert, was der GTR hergibt" …

40 Serien mit je 800 „fairen" Münzen lieferten die folgenden absoluten Häufigkeiten:

393	412	408	390	401
403	387	416	407	399
398	424	389	409	390
383	396	426	399	389
376	387	402	390	411
395	398	409	401	378
398	401	386	392	397
404	414	389	393	403

Abb. 4a

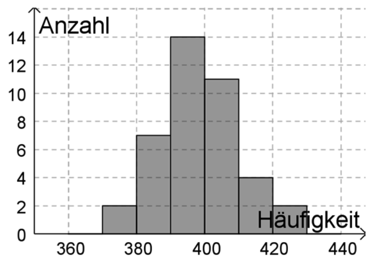


Abb. 4b

Nur zweimal (also in 2/40 = 5 % der Simulationen) kommt es zu Ergebnissen wie beim händischen Reißzweckenwurf – mit einer „folgenschweren" Konsequenz:

Wenn Reißzweckenwerfen tatsächlich „fifty-fifty" wäre, dann wäre das im Unterricht „erwürfelte" Ergebnis also recht unwahrscheinlich. Dann ist Reißzweckenwerfen vielleicht doch nicht „fifty-fifty"? Einstimmig wird umgeschwenkt: Alle bezweifeln „fifty-fifty".

Die Simulation zeigt, dass die Achtklässler für $n = 800$ schon bei $h = 0{,}47$ „fifty-fifty" bezweifeln, dann erst recht bei $h = 0{,}4$ und $h = 0{,}35$. Diskussionsbedarf zu den Vorschlägen von Kathrin, Marco und Christian gibt es nicht mehr, Marco und Kathrin waren viel zu „streng".

Bei den 50er-Serien dagegen lag $h = 0{,}47$ noch „voll im normalen Bereich" und hätte keine Veranlassung geboten, die Fairness der Zwecke zu bezweifeln. Hier wäre Kathrins Setzung sinnvoll gewesen. Man ist mittendrin im „Testen", in der Bewertung von Abweichungen in Abhängigkeit vom Versuchsumfang n, nicht formal, sondern inhaltlich, fallbezogen, erfahrungsbasiert und absolut sinnstiftend.

Die Unterrichtsszene ist als Aufgabe in einem Schulbuch gestaltet (KÖRNER [2015], S. 133).

3 Simulationen

Bei den Vorschlägen von Kathrin, Marco und Christian handelt es sich nicht um Vermutungen, die falsifiziert oder verifiziert werden können, sondern um normative Setzungen, die jede für sich mit unterschiedlichen Konsequenzen und Interpretationen benutzt werden können und bei denen man den Stichprobenumfang n immer mit bedenken muss.

Dies ist Anlass, den Schwankungen der (relativen) Häufigkeiten in Abhängigkeit von n genauer auf die Spur zu kommen. Im Fokus stehen also nicht mehr relative Häufigkeiten „an sich", sondern deren Schwankung (zentrales Moment stochastischen Denkens) um die (angenommene oder vermutete) Wahrscheinlichkeit, hier $p = 1/2$.

Hier entfalten nun Simulationen mit digitalen Werkzeugen ihre ganze Kraft, nicht nur, weil Computer schnell würfeln und Daten auswerten/visualisieren, sondern weil man Wahrscheinlichkeiten über Zufallsgeneratoren exakt vorgeben kann.

Um zunächst nahe am Realexperiment zu bleiben, reicht im Erstzugang das sukzessive Simulieren durch Tastendruck aus. Man erlebt die Schwankungen als Folge („Film") von Werten und die Verkleinerung der Zufallsschwankungen mit wachsendem n wird unmittelbar **erlebbar**. Die affektive (für viele Schülerinnen und Schüler durchaus überraschende) Komponente des Erlebens kann man nicht hoch genug wertschätzen.

