ISTRON - Materialien für einen realitätsbezogenen Mathematikunterricht - Band 6 (2000) S. 138-150

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.

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 oder 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

$$\boxed{P(1.\ \text{Ziffer} = p) = \log_{10}(p+1) - \log_{10} p}$$

beträgt; dies heißt auch „logarithmisches Gesetz“ und wird oft in folgender Form[2] geschrieben: $\boxed{P(1.\ \text{Ziffer} \leq p) = \log_{10}(p+1)}$.

Danach hätten die einzelnen Ziffern die in Tab. 1 angegebenen Wahrscheinlichkeiten (Auftreten als 1. Ziffer).

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

$$\begin{aligned} E(G) = \; & -\; 10\,(0.301 + 0.176) + \\ & +\; 5\,(0.125 + \ldots + 0.046) = -2.155\,. \end{aligned}$$

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

Bemerkung: Zur Urheberschaft bzw. Namensgebung des Gesetzes („BENFORD-Gesetz“) schreibt RAIMI (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!

[2] In weiterer Folge wird die Basis 10 des Logarithmus weggelassen, da nie eine andere Basis auftritt, also $\log \stackrel{\text{def}}{=} \log_{10}$.

2 Die Wahrscheinlichkeiten $P_n(p)$ bei den ersten n natürlichen Zahlen als mögliche Zufallszahlen

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 $P_1(1) = 1$, bei $n = 2$ ist $P_2(1) = \frac{1}{2}$ usw., diese Wahrscheinlichkeit sinkt dann bis $P_9(1) = \frac{1}{9}$ bei $n = 9$. Dann steigt die Wahrscheinlichkeit wieder bis $n = 19$ (auf $P_{19}(1) = \frac{11}{19}$), um dann wieder bis $n = 99$ abzufallen (wieder $P_{99}(1) = \frac{1}{9}$), dann kommt natürlich wieder ein Anstieg bis $n = 199$ (auf $P_{199}(1) = \frac{111}{199}$), dem wieder ein Abfall bis $n = 999$ folgt (siehe Tab. 2 und zur Veranschaulichung Fig. 1).

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 sicher divergent. Einfach durch Grenzwertbildung $(\lim_{n\to\infty})$ kann man also nicht zur gesuchten Wahrscheinlichkeit „$P(1)$“ in ganz $\mathbb{N}$ kommen (falls diese überhaupt existiert), denn $P_n(1)$ schwankt immer zwischen $\frac{1}{9}$ (Untergrenze!) und einer nur fast gleichbleibenden Obergrenze $O_m(1)$ $(\frac{11}{19}, \frac{111}{199}, \frac{1111}{1999}, \ldots)$; bei der m-ten Obergrenze O_m bezeichne m die Anzahl der Ziffern in Zähler und Nenner. Schon an dieser Stelle ist plausibel, daß $P(1) > \frac{1}{9}$ ist, da für „fast alle“ $n \in \mathbb{N}$ der entsprechende Wert $P_n(1) > \frac{1}{9}$ ist (nur bei $n = 10^\ell - 1$

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

Tab. 1: Wahrscheinlichkeiten für die einzelnen Ziffern nach Benford

n	1	9	19	99	199	999	1 999	9 999	19 999
$P_n(1)$	$\frac{1}{1}$	$\frac{1}{9}$	$\frac{11}{19}$	$\frac{11}{99}=\frac{1}{9}$	$\frac{111}{199}$	$\frac{111}{999}=\frac{1}{9}$	$\frac{1\,111}{1\,999}$	$\frac{1\,111}{9\,999}=\frac{1}{9}$	$\frac{11\,111}{19\,999}$

Tab. 2: Wahrscheinlichkeiten $P_n(1)$: Obergrenzen $O_m(1) = \frac{11...1}{19...9}$ und Untergrenze $\frac{1}{9}$

ist $P_n(1) = \frac{1}{9}$). Salopp formuliert: nach jeder neu dazukommenden Stelle wird die Ziffer 1 lange Zeit „bevorzugt" (zwischen $10\ldots0$ und $19\ldots9$), bevor die anderen Ziffern der Reihe nach diesen Rückstand wieder aufholen.

In Fig. 1 ist der Verlauf von $P_n(1)$ als kontinuierliche Funktion in Abhängigkeit von n graphisch dargestellt, so daß die oben schon angesprochenen Phasen des Anstieges bzw. Abfalles von $P_n(1)$ auch *sichtbar* werden. (Die n-Achse wurde logarithmisch skaliert, damit die Abstände zwischen zwei aufeinanderfolgenden Zehnerpotenzen nicht immer größer werden, sondern gleich bleiben; die Werte $P_n(1)$ „schwanken" in einem gewissen Sinn um den Wert $\log 2$.)

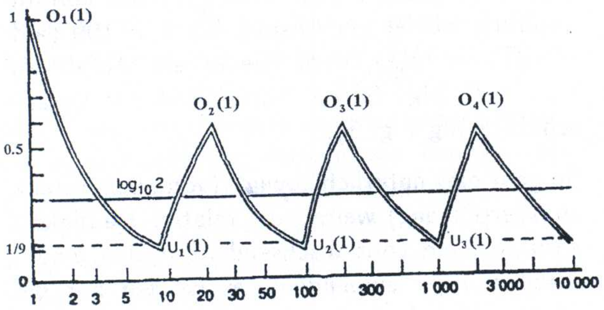

Fig. 1: Graphische Darstellung des Verlaufes von $P_n(1)$ als kontinuierliche Funktion von n

Die Obergrenzen $O_m(1)$ könnten auch im Schulunterricht näher studiert werden. Was passiert mit diesen Obergrenzen? Existiert ein Grenzwert für $m \to \infty$? Eine Frage, die durchaus von Schülern selbständig bearbeitet werden kann (Vernetzung: Grenzwert von Folgen), indem sie die Zahlen $O_m(1) = \frac{11...1}{19...9}$ näher betrachten. Sie sollten z. B. in der Lage sein zu beweisen, daß die Folge dieser Obergrenzen $O_m(1)$ monoton fallend und nach unten beschränkt ist (z. B. durch 0 oder sogar durch $\frac{1}{2}$), wodurch die Konvergenz gesichert wäre. Die Schüler werden auch schnell (z. B. durch Probieren, d. h. Berechnen einiger Werte mit dem Taschenrechner oder Computer) auf die Vermutung kommen, daß es der Wert $0.\dot{5} = \frac{5}{9}$ ist, dem sich diese Zahlen von oben nähern. Diese Tatsache kann nun durch direktes Einsetzen in die Grenzwertdefinition auch bewiesen werden. Dafür ist zu zeigen, daß es für alle $\varepsilon > 0$ ein $M(\varepsilon)$ gibt, so daß für alle $m > M(\varepsilon)$ die Beziehung $|\frac{11...1}{19...9} - \frac{5}{9}| < \varepsilon$ gilt; diese ist für $19\ldots9 > \frac{4}{9\cdot\varepsilon}$ erfüllt (so ist der Grenzwert *ohne* explizite Termdarstellung der Zahlen $O_m(1) = \frac{11...1}{19...9}$ bestimmbar).

Einfacher ist die Grenzwertbestimmung, wenn man einen geschlossenen Term für die Zahlen $O_m(1)$ findet. Auch dies sollte für die Schüler kein Problem darstellen: die Zähler $11\ldots1$ (m Stellen) können z. B. als $\frac{10^m-1}{9}$ und die Nenner $19\ldots9$ z. B. durch $\frac{1}{5}\cdot 10^m - 1$ oder durch $2\cdot 10^{m-1} - 1$ dargestellt werden, wodurch man für $O_m(1)$ die geschlossene Darstellung $O_m(1) = \frac{5}{9}\cdot\frac{10^m-1}{10^m-5}$ und damit den Grenzwert $\frac{5}{9}$ leicht finden kann.

