Hans Humenberger, Wien

Das „BENFORD-Gesetz" über die Verteilung der ersten Ziffer von Zahlen[1]

Zusammenfassung

Uns ist sehr wohl bewußt, daß es in der Schule nicht (oder nur kaum — evtl. in Leistungskursen mit Freiwilligen) möglich sein wird, die doch in den Anfangsgründen der Maßtheorie beheimateten Inhalte zu behandeln. Trotzdem scheinen uns auch diese einfachen maßtheoretischen Überlegungen wichtig und interessant zu sein, z.B. für Lehrerfortbildungen oder für Studierende des Mathematik-Lehramtes, selbst wenn keine wirkliche Vorbildung in dieser Richtung vorliegt. Für eine allfällige Behandlung des Themas in der Schule sei angemerkt, daß wir es doch für möglich und sinnvoll halten, in *Einzelfällen* (!) gewisse Ausblicke auf die Hochschulmathematik (bzw. auf wirkliche mathematische Forschungsinhalte — hier speziell der 60er-Jahre) in der Schule zu vermitteln, selbst wenn bei einem bestimmten Problem vieles nur erzählt werden kann und die dahintersteckende Mathematik eher im Hintergrund bleibt bzw. bleiben muß. Eine Einführung (Schilderung des Problems, geometrische Folgen etc.) enthält aber nur Elementarmathematik und ist insofern u.E. auch als Schulstoff in Leistungskursen denkbar. Die eigentlichen „Zielgruppen" dieses Themas sind jedoch nicht die Schulklassen des Pflichtunterrichts, sondern die Lehrerausbildung (Mathematikstudenten), die Lehrerfortbildung (Seminare), Facharbeiten und Projekte mit interessierten Schülern etc.

1 Einleitung

Die Geschichte des BENFORD-Gesetzes begann mit Beobachtungen von Logarithmentafeln, und zwar berichtete der Physiker Frank BENFORD (1938) - nach ihm wurde das resultierende Gesetz benannt - , daß die Logarithmentafeln in den Bibliotheken auf den ersten Seiten viel dreckiger und abgegriffener wären als auf den hinteren. Dies wäre bei anderen Büchern als Logarithmentafeln in Bibliotheken durchaus erklärbar, denn viele Leute beginnen ein Buch zu lesen (Roman, Gedichte, Theaterstück, Kurzgeschichten, Sachbücher, Fachbücher etc.), hören aber vorzeitig damit wieder auf, weil sie keine Zeit mehr haben, weil es ihnen zu langweilig wird, weil es ihnen zu kompliziert wird (Fachbücher) u.ä. Wenn viele die Lektüre unfertig unterbrechen, ist es klar, daß der Anfang von Büchern abgenützter sein kann als der Schluß. Aber warum soll dies bei Logarithmentafeln der Fall sein - diese werden ja nach anderen Gesichtspunkten benützt. Die einzige Erklärung, die es dafür gibt, ist, daß der Logarithmus von Zahlen mit niedrigen Anfangsziffern (1,2,...) häufiger gesucht wurde als von Zahlen mit hohen Anfangsziffern (9,8,...)! Aber warum? Kommen Zahlen mit niedrigen Anfangsziffern „in der Welt" häufiger vor? Warum sollte die Natur eine Präferenz für die 1 als Anfangsziffer haben?

[1]Das ausgearbeitete Manuskript des Vortrages ist erschienen in *Stochastik in der Schule* **16**, 3, 2–17.

Es sind schon viele empirische Daten erhoben worden, wobei die relative Häufigkeit der einzelnen Anfangsziffern beobachtet wurde. Diese müßte bei Gleichverteilung für alle möglichen Anfangsziffern (1,2,...,8,9) bei ca. $\frac{1}{9} \approx 0,1111$ liegen - mit *Anfangsziffer* sei im folgenden stets *die erste Ziffer ungleich 0* bzw. *erste „signifikante" Ziffer* gemeint, also z.B. 3 in 0,0367. Tatsächlich lag jedoch die relative Häufigkeit von 1 als Anfangsziffer in vielen Datensätzen (insbesondere in jenen von BENFORD) bei ca. 0,3 - abnehmend zu ca. 0,05 bei Ziffer 9.

Es sind z.B. untersucht worden:
Oberflächen vieler Seen, Hausnummern in den Adressen vieler Personen, Halbwertszeiten radioaktiver Substanzen, Energieverbrauchszahlen vieler Haushalte u.v.a.m.

Bemerkung: Es hat natürlich keinen Sinn, Daten zu betrachten, die von vornherein auf einen Bereich eingeschränkt sind, der die Möglichkeiten für die erste Ziffer ziemlich einengt - dies wären z.B. die Anzahl der Buchstaben in den Familiennamen der Bewohner einer Stadt oder eines Landes, das Alter von Studierenden an einer Universität (das Alter generell!), die Anzahl der Schulbildungsjahre, die Anzahl der Sitze in Fahrzeugen, die Wurzeln der ersten 1 000 natürlichen Zahlen usw.

