Absolute Häufigkeiten widersetzen sich dem Gesetz der großen Zahl[1]

RUMA FALK UND AVITAL LAVIE LANN, JERUSALEM

[1] Original: Numbers defy the law of large numbers.
In *Teaching Statistics* 37 (2015) 2, 54–60.
Kürzung, Bearbeitung und Übersetzung: JÖRG MEYER

Zusammenfassung: *Wenn man immer häufiger eine faire Münze wirft, wird das Verhältnis von „Kopf“ und „Zahl“ sich immer mehr der 1 nähern. Das wird von vielen Lernenden falsch interpretiert, indem sie annehmen, dass die Differenz der absoluten Häufigkeiten von „Kopf“ und „Zahl“ sich immer mehr der 0 nähern müsse, obwohl sich die Differenz sehr weit von der 0 entfernen kann. Dieses Missverständnis sollte sehr früh ausgeräumt werden.*

1 Einleitung

Was erwartet man von einer Folge unabhängiger Münzwürfe, wenn die Anzahl der Würfe immer größer wird? Eine pointierte Teilantwort stammt von Weaver (1963, S. 185): „Es ist charakteristisch für solche Folgen, dass sie sich im relativen Sinne immer zahmer verhalten, sich aber im absoluten Sinne immer wilder gebären.“ In diesem Artikel werden wir das statistische Verhalten solcher Münzwurffolgen untersuchen und parallel dazu darstellen, wie Leute die zugehörigen Regeln mitunter (miss-)verstehen.

2 Das Gesetz der großen Zahl

Dies Gesetz handelt von der Tendenz der relativen Häufigkeiten, sich der zugehörigen Wahrscheinlichkeit zu nähern, wenn die Anzahl der Versuchsdurchführungen anwächst. Wir schränken unsere Diskussion auf den Spezialfall des Werfens einer fairen Münze ein. Ein Autor populärer Wissenschaft schreibt dazu:

„Dies Gesetz wurde zuerst von Jakob Bernoulli gefunden und 1713 posthum veröffentlicht. Es besagt, dass die Differenz zwischen 1/2 und dem Quotienten zwischen der Anzahl von „Kopf“ und der Anzahl der Würfe bewiesenermaßen beliebig dicht sich der 0 annähert, wenn die Anzahl der Würfe immer weiter und grenzenlos vergrößert wird.“ (Paulos 1991, S. 227)

Wenn n die Anzahl der Würfe und $n(H)$ die Anzahl von „Kopf“ bezeichnet und wenn die Wahrscheinlichkeit für „Kopf“ 1/2 beträgt, dann gilt für jedes $\varepsilon > 0$, dass

$$p\left\{\left|\frac{n(H)}{n}-\frac{1}{2}\right|<\varepsilon\right\} \xrightarrow{n\to\infty} 1$$

ist. Diese Formel beinhaltet das schwache Gesetz der großen Zahl. Bei Feller (1957), Ross (1984, Kap. 8) und Thomasian (1969, Kap. 12) kann man sich über das schwache und das starke Gesetz der großen Zahl informieren.

3 Die Wichtigkeit des Gesetzes der großen Zahl

Das Gesetz verbindet empirische relative Häufigkeiten und theoretische Wahrscheinlichkeiten. Es tritt im täglichen Leben auf, indem Regularitäten in verschiedenen sozialen und demographischen Massenphänomenen zu beobachten sind. Viele Ereignisse der Welt wie Arbeitslosigkeit oder Selbstmorde erscheinen als erratisch, wenn man nur wenige Fälle betrachtet, zeigen jedoch ein zunehmend reguläres Verhalten, je solcher Fälle in den Blick kommen (Nickerson 2004, S. 238–239; Paulos 1991, S. 227; Stigler 1986, S. 169–174; Weaver 1963, S. 177).

4 Zum populären Verständnis des Gesetzes der großen Zahl

Die Ansicht, dass das Gesetz etwas ausdrückt, was jedermann versteht, wird etwa von Feller (1957, S. 141), Nisbett et al. (1983) und Rosenhouse (2009, S. 86–87) geteilt. Andererseits gibt es einige wenige psychologische und didaktische Studien, die in eine andere Richtung weisen (Kahneman 2011, S. 112 f.).

Im oft zitiertem Krankenhausproblem von Tversky und Kahneman (1974) mussten die Teilnehmer entscheiden, welches Krankenhaus im Jahr mehr als 60 % Knabengeburten habe: ein kleines Krankenhaus mit wenigen Geburten pro Tag oder ein großes mit mehr Geburten pro Tag. Die große Mehrheit der Teilnehmer nahm an, dass die Größe des Krankenhauses egal sei. Die Unsensibilität gegenüber der Versuchsgröße steht nicht im Einklang mit dem Gesetz der großen Zahl, wonach das kleine Krankenhaus an einer größeren Anzahl von Tagen eine größere Abweichung vom Mittelwert der 50 % Knabengeburten zu erwarten hat. Dies Missverständnis wuchs sogar mit dem Alter der Teilnehmer. Unterricht in Stochastik konnte dieses Missverständnis eindämmen.