Die Ober- und Untergrenzen der anderen Ziffern 2,3,. . . ,9

Betrachten wir als nächstes z. B. die Ziffer 9. Verfolgen wir analog die Werte $(P_n(9))_{n\in\mathbb{N}}$, d. h. die Wahrscheinlichkeiten, daß die erste Ziffer einer Zahl 9 ist, wenn aus den ersten n natürlichen Zahlen zufällig gezogen wird:

Bei $n = 1,\ldots,8$ ist $P_n(9) = 0$, bei $n = 9$ ist $P_9(9) = \frac{1}{9}$, diese Wahrscheinlichkeit sinkt dann bis $P_{89}(9) = \frac{1}{89}$ bei $n = 89$. Dann steigt die Wahrscheinlichkeit wieder bis $n = 99$ (auf $P_{99}(9) = \frac{11}{99} = \frac{1}{9}$), um dann wieder bis $n = 899$ abzufallen auf $P_{899}(9) = \frac{11}{899}$), dann kommt wieder eine hundert Zahlen lange Aufholphase bis $n = 999$ auf $P_{999}(9) = \frac{111}{999} = \frac{1}{9}$), der wieder ein Abfall bis $n = 8\,999$ folgt usw. (siehe Tab. 3 und Fig. 2). $P_n(9)$ schwankt immer zwischen $\frac{1}{9}$ (Obergrenze!) und der Untergrenze $U_m(9) = 0, \frac{1}{89}, \frac{11}{899}, \frac{111}{8\,999}, \frac{1\,111}{89\,999}, \ldots$. Es ist auch hier plausibel, daß für $P(9)$ (falls existent) $P(9) < \frac{1}{9}$ gilt, da für „fast alle" $n \in \mathbb{N}$ der entsprechende Wert $P_n(9) < \frac{1}{9}$ ist (nur bei $n = 10^\ell - 1$ ist $P_n(9) = \frac{1}{9}$). Salopp formuliert: nach jeder neu dazukommenden Stelle (also nach $10\ldots0$) hat die Ziffer 9 lange Zeit „Rückstand", bis sie unmittelbar vor der nächsten dazukommenden Stelle den Rückstand wieder aufholt (zwischen $90\ldots0$ und $99\ldots9$). Die Ziffer 9 hat also nie einen „Vorteil" gegenüber einer anderen, sie kann ihren Rückstand nur manchmal wettmachen!

Die Folge der Untergrenzen $U_m(9) = \frac{1\ldots1}{89\ldots9}$ (m sei die Anzahl der Einsen im Zähler) ist monoton wachsend und nach oben beschränkt, sie konvergiert gegen den Wert $\frac{1}{81}$, wie man durch analoge Überlegungen wie oben feststellen kann (explizite Darstellung: $U_m(9) = \frac{1}{81} \cdot \frac{10^m-1}{10^m-1/9}$).

In Fig. 2 sieht man nochmals deutlich: $\frac{1}{9}$ ist der *kleinste* Wert im Verlauf von $P_n(1)$ („das schlimmste, was der Ziffer 1 passiert"), während $\frac{1}{9}$ der *größte* Wert im Verlauf von $P_n(9)$ ist („das beste, was der Ziffer 9 passiert"), so daß in Fig. 2 die Relationen $P(1) > \frac{1}{9}$, $P(9) < \frac{1}{9}$ und $P(1) \gg P(9)$ auch „sichtbar" sind.

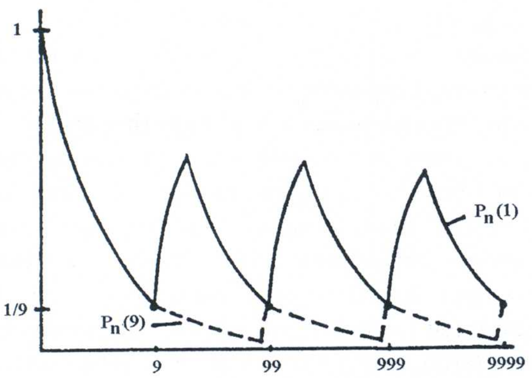

Fig. 2: Gleichzeitige graphische Darstellung der Verläufe von $P_n(1)$ und $P_n(9)$ als kontinuierliche Funktionen von n

Während bei der Ziffer 1 die Untergrenze konstant $U_m(1) = \frac{1}{9}$ war und bei der Ziffer 9 dies die konstante Obergrenze $O_m(9)$ war, sind bei den anderen Ziffern $p = 2, \ldots, 8$ die Werte der jeweiligen Ober- bzw. Untergrenzen $O_m(p)$ bzw. $U_m(p)$ nicht mehr konstant. Tab. 4 gibt einen Überblick über die entsprechenden Werte bei der Ziffer 3.

Die Schüler können ganz selbständig die Form der Unter- bzw. Obergrenzen für jede beliebige Ziffer p erarbeiten: ganz allgemein haben diese für die Ziffer p die Gestalt (m sei dabei die Anzahl der Einsen im Zähler):

$$U_m(p) = \frac{1\ldots1}{(p-1)9\ldots9} = \frac{1}{9\,p} \cdot \frac{10^m - 1}{10^m - \frac{1}{p}} \quad \text{und}$$

$$O_m(p) = \frac{11\ldots1}{p9\ldots9} = \frac{10}{9\,(p+1)} \cdot \frac{10^m - 1}{10^m - \frac{10}{p+1}} .$$

Die Schüler sollten erkennen (und begründen) können, daß die Obergrenzen $O_m(p)$ monoton fallend und die Untergrenzen $U_m(p)$ monoton wachsend sind, wobei auch deren Grenzwerte für $m \to \infty$ nicht schwierig zu finden sind (durch die explizite Darstellung der Zähler und Nenner und somit der Werte $U_m(p)$ und $O_m(p)$ selbst):

$$U(p) \stackrel{\text{def}}{=} \lim_{m\to\infty} U_m(p) =$$

$$= \lim_{m\to\infty} \left(\frac{1}{9\,p} \cdot \frac{10^m - 1}{10^m - \frac{1}{p}} \right) = \frac{1}{9\,p} ,$$

und

$$O(p) \stackrel{\text{def}}{=} \lim_{m\to\infty} O_m(p) =$$

$$= \lim_{m\to\infty} \left(\frac{10}{9\,(p+1)} \cdot \frac{10^m - 1}{10^m - \frac{10}{p+1}} \right) = \frac{10}{9\,(p+1)} .$$

Die so definierten (Grenz-)Werte $U(p)$ bzw. $O(p)$ sind natürlich Schranken für die Wahrscheinlichkeiten $P(p)$, falls diese überhaupt existieren. Man sieht (vgl. auch Tab. 5), daß die Intervalle $[U(p)\ ,\ O(p)]$ für alle $p = 1, \ldots, 9$ den Wert $\frac{1}{9}$ enthalten, mit den beiden Extremfällen: bei $p = 1$ als konstante Untergrenze und bei $p = 9$ als konstante Obergrenze — bei $n = 9,\ 99,\ 999\ldots$ ist ja $P_n(p) = \frac{1}{9}$ für alle $p = 1, \ldots, 9$ (die Wahrscheinlichkeiten $P_n(p)$ „kehren mit wachsendem n immer wieder zum Wert $\frac{1}{9}$ zurück").