	A reiss	B
=	=randsamp({0,1},'n)	
1	1.	23.
2	1.	**50.**
3	0.	0.46

	A reiss	B
=	=randsamp({0,1},'n)	
1	1.	31.
2	0.	**50.**
3	0.	0.62

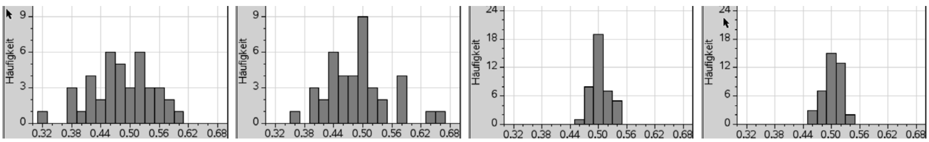
	A reiss	B
=	=randsamp({	
1	0.	428.
2	1.	**800.**
3	0.	0.535

	A reiss	B
=	=randsamp({	
1	1.	398.
2	0.	**800.**
3	0.	0.4975

Abb 5a **b** **c** **d**

In einem zweiten Schritt wird man die Verkleinerung des Bereichs (Intervalls), in dem die Werte liegen, studieren und wie folgt in Worte fassen: „Sechzehnmal so viel: viermal so genau!“

n = 50, 40 Serien n = 800, 40 Serien



0,32 < h < 0,60 0,36 < h < 0,66 0,46 < h < 0,54 0,46 < h < 0,54

Abb. 6a **b** **c** **d**

Mit dem GTR überprüft man das (in numerischer Präzisierung von **Abb. 6**) durch Auszählen von Listen relativer Häufigkeiten.

Mit Dynamischen Geometrieprogrammen bieten sich Visualisierungen wie in **Abb. 7** an. Dort wurden links die Schwankungen der relativen Häufigkeiten an den Stellen $n = 25$, $n = 100$, $n = 400$ im Spurmodus durch „springende“ Punkte und rechts zusätzlich durch Boxplots veranschaulicht. Man erkennt:

Die Länge der Boxen halbiert sich bei Vervierfachung des Versuchsumfangs.

Der Quartilabstand ist für Schüler der Sekundarstufe 1, die noch keine Standardabweichung kennen, ein geniales Streuungsmaß, das die Faustregel prägnant begreifbar macht: Viermal so viel: doppelt genau, sechzehnmal so viel, vierfach genau …

a b

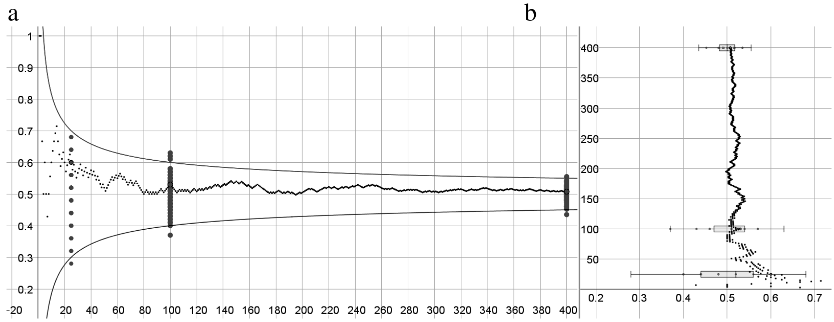


Abb. 7: **a, b:** relative Häufigkeiten von 100 Münzwürfen an den „Stellen“ n = 25, 100 und 400

Stehen ab Klasse 9 Wurzeln zur Verfügung, kann man verallgemeinern und das $\frac{1}{\sqrt{n}}$-Gesetz als Naturgesetz festhalten: Fast immer (in ca. 95 % aller Fälle) liegt die relative Häufigkeit h im Intervall $\left[p-\frac{1}{\sqrt{n}}; p+\frac{1}{\sqrt{n}}\right]$, das in **Abb. 7a** in Abhängigkeit von n als Wurzeltrichter dargestellt ist.[1]

Wenn man bei einer Wahrscheinlichkeitsangabe p nicht ganz sicher ist und eine relative Häufigkeit außerhalb dieses Intervalls beobachtet, dann hat man Grund, an der angenommenen Wahrscheinlichkeit p zu zweifeln.