Hier soll untersucht werden, ob Leute erkennen, dass mit zunehmender Versuchsanzahl der Quotient sich stabilisiert und nicht die Differenz.

5 Quotienten gegen Differenzen

Ein typisches Münzwurfprotokoll ist in Tab. 1 dargestellt. Wenn n wächst, nähert sich der Anteil der „Kopf"-Ergebnisse 1/2, während gleichzeitig die absolute Differenz zwischen der Anzahl der Köpfe und der Zahlen wächst sowie auch sich immer weiter von 1/2 entfernt.

n	$n(H)$	$n(T)$	$\lvert D_n\rvert = \lvert n(H) - n(T)\rvert$	$\frac{n(H)}{n}$
10	6	4	2	0,6
100	57	43	14	0,57
1.000	525	475	50	0,525
10.000	5.044	4.956	88	0,504

Tab. 1: Mögliche Wurfergebnisse

Tabelle 2 zeigt die mathematischen Wahrscheinlichkeiten, dass $n(H) - n(T) = 0$ und dass $|n(H) - n(T)| \leq 2$ ist. Ein Beweis, dass die Wahrscheinlichkeit von $n(H) - n(T) = 0$ für wachsendes n gegen 0 geht, findet sich in Lipkin (2003).

n	$p(D_n = 0)$	$p(\lvert D_n\rvert \leq 2)$
2	0,500	1,000
10	0,246	0,656
20	0,176	0,497
50	0,112	0,328
100	0,080	0,236
500	0,036	0,107
1.000	0,025	0,076

Tab. 2: Wahrscheinlichkeiten für (fast) gleiche Anzahlen von Köpfen und Zahlen

Für wachsendes n tendiert $\frac{n(T)}{n(H)}$ gegen 1, also wird die Differenz ihrer Logarithmen gegen 0 tendieren, obwohl $|n(H) - n(T)|$ beliebig groß werden kann (Székely 1986, S. 35).

6 Untersuchung

187 Studenten der hebräischen Universität wurden getestet, die alle noch nicht auf den entscheidenden Unterschied zwischen Quotienten und Differenzen hingewiesen worden waren. Es gab 5 verschiedene Fragebögen (EN, CN, GN, CP, GP). Der gemeinsame Text war:

„Wir werfen eine faire Münze sehr oft. Du bekommst einen Preis, wenn

EN: $n(T) = n(H)$

CN: $|n(T) - n(H)| \leq 2$

GN: $|n(T) - n(H) > 2$

CP: $0{,}4 \leq \frac{n(H)}{n} \leq 0{,}6$

GP: $\frac{n(H)}{n} < 0{,}4$ oder $0{,}6 < \frac{n(H)}{n}$

Wie oft möchtest du, dass geworfen wird? 10-mal oder 100-mal oder ist es dir egal?"

Es gab auch parallele Fragebögen, bei denen die 10 und 100 ersetzt wurden durch 500 und 1.000, um zu verhindern, dass jemand bei 10 Würfen die Wahrscheinlichkeiten ausrechnen würde. Insgesamt gab es also 10 verschiedene Fragebögen. Die meisten Teilnehmer bekamen zwei verschiedene Fragebögen.

Die richtigen Antworten waren:

EN, CN, GP: wenig Würfe

GN, CP: viele Würfe

7 Resultate

Es gab keine erkennbaren Einflüsse akademischer Bildung. 36,3 % der Fragebögen wiesen die Antwort auf, dass die Anzahl der Würfe egal sei.

Diese Unsensibilität gegenüber der Größe der Stichprobe entspricht den Untersuchungen von Tversky und Kahneman (1974) sowie von Fischbein und Schnarch (1997). Die sich nun anschließende Darstellung der Resultate bezieht sich auf die Fragebögen, bei denen die Anzahl der Würfe nicht egal war.

Die Anzahl der richtigen Antworten betrug 61 %, wenn man zwischen 10 und 100 Würfen wählen konnte, und 52 %, wenn man zwischen 500 und 1.000 Würfen wählen musste.

Die folgende Auflistung gibt die Resultate genauer wieder, und zwar als Verhältnisse zwischen der Anzahl der richtigen Antworten und der Anzahl aller derjenigen Antworten, die sich zwischen 10 und 100 bzw. zwischen 500 und 1.000 entschieden hatten.