Als Schätzwert für $P(p)$ böte sich in erster Näherung z. B. der jeweilige Mittelwert $P(p) \approx \frac{U(p)+O(p)}{2}$ an, eine Schätzung, die – wie wir sehen werden – gar nicht so schlecht ist. Tab. 5 gibt einen Überblick über die jeweiligen Werte von $U(p)$ und $O(p)$ (als Intervallgrenzen) in exakter Form und in Dezimaldarstellung (auf drei

n	8	9	89	99	899	999	8 999	9 999	89 999
$P_n(9)$	0	$\frac{1}{9}$	$\frac{1}{89}$	$\frac{11}{99}=\frac{1}{9}$	$\frac{11}{899}$	$\frac{111}{999}=\frac{1}{9}$	$\frac{111}{8\,999}$	$\frac{1\,111}{9\,999}=\frac{1}{9}$	$\frac{1\,111}{89\,999}$

Tab. 3: Wahrscheinlichkeiten $P_n(9)$: Obergrenze $\frac{1}{9}$ und Untergrenzen $U_m(9) = \frac{1...1}{89...9}$

n	2	3	29	39	299	399	2 999	3 999	29 999
$P_n(3)$	0	$\frac{1}{3}$	$\frac{1}{29}$	$\frac{11}{39}$	$\frac{11}{299}$	$\frac{111}{399}$	$\frac{111}{2\,999}$	$\frac{1\,111}{3\,999}$	$\frac{1\,111}{29\,999}$

Tab. 4: Wahrscheinlichkeiten $P_n(3)$: Obergrenzen $O_m(3) = \frac{11...1}{39...9}$ und Untergrenzen $U_m(3) = \frac{1...1}{29...9}$

Dezimalen gerundet); weiters ist der jeweilige Intervallmittelpunkt $\frac{U(p)+O(p)}{2}$ angegeben und der wirkliche Wert $P(p)$ (auf drei Dezimalen gerundet; siehe Tab. 1). Ein Vergleich dieser beiden Werte zeigt, daß der jeweilige Schätzwert durch den Intervallmittelpunkt eine Differenz zum jeweils exakten Wert von nur 0.016 (bei $p = 9$) bis 0.037 (bei $p = 2$) aufweist.

Ein möglicher Schritt zur Verbesserung der Genauigkeit der durch die Intervallmittelpunkte gegebenen Schätzwerte könnte folgender sein: Da die Summe dieser Schätzwerte (vierte Spalte in Tab. 5) nicht 1, sondern 1.229 ergibt, so ist es naheliegend, alle diese Werte durch 1.229 zu dividieren, was bei $p = 2, \ldots, 9$ die Genauigkeit beträchtlich und bei $p = 1$ auch ein wenig erhöhte (diese Werte sind die Werte in Klammer in der vierten Spalte von Tab. 5). Die Differenz zum exakten Wert beträgt dann nur 0.003 bis 0.006 bei $p = 2, \ldots, 9$ und 0.030 bei $p = 1$. So könnten plausible und passable Schätzwerte für die Wahrscheinlichkeiten $P(p)$ gewonnen werden — und zwar *a priori*, ohne folgende (oder andere) Überlegungen, die den mathematischen Hintergrund erst etwas erleuchten.

Nun verlassen wir den Bereich der natürlichen Zahlen und wenden uns den (positiven) reellen Zahlen zu (als „Topf, aus dem eine Zufallszahl gezogen wird"). Zunächst betrachten wir Größen, die *zeitlichen Veränderungen* unterworfen sind; dazu ein Zitat, das (nicht mathematisch) plausibel macht, warum die 1 als führende Ziffer dort ein Übergewicht hat: „Die Eins ist auf der Zahlenskala nicht weiter entfernt als die Fünf von der Sechs. Für die wirklichen Dinge allerdings, die gezählt, gemessen oder gewogen werden, kann der Weg von der Eins zur Zwei sehr lang sein: Um ihn zurückzulegen, müssen sie auf das Doppelte wachsen. Einer Fünf fehlt dagegen nur ein Fünftel, um zur Sechs zu werden. [...] Angenommen der Deutsche Aktienindex stünde gerade bei 1000 Punkten, dann müßten sich die Aktienkurse im Schnitt verdoppeln, ehe der DAX die 2000 erreicht. Solange bliebe die führende Eins erhalten, solange erschiene sie auf allen Listen. Stünde der DAX aber bei 5000 Punkten, so müßten die Werte nur noch um 20 Prozent steigen, ehe mit 6000 die Fünf als erste Ziffer abgelöst wird. Noch kleiner ist im Verhältnis der Schritt von 9000 auf 10 000. [...] Was wächst oder schrumpft, verharrt deshalb relativ lang im Bereich der führenden Eins." (DWORSCHAK 1998, S.229).

von der Zwei

In diesem Spiegel-Artikel berichtet DWORSCHAK von einer echten *Anwendung* des Benford-Gesetzes (mögliche Antwort auf die Frage *Was nützt es, das Benford-Gesetz zu kennen oder gar zu verstehen?*): Auch Zahlen in umfangreichen Steuererklärungen gehorchen normalerweise dem logarithmischen Gesetz, außer es handelt sich um ausgedachte, von Menschenhand manipulierte Zahlenkolonnen. Ein amerikanischer Mathematiker (Mark Negrini) entwickelte eine Software, die Abweichungen von Benfords Gesetz in großen Datensätzen aufspürt, wodurch bei Steuersündern der Verdacht auf Manipulation erhoben werden kann und bei positivem Test genauere Untersuchungen sich anschließen können. Die Steuerbehörden mehrerer amerikanischer Bundesstaaten verwenden diese Software wirklich, aber auch etliche Großunternehmen haben ihre interne Buchhaltung damit ausgerüstet. Tatsächlich seien damit schon einige krumme Transaktionen in Amerika und Deutschland aufgedeckt worden!

p	$[U(p)\,,\,O(p)]$ Brüche	$[U(p)\,,\,O(p)]$ 3 Dezimalen	$(U(p)+O(p))/2$ 3 Dezimalen	$P(p)=\log(p+1)-\log p$ 3 Dezimalen
1	$\left[\frac{1}{9}\,,\,\frac{5}{9}\right]$	[0.111 , 0.555]	0.333 (0.271)	0.301
2	$\left[\frac{1}{18}\,,\,\frac{10}{27}\right]$	[0.056 , 0.370]	0.213 (0.173)	0.176
3	$\left[\frac{1}{27}\,,\,\frac{5}{18}\right]$	[0.037 , 0.278]	0.157 (0.128)	0.125
4	$\left[\frac{1}{36}\,,\,\frac{2}{9}\right]$	[0.028 , 0.222]	0.125 (0.102)	0.097
5	$\left[\frac{1}{45}\,,\,\frac{5}{27}\right]$	[0.022 , 0.185]	0.104 (0.085)	0.079
6	$\left[\frac{1}{54}\,,\,\frac{10}{63}\right]$	[0.019 , 0.159]	0.089 (0.072)	0.067
7	$\left[\frac{1}{63}\,,\,\frac{5}{36}\right]$	[0.016 , 0.139]	0.077 (0.063)	0.058
8	$\left[\frac{1}{72}\,,\,\frac{10}{81}\right]$	[0.014 , 0.123]	0.069 (0.056)	0.051
9	$\left[\frac{1}{81}\,,\,\frac{1}{9}\right]$	[0.012 , 0.111]	0.062 (0.050)	0.046

Tab. 5: Überblick über die Werte von $U(p)$ bzw. $O(p)$

Auch für Größen, die sich im Lauf der Zeit nicht wesentlich ändern, kann man intuitive Überlegungen anstellen: „Es gibt einfach mehr Pfützen als Tümpel, mehr Tümpel als Ozeane. Folglich gibt es wahrscheinlich auch mehr Gewässer zwischen 10 und 20 Hektar als zwischen 20 und 30, mehr zwischen 100 und 200 als zwischen 200 und 300 – und so fort. Damit ist Benfords Gesetz vollends auf dem Weg zur Weltenformel. Denn es gibt auch mehr Kieselsteine als Felsbrocken und überhaupt mehr kleine Dinge als große. Warum sich dies so verhält, ist wieder eine andere Frage.“ (DWORSCHAK 1998, S.229)

3 Geometrische Folgen

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).