4 Die Suche geht weiter – die „kopernikanische Wende" in der Didaktik der Statistik

Nachdem „fifty-fifty" einstimmig „für die Tonne" war, wird die Frage nach einer „besseren" (der „tatsächlichen"?) Wahrscheinlichkeit virulent: Vielleicht 0,48 – oder 0,47 – oder 0,46? Keine Chance! Der Funke ist übergesprungen! Schenken Sie „Ihren" Kindern die Viertelstunde. Länger braucht es nicht für *eine* relative Häufigkeit h aus $n = 4000$ echten Zweckenwürfen.

Und dann wird simuliert. Arbeitsteilig! In Gruppen! Für die Wahrscheinlichkeiten $p = 0{,}49$, 0,47, 0,45, 0,43 … immer 4000 Würfe! Zu welchen Wahrscheinlichkeiten passt die „erwürfelte" relative Häufigkeit h? Bei welchen liegt sie „mittendrin", bei welchen am Rand? Bei welchen liegt sie so, dass man das angenommene p nicht bezweifeln muss, sodass p als brauchbares Modell für die Wirklichkeit Bestand hat? Alle diejenigen Gruppen, bei denen das der Fall ist, erheben sich und reihen sich der Größe (von p!) nach auf. Das Konfidenzintervall ist (numerisch etwas ungenau, aber begrifflich absolut sauber, intellektuell ehrlich) gefunden! In Klasse 8!
Und die Erkenntnis mit nahezu philosophischer Tiefe: Uns Menschen wird trotz des „je mehr desto genauer" auf ewig „die wahre Wahrscheinlichkeit" verborgen bleiben.
Die kennt nur Gott! … Aber auch das ist nicht ganz sicher.
Nutzen die Zwecken sich nicht ab, wenn man von Ewigkeit zu Ewigkeit würfelt?

Man vergleiche den spannungsgeladenen Unterrichtsgang, der in JEDER Klasse zum Selbstläufer wird, weil er mit einer echten Frage für Spannung sorgt, mit dem „klassisch-frequentistischen", in dem der Lehrer verzweifelt … und vergeblich … versucht, Wahrscheinlichkeit nicht als Modell der Wirklichkeit, sondern als objektiv existenten Grenzwert relativer Häufigkeiten in den Köpfen der Kinder zu verankern. Der Schüler Marco (vgl. Kasten) bringt es auf den Punkt: „Die Wahrscheinlichkeit hängt vom Zufall ab." Recht hat er! Das Modell, das wir uns von der Zwecke machen, hängt von den experimentellen Ergebnissen ab:

„Aus Erfahrung wird Erwartung!" Aber das hat der fragend-entwickelnd arbeitende Lehrer nicht verstanden, der Modellbildungskreislauf passt nicht in sein Weltbild objektiver Wahrscheinlichkeit.

[1] Für allgemeines p begründet man in der Sekundarstufe 2 mit den Sigmaregeln der Binomialverteilung: in ca. 95 % aller Fälle liegt die relative Häufigkeit im „Prognose"-Intervall

$$\left[p-2\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p+2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subseteq \left[p-\frac{1}{\sqrt{n}}; p+\frac{1}{\sqrt{n}}\right].$$

Die Abschätzung ist für die Sekundarstufe I zur Entwicklung von Grundvorstellungen und als Faustregel fürs Kopfrechnen völlig ausreichend, insbesondere, wenn man sich auf Wahrscheinlichkeiten $p \approx 0{,}5$ konzentriert. Vgl. RIEMER [1991].

Klasse 8, Gymnasium, 2. Stunde Wahrscheinlichkeitsrechnung.

An der Tafel stehen (aufsummierte) relative Häufigkeiten für das Ergebnis „Spitze“ beim Werfen einer Reißzwecke.