EN 27 %

CN 57 %

CP 75 %

GN 58 %

GP 79 %

Insgesamt waren 57 % aller Antworten richtig. Der Prozentsatz richtiger Antworten war größer, wenn es um Anteile ging (CP und GP). Die Teilnehmer konnten den Inhalt des Gesetzes der großen Zahl besser beurteilen, wenn es um Anteile bzw. um Prozentzahlen ging, und waren eher verwirrt, wenn es um die absoluten Häufigkeiten ging. Am kleinsten war die Erfolgsquote bei EN; die falsche Antwort wurde erläutert mit „Je häufiger man wirft, um so größer ist die Chance, gleich viele Köpfe wie Zahlen zu bekommen".

Insgesamt war das Format (absolute Häufigkeiten oder Prozentangaben) entscheidend. Bei Prozentangaben war die Erfolgsquote deutlich höher.

8 Diskussion

Über ein Drittel der Versuchsteilnehmer meinte, dass die Anzahl n der Versuche beim Verhältnis $n(H)/n(T)$ keine Rolle spielen würde. Diejenigen, denen der Einfluss von n klar war, konnten mit den Aufgaben gut umgehen, wenn sie im Prozent-Format gestellt waren. Über 40 % der Teilnehmer, denen der Einfluss von n klar war, hatten Schwierigkeiten mit EN, CN und GN, also mit den Aufgaben, die im Format der absoluten Häufigkeiten gestellt waren. Teilnehmer, die CP und GP korrekt hatten, hatten gleichwohl EN, CN und GN falsch.

Es scheint so, als würden viele Leute glauben, dass sich absolute und relative Häufigkeiten gleich verhalten würden. Sie beachten nicht, dass das bei wachsendem n überhaupt nicht der Fall zu sein braucht und dass sich große Unterschiede zwischen $n(H)$ und $n(T)$ bei Division durch ein noch viel größeres n kaum noch bemerkbar machen.

Das Denken in absoluten Häufigkeiten fällt leichter als das Denken in relativen Häufigkeiten, wie Gigerenzer und Hoffrage (1995) deutlich gemacht haben. Beim Lösen von Bayes-Problemen kann dies fruchtbar gemacht werden.

9 Didaktische Konsequenzen

Da Experimente überzeugender sind als alles Reden, ist es sinnvoll, eine Tabelle wie Tabelle 1 im Unterricht zu erzeugen und ganz deutlich darauf aufmerksam zu machen, wie unterschiedlich sich die Quotienten und die Differenzen verhalten.

Literatur

Feller, W. (1957): An Introduction to Probability Theory and Its Applications (2nd edn) Vol. 1. New York: Wiley.

Fischbein, E.; Schnarch, D. (1997): The evolution with age of probabilistic, intuitively based misconceptions. In: *Journal for Research in Mathematics Education*, 28(1), 96–105.

Gigerenzer, G.; Hoffrage, U. (1995): How to improve Bayesian reasoning without instruction: Frequency formats. In: *Psychological Review*, 102(4), 684–704.

Kahneman, D. (2011): Thinking Fast and Slow. London: Penguin.

Lipkin, L. (2003): Tossing a fair coin. In: *College Mathematics Journal*, 34(2), 128–133.

Nickerson, R. S. (2004): Cognition and Chance: The Psychology of Probabilistic Reasoning. Mahwah, NJ: Lawrence Erlbaum.

Nisbett, R. E.; Krantz, D. H.; Jepson, C.; Kunda, Z. (1983): The use of statistical heuristics in everyday inductive reasoning. In: *Psychological Review*, 90(4), 339–363.

Paulos, J. A. (1991): Beyond Numeracy: Ruminations of a Numbers Man. New York: Alfred A. Knopf.

Rosenhouse, J. (2009): The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser. Oxford: Oxford University Press.

Ross, S. (1984): A First Course in Probability (2nd edn). New York: Macmillan.

Stigler, S. M. (1986): The History of Statistics: The Measurement of Uncertainty Before 1900. Cambridge, MA: The Belknap Press of Harvard University Press.

Székely, G. J. (1986): Paradoxes in Probability Theory and Mathematical Statistics. Dordrecht, Holland: Reidel.

Thomasian, A. J. (1969): The Structure of Probability Theory with Applications. New York: McGraw-Hill.

Tversky, A.; Kahneman, D. (1974): Judgment under uncertainty: Heuristics and biases. In: *Science*, 185, 1124–1131.

Weaver, W. (1963): Lady Luck: The Theory of Probability. New York: Dover.

Anschrift der Verfasser

Ruma Falk; Avital Lavie Lann
The Hebrew University of Jerusalem, Israel
rfalk@cc.huji.ac.il
avital.a8@gmail.com