Stellen wir uns eine geometrische Folge vor $(A\,q^n)_{n\in\mathbb{N}_0} = (A,\ A\,q,\ A\,q^2,\ A\,q^3,\ \ldots)$ mit $A, q > 0$ und eine *logarithmische* Skala, auf der wir die einzelnen Folgenglieder einzeichnen wollen. Der Abstand von 1 nach 2 ist dabei natürlich ein größerer als von 2 nach 3; der kleinste Abstand befindet sich zwischen 9 und 10 (bzw. 1) – wie z. B. bei einem Rechenstab. Auf dieser Skala denken wir uns irgendwo einen Startpunkt A, wobei wir beim Einzeichnen der weiteren Folgenglieder jeweils um $\log q$ nach rechts zu gehen haben (wenn $q > 1$, sonst nach links), denn es gilt ja $\log(A\,q^{n+1}) = \log(A\,q^n) + \log q$. Jede Multiplikation mit q „bewegt“ den vorigen Punkt um ein konstantes Stück nach rechts (bzw. links), nämlich um $\log q$. Kommt man hierbei an das Ende der Skala, d. h. über 10 bzw. unter 1, so kann man – eine 10er-Potenz höher (niedriger) – die Skala rechts (links) verlassen und links (rechts) wieder „betreten“. Um dieses Verlassen der Skala („Zurückspringen“) zu vermeiden, denken wir uns diese zu einem Kreis gebogen und die beiden Enden zusammengeschweißt („kreisförmiger Rechenstab“ – zur Veranschaulichung siehe Fig. 3).

Anhand der Fig. 3 ist es nicht schwierig, sich zu überzeugen, daß viele geometrische Folgen das Gesetz von BENFORD befolgen, wenn ihre Glieder bis zu einem hinreichend großen n betrachtet werden. Der Umfang des Kreises beträgt $\log 10 = 1$ und der Abschnitt z. B. zwischen 1 und 2 (dort liegen alle Werte, die mit 1 beginnen) hat Länge $\log 2 - \log 1 = \log 2$. Damit dies wirklich die Wahrscheinlichkeit darstellt, daß ein zufällig gewähltes Folgenglied mit Ziffer 1 beginnt, muß vorausgesetzt werden, daß die Folgenglieder *gleich dicht* in den einzelnen Abschnitten zu liegen kommen, daß

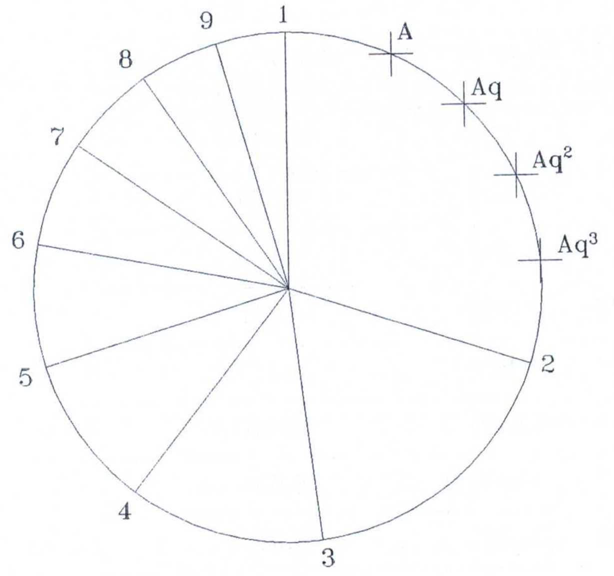

Fig. 3: Kreisförmige logarithmische Skala

kein Abschnitt von den Folgengliedern „bevorzugt" bzw. „vernachlässigt" wird – eine Art „geometrische Wahrscheinlichkeit", d. h. Länge des günstigen Intervalls durch die Länge des möglichen Intervalls (des Kreises mit Umfang 1). Ist z. B. $q = 10^{\frac{1}{3}}$, so werden überhaupt nur 3 verschiedene Punkte am Kreis durch die Folge erreicht, hier kann also von einer gleichen Dichte in den Abschnitten wohl nicht die Rede sein und diese Folge wird auch das BENFORD-Gesetz nicht befolgen. Ist hingegen $q = 10^{\varepsilon}$ mit $|\varepsilon| \ll 1$, so daß bei jeder Multiplikation mit q nur um ein winziges Stück (ε) weitergerückt werden muß, so kann man sich gut vorstellen, daß (bei einer ganzen Umrundung) die Folgenglieder (Punkte) in jedem „Abschnitt" gleich dicht liegen und daher die Wahrscheinlichkeiten für die jeweilige Anfangsziffer p eines zufällig gewählten Folgengliedes der jeweiligen Bogenlänge entspricht ($\log(p+1) - \log p$, also das BENFORD-Gesetz). Wenn man nicht nur ein Mal und auch nicht nur einige Male den Kreis umrundet, sondern viele Tausend oder sogar Millionen Mal, so wird es auch keine Rolle spielen, wenn die Folge nicht nach einer ganzen Umrundung abgebrochen wird (bei $1\frac{1}{2}$ oder $2\frac{1}{2}$ Umrundungen wäre die Dichte in den zwei verschiedenen Kreishälften ja wohl nicht gleich!).

Genauer: • Bei allen *rationalen* Potenzen von 10 als q (also $q = 10^{\frac{a}{b}}$ mit $a \in \mathbb{Z}$, $b \in \mathbb{N}$) wird der Ausgangspunkt A wieder erreicht (nach b Schritten bzw. a Umrundungen) und es entsteht immer wieder das gleiche „Muster". In diesem Fall nützt es daher auch nichts, nach der Rückkehr zum Ausgangspunkt, weitere Folgenglieder in den „Topf", aus dem eine Zahl zufällig gewählt wird, aufzunehmen – die Annäherung an das logarithmische Gesetz kann dadurch nicht besser werden!

• Bei *irrationalen* Potenzen von 10 muß jeweils um ein irrationales Stück weitergerückt werden, es kann daher kein Punkt zweimal „getroffen" werden, neue Punkte fallen immer zwischen zwei schon vorhandene! Weiters gilt in diesem Fall – dies ist intuitiv zwar einleuchtend, bedarf aber streng genommen eines relativ schwierigen Beweises – , daß die Dichte bei uneingeschränkter Fortführung der „Produktion" von Folgengliedern in jedem der fraglichen Abschnitte gleich groß wird, so daß die Annäherung an das logarithmische Gesetz immer besser wird.

Die unendlichen geometrischen Folgen erfüllen in den „meisten" Fällen das Gesetz von BENFORD, nämlich bei irrationalen Potenzen von 10 als Faktor – es gibt bekanntlich nur abzählbar viele rationale Zahlen im Gegensatz zu den überabzählbar vielen Irrationalzahlen (das LEBESGUE-Maß von $\mathbb{Q}$ ist Null!); bei z. B. hinreichend kleinen rationalen Werten $|\frac{a}{b}|$ wird das Gesetz ebenfalls in guter Näherung erfüllt!