Bei den meisten der Untersuchungen hat sich eine abnehmende relative Häufigkeit von 1 bis 9 als Anfangsziffer ergeben. Wenn obige Werte „tatsächlich stimmen" (ungefähr die theoretischen Wahrscheinlichkeiten darstellen, d.h. $P(1) \approx 0,3$ und $P(9) \approx 0,05$), ist es einleuchtend, daß bei einer 9-seitigen Logarithmentafel die erste Seite abgenützter ist als die letzte (ca. sechsmal so stark!).

I. STEWART (1994, S.1) berichtet sogar von einem Jahrmarktspiel mit einem an Daten reichen Computer: Der Spieler kann selber irgendwelche Daten anklicken (z.B. Einwohnerzahlen von Verwaltungsbezirken auf den Malediven oder in Lettland, Energieverbrauch in den Ländern der Welt, und viele andere „exotische" - jedenfalls i.a. nicht vorhersehbare Werte). Entscheidend für den Spielausgang ist die erste Ziffer der erscheinenden Zahl. Ist diese 1 oder 2, dann gehören die 10 DM Einsatz dem Spielbudenbesitzer, bei 3, 4, 5, 6, 7, 8 und 9 erhält der Spieler seine 10 DM Einsatz zurück plus 5 DM Gewinn. Auf den ersten Blick sehen die Gewinnchancen für den Spieler nicht schlecht aus, wenn man die Wahrscheinlichkeiten mit jeweils $\frac{1}{9}$ annimmt. Der Erwartungswert des Gewinnes für den Spieler beträgt dann $-10 \cdot \frac{2}{9} + 5 \cdot \frac{7}{9} = \frac{15}{9} \approx 1,66$ DM. Wenn der Spielbudenbesitzer wirklich auf lange Sicht 1,66 DM pro Spiel bezahlen müßte, so wäre dies für ihn nicht besonders lukrativ und sein Bankrott gleichsam vorprogrammiert. Aber der Spielbudenbesitzer vertraut offenbar auf das Gesetz von BENFORD, demzufolge die Wahrscheinlichkeit für das Auftreten einer Ziffer p $(1 \leq p \leq 9)$ als erste Ziffer den Wert $\log_{10}(p+1) - \log_{10} p$ hat. In weiterer Folge wird die Basis 10 des Logarithmus weggelassen, da nie eine andere Basis auftritt, also $\log \stackrel{\text{def}}{=} \log_{10}$. Danach hätten also die einzelnen Ziffern die in Tab. 1 angegebenen Wahrscheinlichkeiten (für das Auftreten als erste Ziffer).

Wenn diese Wahrscheinlichkeiten zur Berechnung des Gewinnerwartungswertes $E(G)$ obigen Spielers herangezogen werden, so ergibt sich

$$E(G) = -10\,(0,301 + 0,176) + 5\,(0,125 + 0,097 + \ldots + 0,051 + 0,046) = -2,155\ .$$

Unter dieser Voraussetzung ist das wirtschaftliche Überleben der Spielbude natürlich relativ gesichert (wenn sich genügend Kunden darauf einlassen).

1	2	3	4	5	6	7	8	9
0,301	0,176	0,125	0,097	0,079	0,067	0,058	0,051	0,046

Tabelle 1: Wahrscheinlichkeiten für die einzelnen Ziffern nach BENFORD

2 Wahrscheinlichkeitswerte in Abhängigkeit von der Grundgesamtheit der Daten

Zunächst ist klar, daß die Wahrscheinlichkeiten für die einzelnen Ziffern, als erste Ziffer einer Zufallszahl zu stehen, von der Grundgesamtheit des „Topfes" abhängen, aus dem die Zahl zufällig gezogen wird. Wenn z.B. aus den ersten 20 natürlichen Zahlen zufällig gezogen wird, so ist offenbar die 1 als erste Ziffer ziemlich übermächtig (11 von 20 möglichen), 2 steht in zwei von 20 Fällen an erster Stelle und jede andere Ziffer $(p \neq 0)$ genau einmal!

Wir erkennen auch sofort, daß bei natürlichen Zahlen die Wahrscheinlichkeit $\frac{1}{9}$ nur dann bei jeder Ziffer $p = 1, 2, \ldots, 9$ auftritt, wenn die Grundgesamtheit aus den ersten 9, 99, 999, 9 999 usw. Zahlen besteht. Verfolgen wir z.B. einmal die Wahrscheinlichkeit von 1 als erste Ziffer, wenn der „Topf" aus den ersten n natürlichen Zahlen besteht, d.h. wir betrachten die Folge $(P_n(1))_{n \in \mathbb{N}}$:

Bei $n = 1$ ist $P(1) = 1$, bei $n = 2$ ist $P(1) = \frac{1}{2}$ usw., diese Wahrscheinlichkeit sinkt dann bis $P(1) = \frac{1}{9}$ bei $n = 9$. Dann steigt die Wahrscheinlichkeit wieder bis $n = 19$ (auf $P(1) = \frac{11}{19}$), um dann wieder bis $n = 99$ abzufallen (wieder $P(1) = \frac{1}{9}$), dann kommt natürlich wieder ein Anstieg bis $n = 199$ (auf $P(1) = \frac{111}{199}$), dem wieder ein Abfall bis $n = 999$ folgt. So geht dies natürlich „ewig" (von 10er-Potenz zu 10er-Potenz) weiter, die relative Häufigkeit wird sich nie (mit wachsendem n) bei einem stabilen Wert einpendeln, es werden nur die Phasen des Anstiegs bzw. des Abfalles klarerweise länger, die Folge $(P_n(1))_{n \in \mathbb{N}}$ ist also sicher divergent.