Lehrer: „54 %, 60 %, 54 %, 51 %, 55 %, 60 % … Sind das Wahrscheinlichkeiten?“

Schüler: „Weiß nicht.“

Lehrer: „Was sind denn das für Dinger?“

Schüler: „Wahrscheinlichkeiten.“

Einige Minuten später …

Marco: „Die Wahrscheinlichkeit hängt vom Zufall ab.“

Lehrer: „Die Wahrscheinlichkeit hängt nicht so sehr vom Zufall ab. Die Versuchsausgänge hängen vom Zufall ab.“

Lehrer: „Was hat jetzt die relative Häufigkeit mit Wahrscheinlichkeit zu tun?“

Schüler: Erklärt, wie man die relative Häufigkeit ausrechnet.

Lehrer: „Ja, aber ist jetzt Wahrscheinlichkeit dasselbe wie relative Häufigkeit?“

Stefan: „Ich würde sagen ja. Von der Herkunft würde ich sagen ja. Nur relative Häufigkeit ist im Vergleich zur Wahrscheinlichkeit ungenauer.“

Lehrer: (deutet auf eine relative Häufigkeit): „Bist du denn sicher, dass das die richtige Wahrscheinlichkeit ist?“

Stefan: „Nein. Die kann man gar nicht feststellen.“

Lehrer: „Die kann man gar nicht feststellen. So etwas gibt es, aber feststellen können wir sie nicht. Was können wir machen für die Wahrscheinlichkeit?“

Schüler: „Überhaupt nichts.“ BAUERSFELD/VOIGT [1986]

Durch die Idee, Wahrscheinlichkeiten als vom Menschen gesetzte Modelle der Wirklichkeit zu begreifen, die man in Modellbildungskreisläufen nachjustiert und optimiert, schlägt man mehrere Fliegen mit einer Klappe:

- man knüpft (über das vor jedem Experiment kultivierte Schätzen von Chancen) an subjektivistische Vorerfahrungen aus der Grundschule/dem Alltag an
- man braucht keine unendlich langen Versuchsreihen, die es in Wirklichkeit nicht gibt
- man verwirft/akzeptiert Hypothesen nicht mehr, man bezweifelt und verbessert Modelle.

Und das könnte man als „kopernikanische Wende“ oder als didaktisches „Ei des Kolumbus“ bezeichnen: Nicht mehr die in vielen Schulbüchern kultivierte Idee des R. v. MISES, Wahrscheinlichkeit als „Grenzwert relativer Häufigkeiten“ zu „definieren“ steht im Mittelpunkt unseres stochastischen Weltbildes, sondern die hinter A. N. KOLMOGOROFFS Axiomatik steckende Sicht, Wahrscheinlichkeit als vom Menschen gesetztes Modell zu begreifen, das – wie alle Modelle – einem stetigen Prozess der Optimierung („Modellierungskreislauf“) unterliegt. Die fast schon dialektische Pointe ist, dass hier strukturmathematische[2] und fachdidaktische Intentionen – wenn auch aus unterschiedlichen Beweggründen – zusammenkommen.

5 Forschungsprotokoll

Forschende Kinder sind dankbar, wenn man den durchlebten Gedankengang und die Schritte der „verzwickten“ Forschung in wenigen knackigen Sätzen z. B. wie folgt zusammenfasst:

[2] Leider trug die in den 1970er-Jahren während der „strukturmathematischen Welle“ kultivierte „Ereignisalgebra“ zur Förderung stochastischer Grundvorstellungen wenig bei und hat den Namen KOLMOGOROFF – der in keinem Lehrplan mehr auftaucht – in schulischen Gefilden völlig zu Unrecht diskreditiert.

Wahrscheinlichkeiten prüfen – Mit Wahrscheinlichkeiten Wirklichkeit modellieren.

