Das „BENFORD-Gesetz" — warum ist die Eins als führende Ziffer von Zahlen bevorzugt?

Hans Humenberger, Wien

In diesem Aufsatz wird gezeigt, wie eine Behandlung des in letzter Zeit immer populärer werdenden BENFORD-Gesetzes (siehe die neuen populärwissenschaftlichen Literaturzitate von DWORSCHAK 1998 oder MATTHEWS 1999) mit elementaren Mitteln möglich erscheint. Es wäre dabei auch durchaus vorstellbar, daß die doch etwas komplexeren mathematischen Begründungen in Abschnitt 4.3 weggelassen werden. Eine „außermathematische Anwendung" des prima vista vielleicht höchst theoretisch scheinenden Gesetzes wurde durch Mark NEGRINI realisiert, der mittels dieses Gesetzes Steuersündern auf die Spur gekommen ist. Internationale Konzerne und Finanzbehörden interessieren sich mittlerweile für das Programm (Software) von M. NEGRINI (Universität Halifax).

1 Einleitung

[1] 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 1 als Anfangsziffer haben?

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 in vielen Datensätzen (insbesondere in jenen von BENFORD) die relative Häufigkeit von 1 als Anfangsziffer 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 — 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

[1] Dieser Aufsatz ist eine kombinierte und überarbeitete Version von HUMENBERGER 1996 und 1997.