Wenn also eine Zahl zufällig aus einer unendlichen geometrischen Folge gewählt wird, ist das Zustandekommen der ungleichen Verteilung der ersten Ziffer (in den meisten Fällen) durch obige Gedanken erklärt, aber warum soll es auch dann gelten, wenn man irgendeine reelle positive Konstante wählt?

STEWART (1994, S.20) schreibt dazu: „Wir Menschen zählen in arithmetischer Folge 1, 2, 3, ... und wundern uns, ungleiche Wahrscheinlichkeiten für die Anfangsziffern zu finden. Aber das läßt sich dadurch erklären, daß die Natur mit gleichen Wahrscheinlichkeiten unter den Termen einer geometrischen Folge wählt x, x^2, x^3, ...". Aber *warum* soll die Natur gleichsam *geometrisch* statt *arithmetisch* zählen?

Auch BENFORD selbst gab keine tiefen mathematischen Erklärungen dazu ab, sondern betrachtete es als eine Art Naturgesetz, das er und andere empirisch belegt hatten.

4 Mathematische Erklärungen

4.1 Das Modell von PINKHAM

Einer der Mathematiker, die sich mit diesem Thema beschäftigt haben, war Roger S. PINKHAM. Er betrachtete $\mathbb{R}^+$ als das potentielle Universum aller möglichen physikalischen Konstanten – auf das Problem, daß einige physikalische Werte negativ sein können (z. B. Temperaturen), gehen wir hier nicht ein. Wenn nun irgendein Gesetz der Verteilung der ersten Ziffer im Reich dieser möglichen Werte existieren sollte, so kann dies laut PINKHAM nur mit einem **Invarianzprinzip** verbunden sein: Die verwendeten *Einheiten* der physikalischen Größen sind ja nicht von der Natur vorgegeben, sondern höchst willkürlich von Menschen geschaffen. Es sollte also nicht sein, daß die Verteilung der ersten Ziffer in der Welt aller möglichen physikalischen Werte davon abhängt, ob die Entfernungen in Meter oder in Zoll gemessen werden, Geschwindigkeiten in Meilen pro Stunde oder in Meter pro Sekunde usw. Umwandlungen in eine andere Einheit entsprechen i. a. Multiplikationen mit einer gewissen positiven Zahl k. Die Werte einiger physikalischer Größen hängen aber nicht nur von der Größe der jeweiligen Einheiten ab, sondern auch von der Wahl des Nullpunktes, z. B. Temperatur. Auch die Transformation von z. B. Grad CELSIUS in FAHRENHEIT entspricht nicht einer reinen Multiplikation mit einem konstanten Faktor, sondern einer *affinen* Transformation. Beide Tatsachen machen keine großen Schwierigkeiten, sie sollen hier aber gänzlich ausgeklammert werden.

Mit der Forderung nach **Skaleninvarianz** und nach der *Existenz* einer Verteilungsfunktion $F : \mathbb{R}^+ \to [0\,,\,1]$ für alle möglichen physikalischen Konstanten bewies PINKHAM, daß die einzig mögliche Verteilungsfunktion für die erste Ziffer $V(p) = \log(p+1)$ ist (Wahrscheinlichkeit, daß die erste Ziffer $\leq p$ ist für $p = 1, 2, \ldots, 9$), also BENFORDs Gesetz.

Wir möchten hier nicht auf die Einzelheiten des Beweises von PINKHAM eingehen, sondern nur auf einen Widerspruch hinweisen, den RAIMI (1969, S.344) darin entdeckt hat. PINKHAMs Resultat wird zwar anerkannt, aber die Voraussetzung über die Existenz einer Verteilungsfunktion aller möglichen physikalischen Konstanten ist leicht zu widerlegen – dies ist ein sehr gutes Beispiel, daß ein Resultat trotz falscher Voraussetzungen richtig sein kann.

PINKHAM fordert die Existenz einer Verteilungsfunktion $F :\ \mathbb{R}^+ \to [0\,,\,1]$ mit

1) $\lim\limits_{x\to 0+} F(x) = 0$ und $\lim\limits_{x\to\infty} F(x) = 1$,

2) F ist monoton wachsend und *stetig* (sonst hätte ein einzelner Wert eine *positive* Auftrittswahrscheinlichkeit bei zufälliger Wahl einer Konstanten aus $\mathbb{R}^+$),

wobei dann F bekanntlich ein (abzählbar additives) Wahrscheinlichkeitsmaß Φ auf $\mathbb{R}^+$ induziert. Sei nämlich $[a, b) \subset \mathbb{R}^+$ (bzw. $(a, b] \subset \mathbb{R}^+$) ein halboffenes Intervall, dann kann $\Phi([a, b)) \stackrel{\text{def}}{=} F(b) - F(a) = \Phi((a, b])$ als Wahrscheinlichkeit interpretiert werden, daß eine zufällig gewählte positive reelle Zahl im Intervall $[a, b)$ bzw. $(a, b]$ liegt.

Betrachten wir zunächst einmal die Menge D_p aller positiven reellen Zahlen in Dezimalschreibweise, die eine Ziffer $\leq p$ als erste Ziffer haben. Diese Menge kann offenbar als folgende abzählbare Vereinigung paarweise disjunkter Mengen geschrieben werden

$$D_p = \bigcup_{n=-\infty}^{\infty} [\,10^n\,,\,(p+1)\,10^n\,) \quad p = 1, \ldots, 9.$$

Für die Wahrscheinlichkeit $P(x \in D_p) = \Phi(D_p)$ ergibt sich daher

$$\Phi(D_p) = \sum_{n=-\infty}^{\infty} (\,F((p+1)\,10^n) - F(10^n)\,)\ ,$$

ein Wert, der $\log(p+1)$ ergeben soll. Nun kommt die Skaleninvarianz ins Spiel. Die wohl einfachste Art, die Skaleninvarianz zu „modellieren“, ist zu verlangen, daß

$$\Phi(D_p) = \Phi(k \cdot D_p) \qquad \forall k \in \mathbb{R}^+$$

gilt, wobei $k \cdot D_p$ die Menge aller mit k multiplizierten Elemente von D_p ist:

$$k \cdot D_p \stackrel{\text{def}}{=} \bigcup_{n=-\infty}^{\infty} [\,k \cdot 10^n\,,\,k \cdot (p+1)\,10^n\,)\ .$$

Wenn ein physikalischer Wert in einem bestimmten Einheitensystem in D_p liegt, dann wird dieser Wert, wenn eine andere Skala zugrundegelegt wird (Multiplikation mit k), in $k \cdot D_p$ liegen. Skaleninvarianz muß also bedeuten, daß die Wahrscheinlichkeit, daß eine Zahl in D_p liegt, genauso groß ist wie die Wahrscheinlichkeit, daß eine Zahl in $k \cdot D_p$ liegt; das Maß von D_p muß dem Maß von $k \cdot D_p$ gleich sein: $P(x \in D_p) = P(x \in k \cdot D_p)$ bzw. $\Phi(D_p) = \Phi(k \cdot D_p)$ für $p = 1, 2, \ldots, 9$ und $k > 0$.

Aus der geforderten Skaleninvarianz folgert aber RAIMI (als Kritik an PINKHAMs Methode) sofort, daß eine Verteilungsfunktion F aller physikalischen Konstanten, wie sie PINKHAM postuliert, nicht existieren kann. Denn wenn die geforderte Skaleninvarianz (SI) in bezug auf die erste Ziffer gelten soll, d. h. für die Mengen D_p, dann scheint es vernünftig, zu fordern, daß diese Invarianzeigenschaft auch für beliebige (meßbare) Teilmengen $A \subset \mathbb{R}^+$ gelten soll, d. h. $P(x \in A) = \Phi(A) = \Phi(k \cdot A) = P(x \in k \cdot A)$ – mit welchem Recht könnte diese Eigenschaft auf die Mengen D_p eingeschränkt werden? Damit ergibt sich jedoch unmittelbar für alle $k > 0$

$$F(1) = P(x \in (0,1]) \overset{\text{SI}}{=} P(x \in (0,k]) = F(k) \,.$$

Das hieße jedoch, daß F konstant ist, was natürlich nicht sein kann!