Einfach durch Grenzwertbildung $(\lim_{n \to \infty})$ kann man also nicht zur gesuchten Wahrscheinlichkeit $P(1)$ in ganz $\mathbb{N}$ kommen, denn $P_n(1)$ schwankt immer zwischen $\frac{1}{9}$ (Untergrenze) und einer nur fast gleichbleibenden Obergenze $(\frac{11}{19}, \frac{111}{199}, \frac{1111}{1999}, \ldots)$.

Durch fortgesetzte kumulative Mittelwertbildung - sicherlich kein Thema für den Schulunterricht - konnte z.B. B.J. FLEHINGER (1966) zeigen, daß es in einem gewissen Sinn der Wert $\log 2$ ist, um den $P_n(1)$ schwankt. Bei ihrem Modell wird der ganze Bereich aller möglichen (positiven) Konstanten der Welt (positive reelle Zahlen) durch $\mathbb{N}$ repräsentiert, da für das Problem der ersten Ziffer, die Stellung des Kommas keine Bedeutung habe und jede positive relle Zahl in Dezimalschreibweise bei Weglassen des Kommas einer natürlichen Zahl entspreche. So argumentierte FLEHINGER beim nach BENFORD benannten Gesetz des Logarithmus für alle denkbaren positiven rellen Zahlen (Konstanten).

Zur Urheberschaft bzw. Namensgebung schreibt RAIMI allerdings (1976, S.522), daß BENFORD zwar dieses Problem der Verteilung der ersten Ziffer berühmt gemacht hat, daß aber schon 57 Jahre vor ihm (1881) Simon NEWCOMB dieses „Gesetz" beschrieben hat. Es ist also in einem gewissen Sinn zu unrecht nach BENFORD benannt - ein Phänomen, das ja häufig auftritt!

Am Beispiel bestimmter *geometrischer Folgen* ist das BENFORD-Gesetz besonders gut und einfach zu veranschaulichen, obwohl dies natürlich noch nichts zur Klärung beiträgt, warum auch die anderen Zahlen (Konstanten) das logarithmische Gesetz befolgen sollten - vgl. RAIMI (1969b, S.110f).

Für eine genauere Darstellung geometrischer Folgen und der allgemeinen mathematischen Hintergründe kann hier nur auf die ausführlichere Version dieses Aufsatzes und z.B. auf die unten zitierte Literatur verwiesen werden.

Literatur:

1. BENFORD, F. (1938): The law of anomalous numbers. In: *Proceedings of the American Philosophical Society* **78**, 551–572.
2. BUCK, B.; A.C. MERCHANT; S.M. PEREZ (1993): An Illustration of Benford's first digit law using alpha decay half lives. In: *European Journal of Physics* **14**, 59–63.
3. FLEHINGER, B.J. (1966): On the probability that a random integer has initial digit *a*. In: *American Mathematical Monthly* **73**, 1056–1061.
4. HUMENBERGER, H. (1996): Das „BENFORD-Gesetz" über die Verteilung der ersten Ziffer von Zahlen. In: *Stochastik in der Schule* **16**, 3, 2–17. „Eine Ergänzung zum BENFORD-Gesetz — weitere mögliche schulrelevante Aspekte" ist bei derselben Zeitschrift zur Publikation eingereicht.
5. KRÄMER, W. (1990): Das Gesetz der abnormalen Zahl. In: *Stochastik in der Schule* **10**, 3, 48–52.
6. PINKHAM, R.S. (1961): On the distribution of first significant digits. In: *Annals of Mathematical Statistics* **32**, 1223–1230.
7. RAIMI, R.A. (1969a): On the distribution of first significant figures. In: *American Mathematical Monthly* **76**, 342–348.
8. RAIMI, R.A. (1969b): The peculiar distribution of first digits. In: *Scientific American* **221**, Dezember 1969, 109–120.
9. RAIMI, R.A. (1976): The first digit problem. In: *American Mathematical Monthly* **83**, 521–538.
10. ROGERS, C.A. (1964): Packing and Covering. Cambridge.
11. STEWART, I. (1994): Mathematische Unterhaltungen. In: *Spektrum der Wissenschaft* (April 1994), 16–20.

Anschrift des Verfassers:

Hans HUMENBERGER, Institut für Mathematik und Angewandte Statistik, Universität für Bodenkultur, Gregor Mendel-Straße 33, A - 1180 Wien.