- Die relativen Häufigkeiten schwanken zufallsbedingt um die Wahrscheinlichkeit p.
- Die Schwankungen sind umso kleiner, je größer der Versuchsumfang n ist.
- Welche Schwankungen bei einen vorgegebenen Versuchsumfang n noch normal sind, beantworten entweder Simulationen oder das $1/\sqrt{n}$ -Gesetz, dabei wird das, was als normal gelten soll, durch eine Setzung („Wie viele Werte sollen innerhalb eines bestimmten Bereichs liegen?“ – meist sind es 95 %) festgelegt.
- Wenn man bei der Wahrscheinlichkeitsangabe p unsicher ist und die relativen Häufigkeiten deutlich stärker von p abweichen als normal, verspürt man ein „Grummeln im Bauch“: Man bezweifelt die Wahrscheinlichkeitsangabe und sucht (vielleicht) eine bessere oder auch ein Intervall von mehreren besseren Wahrscheinlichkeiten.
- Je mehr Versuche man gemacht hat, desto kleiner wird das Intervall der in Frage kommenden Wahrscheinlichkeiten.

In unserem Beispiel gab es folgendes Ergebnis: $p(\text{Kopf}) = 0{,}5$ beschreibt das Würfeln mit Zwecken „für den Hausgebrauch“ oder „Pi mal Daumen“ recht gut, aber eben nicht „richtig gut“. Das erkennt man, wenn man den Stichprobenumfang von $n = 50$ auf $n = 800$ ver-16-facht.

6 Resüme und Ausblick

Mit einem Unterrichtsgang der vorgestellten Form hat man in der Sekundarstufe I das (zweiseitige) Testen von Hypothesen der Sekundarstufe II jenseits aller Rezeptitis inhaltlich und so sinnstiftend grundgelegt, dass es den Unterzeichnern der Petition „Retire statistical significance“ [Amrhein u. a. 2019] vermutlich warm ums Herz wird. Und wer in der Sekundarstufe, statt Hypothesen zu testen, mit Konfidenzintervallen Wahrscheinlichkeiten schätzen darf, der hat seine Schüler mit dem Reißzweckenproblem aufs allersolideste vorbereitet, denn in den Konfidenzintervallen werden in der Sekundarstufe 2 alle diejenigen Wahrscheinlichkeiten gesammelt, die man nach dem obigen Konzept nicht zu bezweifeln braucht.[3] Dann wird man sagen.: „Das Konfidenzintervall der unbekannten Zweckenwahrscheinlichkeit überdeckt bei $n = 50$ Versuchen den Wert $p = 0{,}5$, nicht aber bei $n = 800$ Versuchen. Denn bei Ver-16-fachung des Versuchsumfanges schrumpft die Länge des Konfidenzintervalls auf ca. ¼ zusammen.“

Und noch eine Anmerkung zum Schluss: Die Reißzwecke steht nicht für Würfelbudenmathematik. Mit dem Vorzeichentest (auf $p = 1/2$) prüft man, ob sich Reaktionszeiten (durch Konsum von Muntermachern) verbessern, ob Gewichtsangaben eingehalten werden, ob beliebige Wahrscheinlichkeitsverteilungen als gleich durchgehen können …

Aber das ist dann ein anderes Thema.

Literatur

[1] Amrhein, V. Greenland, S. McShane, B. (2019): Retire statistical significance. Nature 567, S. 305 – 307. 2019

[2] Bauersfeld, H., Voigt. J. (1986): Den Schüler abholen, wo er steht! In: Lernen – Ereignis und Routine. Friedrich Jahresheft IV 1986: 18 – 20.

[3] Körner, H., Lergenmüller, A., Schmidt, G., Zacharias, M. (Hrsg.): Mathematik Neue Wege Niedersachsen 8, Braunschweig 2019.

[4] Riemer, W. (1991): Das „Eins durch Wurzel aus n“-Gesetz. Einführung in statistisches Denken auf der Sekundarstufe I. Stochastik in der Schule 11 (1991), Nr. 3, 24 – 36.

[3] Zum Vergleich für eigene Experimente: Flache Lego-Vierer: 95 %-Konfidenzintervall für p (Fußlage): [0,406;0,439] bei n = 3597, Reißzwecke: 95 %-Konfidenzintervall für p (Rückenlage): [0,349;0,509] bei $n = 800$.