RAIMI führt auch noch einen philosophischen Grund gegen die Existenz einer solchen Verteilungsfunktion aller physikalischen Konstanten an, der auch dann greift, wenn man nicht bereit ist, der angesprochenen Invarianzeigenschaft über die Mengen D_p hinaus Gültigkeit zu verleihen. Wenn so eine Verteilungsfunktion F existierte, so gäbe es auch einen Wert T mit $F(T) = P(x \leq T) = P(x \in (0,T]) = \frac{1}{2}$. D. h. es gäbe einen Wert T mit der Eigenschaft, daß er genau in der Mitte des Universums aller denkbaren Konstanten aller denkbaren Welten läge. RAIMI schreibt dazu: „T takes on a mystical significance. I cannot bring myself to believe in such a number, yet this is what we must believe when we admit that scale-invariance is not to be demanded for sets like $\{x \leq T\}$." (1969, S.344). Es wäre in der Tat auch u. E. sehr „merkwürdig", wenn ein T als *Median aller denkbaren Konstanten* existierte, so daß genau die Hälfte größer und die andere Hälfte kleiner als dieser wären.

Noch ein Argument: Wenn so eine stetige Verteilungsfunktion existierte, so gäbe es nicht nur den genannten Wert T mit $P(x \leq T) = P(0 < x \leq T) = \frac{1}{2}$, sondern auch ein (nicht notwendigerweise symmetrisches) endliches Intervall $I = (T - a\,,\ T + b]$ mit $a < T$ und $b < \infty$, so daß $P(x \in I) = \frac{1}{2}$ ist – statt des Wertes $\frac{1}{2}$ könnte hier jeder beliebige Wert $0 < c < 1$ (z. B. 0.8) stehen. Wenn jedoch alle denkbaren Einheiten „gleichberechtigt" sein sollten, so gäbe es „viel mehr" Einheiten nahe Unendlich und nahe Null – d. h. außerhalb jedes vorgegebenen endlichen Intervalls mehr als innerhalb desselben – , so daß die numerischen Einträge in die daraus resultierenden Listen „fast alle" nahe Null bzw. Unendlich wären. Obige Aussage $P(x \in I) = \frac{1}{2}$ stünde dazu offenbar im Widerspruch.

4.2 Eine für den Unterricht mögliche Plausibilitätsbetrachtung

Ein einfaches Beispiel: Die Lüftung eines Tunnels wird automatisch in Betrieb gesetzt und wieder ausgeschaltet; und zwar ist sie von jeder vollen Stunde an 25 Minuten lang in Betrieb (d. h. sie wird z. B. um 10.00 Uhr eingeschalten und um 10.25 Uhr wieder ausgeschaltet). Wie groß ist die Wahrscheinlichkeit, daß ein zu einem *zufälligen* Zeitpunkt[3] in den Tunnel einfahrendes Auto die Lüftung in Betrieb vorfindet?

Es wird hiebei wahrscheinlich vielen Schülern selbständig gelingen, zu einer Lösung zu kommen: innerhalb jeder vollen Stunde spielt sich dasselbe Szenario ab (25 von 60 Minuten Betrieb); wegen dieser offensichtlichen Periodizität wird man auch intuitiv den möglichen großen Stichprobenraum $\mathbb{R}$ einschränken auf $[0;1)$ (in Stunden) und dort die Wahrscheinlichkeit mit $\frac{25}{60}$ ausrechnen. Dabei ist ebenfalls intuitiv klar, daß es keine Rolle spielt, ob die Lüftung jeweils zu den vollen Stunden oder jeweils zu irgendeinem anderen fixen Zeitpunkt 25 Minuten lang pro Stunde eingeschalten wird (z. B. jeweils um *.15 Uhr, „Translationsinvarianz der Wahrscheinlichkeit").

Man wird also einsehen und es als plausibel erachten: die Wahrscheinlichkeit, daß ein zufällig gewählter Zeitpunkt (bzw. eine zufällig gewählte reelle Zahl) in der unbeschränkten Menge

$$K_{\frac{25}{60}} \overset{\text{def}}{=} \bigcup_{n=-\infty}^{\infty} \left[n\,,\ n + \frac{25}{60}\right)$$

liegt, beträgt $\frac{25}{60}$. Dasselbe[4] gilt auch für jeden anderen Wert $0 < a \leq 1$:

$$P(x \in K_a) = a \,.$$

Wir erinnern uns nun an die Mengen D_p: die Menge aller *positiven* reellen Zahlen, die eine Anfangsziffer $\leq p$ haben:

$$D_p = \bigcup_{n=-\infty}^{\infty} [10^n\,,\ (p+1)\,10^n) \quad p = 1, \ldots, 9.$$

[3] „Zufällig" soll hier heißen, daß die Ankunftszeiten der Autos gleichverteilt angenommen werden: kein Zeitpunkt soll bevorzugt werden; wenn dies nicht extra angegeben ist, wird man dies vielleicht auch intuitiv – bewußt oder unbewußt – annehmen, obwohl dies vom Standpunkt eines möglichen Realitätsbezuges, auf den es hier allerdings nicht ankommt, zweifelhaft erschiene.

[4] Für einen theoretischen Beweis dieser Beziehung würde z. B. die oben erwähnte „Translationsinvarianz der Wahrscheinlichkeit" eine wichtige Rolle spielen.

Unser Ziel ist es, plausibel zu machen, daß $P(x \in D_p) = \log(p+1)$ ist. Für $x \in \mathbb{R}^+$ ist $\log x \in \mathbb{R}$ und für $\log D_p$ (d. h. die Menge aller $\log x$ mit $x \in D_p$) erhalten wir eine Menge, deren Elemente sich über ganz $\mathbb{R}$ erstrecken

$$\log D_p = \bigcup_{n=-\infty}^{\infty} [\, n\, ,\, n + \log(p+1)\,)\ .$$

$\log D_p$ ist also nichts anderes als in obiger Schreibweise $K_{\log(p+1)}$, daher ergibt sich mit obigen plausiblen Überlegungen $P(\log x \in \log D_p) = \log(p+1)$. Wegen $\log x \in \log D_p \Leftrightarrow x \in D_p$ ergibt sich dadurch[5] das gewünschte Resultat

$$\boxed{P(x \in D_p) = \log(p+1)\ .}$$

Dieses Resultat haben wir eher auf intuitivem Wege erhalten, was für den Schulunterricht auch genügt. Wenn es etwas genauer („axiomatisch“, maßtheoretisch) sein sollte, so kommen z. B. folgende Überlegungen von RAIMI in Betracht.

4.3 Die Methode von RAIMI

RAIMI (1969) selbst benützte sogenannte BANACH-Maße und skalierte Maße, um die Existenz einer oben angesprochenen Verteilungsfunktion aller möglichen positiven Konstanten und eines daraus resultierenden „Mittelpunktes“ all dieser zu vermeiden. Sein Ziel war dasselbe, zu beweisen:
Wenn die Wahrscheinlichkeit $P(1.\ \text{Ziffer} \leq p)$ **skaleninvariant** sein soll, so kommt dafür nur eine logarithmische Verteilungsfunktion in Frage: $P(1.\ \text{Ziffer} \leq p) = \log(p+1)$.

Maße von gewissen Mengen (Teilmengen von $\mathbb{R}$) sollen, worauf auch der Name schon hindeutet, mittels nichtnegativer Zahlenwerte die Größe der in Rede stehenden Teilmengen beschreiben. Diese „Leistung“ kennen wir ja auch durch den ganz naiven Begriff und Gebrauch von Maßen: Gewisse Größen (Längen, Flächeninhalte, Gewichte, etc.) werden mittels nichtnegativer reeller (rationaler) Zahlen und dazugehöriger Einheiten „gemessen“. Die „Maßtheorie“ ist ein eigenes Teilgebiet der Mathematik, in dem Fragen behandelt werden, die Schülern (mit einem klarerweise naiven „Maßbegriff“) i. a. kaum zugänglich sind.

Insbesondere für Lehreraus- und -fortbildungen, Facharbeiten, Seminare etc. wollen wir versuchen, mit so elementar wie möglich gehaltenen maßtheoretischen Betrachtungen die Überlegungen von RAIMI nachzuvollziehen.

Da ein Maß die „Größe“ (magnitude) gewisser Teilmengen von $\mathbb{R}$ (bzw. später $\mathbb{R}^+$) durch nichtnegative Zahlen angeben soll, kommt es zu Definitionen folgender Art. Die Forderungen (Bedingungen) sind ja elementar einsichtig, der neue Name für so eine Funktion („Maß“) soll nicht allzusehr abschrecken, dadurch wird nur die Bedeutung solcher Begriffe unterstrichen (die für einen Schüler an dieser Stelle natürlich nicht zu erkennen ist).

Definition: Eine auf den Teilmengen von $\mathbb{R}$ definierte reellwertige Funktion Θ heißt BANACH-Maß, wenn sie vier Bedingungen erfüllt:

(a) Normierung: $\Theta(\mathbb{R}) = 1$

(b) Nichtnegativität: $\Theta(A) \geq 0 \qquad A \subset \mathbb{R}$

(c) Endliche Additivität: $A \cap B = \emptyset \implies \Theta(A \cup B) = \Theta(A) + \Theta(B)$

(d) Translationsinvarianz: $\Theta(A + s) = \Theta(A) \qquad s \in \mathbb{R},\ A \subset \mathbb{R}.$

Bemerkungen: • Die Bedingungen (b) und (c) sind wohl an jedes „Maß“ zu stellen!

• Es gibt viele solcher BANACH-Maße, wobei man eine Teilmenge von $\mathbb{R}$ dann „BANACH-meßbar“ nennt, wenn alle BANACH-Maße dieser Menge denselben Wert haben. Es gibt nämlich durchaus Mengen, bei denen verschiedene BANACH-Maße verschiedene Werte liefern, z. B. alle Intervalle $(-\infty, x]$ bzw. $[x, \infty)$ – diese sind also nicht BANACH-meßbar (ohne Beweis, RAIMI 1969, S.345).

• Jedes BANACH-Maß Θ eines beschränkten Intervalles $(a, b] \subset \mathbb{R}$ bzw. $[a, b) \subset \mathbb{R}$ muß natürlich 0 sein, weil sonst aufgrund der Translationsinvarianz ein Widerspruch zu $\Theta(\mathbb{R}) = 1$ entstünde: wenn einem endlichen Intervall ein Maß $\varepsilon > 0$ zukommen sollte, so müßte auch allen anderen endlichen Intervallen selber Länge dasselbe Maß ε zukommen (Translationsinvarianz!). Zur Ausschöpfung von ganz $\mathbb{R}$ braucht man klarerweise *unendlich* viele solche disjunkte Intervalle dieser endlichen (!) Länge; das Maß 1 erreicht man aber schon mit der Verinigung bzw. Summe *endlich* (nämlich $\frac{1}{\varepsilon}$) vieler

[5] Darin liegt aber ein sehr wesentlicher Schritt, der hier vielleicht gar nicht so drastisch erscheint: genau genommen soll ja nicht zufällig ein Wert $\log x \in \mathbb{R}$, sondern ein Wert $x \in \mathbb{R}^+$ gewählt werden, wobei man sich für $P(x \in D_p)$ interessiert! Es wird sich herausstellen (siehe nächster Abschnitt), daß dieser Übergang $[\ P(\log x \in \log D_p)$ statt $P(x \in D_p)\ ]$ gerade durch die geforderte *Skaleninvarianz* gerechtfertigt ist: es soll ja $P(x \in D_p) = P(x \in k \cdot D_p)$ sein. Die Skaleninvarianz haben wir bei dieser Plausibilitätsbetrachtung für den Unterricht gar nicht berücksichtigt.

solcher Intervalle (siehe Bedingung (c) – endliche Additivität) $\rightarrow$ Widerspruch!

• Wir haben bis jetzt kein konkretes BANACH-Maß angegeben. Dies deswegen, weil wir uns überhaupt keine Gedanken zu machen brauchen, wie ein konkretes BANACH-Maß „aussieht“; es wird auf die konkrete „Gestalt“ des BANACH-Maßes überhaupt nicht ankommen! Wichtig ist nur, daß solche Funktionen (d. h. BANACH-Maße) überhaupt existieren, und das tun sie (den zugehörigen Beweis deuten wir gar nicht an).

Folgendes Beispiel einer BANACH-meßbaren Menge mit einem Maß $\neq 0$ wird noch sehr wichtig werden; die unbeschränkte Menge K_a sei gegeben durch:

$$K_a \stackrel{\text{def}}{=} \bigcup_{n=-\infty}^{\infty} [n\,,\, n{+}a) \qquad \text{mit} \qquad 0 < a \leq 1\,.$$

Diese Vereinigung halboffener Intervalle (rechtsoffen) ist in Fig. 4 dargestellt und entspricht der Menge $K_{\frac{25}{60}}$ im letzten Abschnitt.

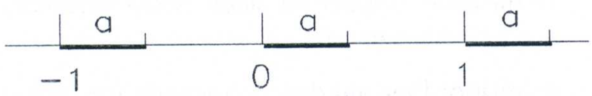

Fig. 4: Die Menge K_a mit Banach-Maß a

Hier ist wohl auch intuitiv zu vermuten: wenn dieser Menge ein Maß zukommen soll, dann kann dies nur a sein, denn in jedem Intervall $[n, n+1)$ spiegelt sich das Verhältnis von K_a zu $\mathbb{R}$ wider und dort ist es mit $a : 1$ abzulesen. Anders formuliert: wegen der augenscheinlichen Periodizität kann der große Bereich $\mathbb{R}$ auch *intuitiv* auf das Intervall $[0, 1)$ eingeschränkt werden. Die Wahrscheinlichkeit, daß eine „zufällig“ gewählte reelle Zahl in K_a liegt, ist daher auch ohne wirkliche Exaktifizierung (z. B. durch maßtheoretische Überlegungen) mit a zu quantifizieren (siehe oben – Gleichverteilung vorausgesetzt oder intuitiv angenommen).

Es ist eine gute Übungsaufgabe für Studenten, $\Theta(K_a) = a$ ausschließlich aus den obigen vier Eigenschaften abzuleiten. Wenn wir allein aus den obigen vier Forderungen an ein BANACH-Maß zeigen können, daß das Maß von K_a den Wert a haben muß, so haben wir auch gezeigt, daß K_a überhaupt BANACH-meßbar ist. RAIMI schreibt zum Thema $\Theta(K_a) = a$ nur lapidar: „The BANACH-measure of K_a is a, as may easily be deduced from properties (a)-(d).“ (1969, S.345). Wir wollen im folgenden eine Beweisskizze geben:

Aus der endlichen Additivität (c) und aus der Translationsinvarianz (d) folgt zunächst für $a + b \leq 1$ $(a, b > 0)$ die Beziehung $\Theta\left(K_{a+b}\right) = \Theta\left(K_a + (K_b + a)\right) = \Theta(K_a) + \Theta(K_b + a) = \Theta(K_a) + \Theta(K_b)$, also

$$\Theta\left(K_{a+b}\right) = \Theta(K_a) + \Theta(K_b)\,, \tag{1}$$

was natürlich sofort auf endlich viele Summanden verallgemeinerbar ist. Mit $K_1 = \mathbb{R}$ und (a) ergibt sich daraus unmittelbar für $k \in \mathbb{N}$

$$1 = \Theta(K_1) = \Theta\left(K_{\frac{1}{k}+\frac{1}{k}+\ldots+\frac{1}{k}}\right) = k \cdot \Theta\left(K_{\frac{1}{k}}\right)\,,$$

woraus wir

$$\Theta\left(K_{\frac{1}{k}}\right) = \frac{1}{k} \qquad \forall k \in \mathbb{N} \tag{2}$$

erhalten. Mit Hilfe von (1) und (2) können wir das BANACH-Maß von K_a für alle rationalen Werte von a bestimmen. Mit $a = \frac{b}{c}$ ergibt sich

$$\Theta\left(K_{\frac{b}{c}}\right) = b \cdot \Theta\left(K_{\frac{1}{c}}\right) = \frac{b}{c}\,. \tag{3}$$

Wir haben bis jetzt also $\Theta(K_a) = a$ für rationale a gezeigt und wollen dies nun auch für irrationale Werte von a beweisen. Sei also $0 < a < 1$ eine irrationale Zahl; dann gibt es bekanntlich eine monoton nicht fallende Folge rationaler Zahlen $(a_n)_{n=1}^{\infty}$ mit $a_n \in \mathbb{Q}$ und $a_n \nearrow a$ (z. B. nach dem „Intervallschachtelungsprinzip“: Intervallfolge mit rationalen Endpunkten $([a_n\,,\, b_n])_{n=1}^{\infty}$ mit innerstem Punkt a).

Mit $a_n \nearrow a$ und $b_n \searrow a$ $(a_n, b_n \in \mathbb{Q})$ gilt zunächst $a_n \leq a \leq b_n$ und somit $K_{a_n} \subset K_a \subset K_{b_n}$ bzw.

$$\Theta\left(K_{a_n}\right) \leq \Theta(K_a) \leq \Theta(K_{b_n})\,,$$

denn aus der Nichtnegativität (b) und aus der Additivität (c) ergibt sich allgemein:

$$A \subset B \quad \Rightarrow \quad \Theta(A) \leq \Theta(A) + \Theta(B \setminus A) = \Theta(B)\,,$$

weil $\Theta(B \setminus A) \geq 0$ ist. Mit $\Theta(K_{a_n}) = a_n$ und $\Theta(K_{b_n}) = b_n$ ergibt sich daraus unmittelbar

$$a_n \leq \Theta(K_a) \leq b_n \qquad \forall n \in \mathbb{N}\,.$$

Da a der einzige (eindeutige) innere Punkt der Intervallschachtelung $([a_n\,,\, b_n])_{n=1}^{\infty}$ ist, muß auch für irrationale a gelten: $\Theta(K_a) = a$.

Damit haben wir $\Theta(K_a) = a$ für alle reellen a mit $0 < a \leq 1$ bewiesen, ein Ergebnis, das später von Bedeutung werden wird.

Da das Universum aller möglichen physikalischen Konstanten als $\mathbb{R}^+$ (und nicht als ganz $\mathbb{R}$) angesehen wird, und *Skaleninvarianz* vorausgesetzt werden soll, muß man sich was einfallen lassen, diese beiden Modellbedingungen wirklich ins Spiel zu bringen; man wird daher von unten folgender Definition (leider noch ein neuer Name!) nicht völlig überrascht sein.

Definition: Eine auf den Teilmengen von $\mathbb{R}^+$ definierte reellwertige Funktion Ψ heißt **skaliertes Maß**, wenn sie vier Bedingungen erfüllt :

(a') Normierung: $\Psi(\mathbb{R}^+) = 1$

(b') Nichtnegativität: $\Psi(A) \geq 0 \quad A \subset \mathbb{R}^+$

(c') Endliche Additivität: $A \cap B = \emptyset \implies \Psi(A \cup B) = \Psi(A) + \Psi(B)$

(d') Skaleninvarianz:
$\Psi(A\,s) = \Psi(A) \qquad s \in \mathbb{R}^+,\ A \subset \mathbb{R}^+.$

Ein **skaliertes Maß** auf $\mathbb{R}^+$ ist also das „Pendant" zum BANACH-Maß auf $\mathbb{R}$ in folgendem Sinn: Zu jedem BANACH-Maß Θ auf $\mathbb{R}$ kann ganz einfach das *zugehörige* skalierte Maß Ψ auf $\mathbb{R}^+$ konstruiert werden: $\Psi(A) \stackrel{\text{def}}{=} \Theta(\log(A))$ – die Existenz von skalierten Maßen ist an die Existenz von BANACH-Maßen gebunden und umgekehrt. (Der Existenznachweis von BANACH- und skalierten Maßen sei hier absichtlich ausgeblendet.) Man überlege, daß für Ψ mit $\Psi(A) \stackrel{\text{def}}{=} \Theta(\log(A))$ wirklich (a')–(d') erfüllt sind, wenn Θ ein beliebiges BANACH-Maß ist, also (a)–(d) erfüllt; die Skaleninvarianz von Ψ ist dank der Logarithmusfunktion äquivalent zur Translationsinvarianz von Θ.

Man kann sogar relativ leicht zeigen, daß *jedes* skalierte Maß vermöge der Logarithmusfunktion „von einem BANACH-Maß herkommt" [dazu sei nur bemerkt: die Abbildung $\log_{10} : \mathbb{R}^+ \to \mathbb{R}$ bildet die multiplikative Gruppe $(\mathbb{R}^+, \cdot)$ *isomorph* auf die additive Gruppe $(\mathbb{R}, +)$ ab, d. h. die Abbildung $\log$ ist bijektiv und „strukturerhaltend": $\log(a \cdot b) = \log a + \log b$].

Solche skalierten Maße haben für unser Problem drei wichtige Eigenschaften:

1) Sie wirken auf $\mathbb{R}^+$, dem potentiellen Universum der physikalischen Konstanten.

2) Sie können als **Wahrscheinlichkeitsmaße** interpretiert werden, denn durch die ersten drei Forderungen an ein BANACH- bzw. skaliertes Maß – (a), (b), (c) bzw. (a'), (b'), (c') – sind ja die KOLMOGOROFF-Axiome für eine Wahrscheinlichkeitsfunktion bzw. für ein Wahrscheinlichkeitsmaß erfüllt. D. h. wenn eine Teilmenge $A \subset \mathbb{R}^+$ ein skaliertes Maß $\Psi(A)$ hat, so kann dieser Wert $\Psi(A)$ als **Wahrscheinlichkeit** interpretiert werden, daß eine „zufällig" gewählte positive reelle Zahl x in A liegt („günstig durch möglich"):

$$P(x \in A) = \frac{\Psi(A)}{\Psi(\mathbb{R}^+)} = \Psi(A)\ .$$

3) Skalierte Maße sind per definitionem *skaleninvariant*, eine Eigenschaft die wir von der Wahrscheinlichkeitsfunktion $P(x \in A)$ a priori verlangt haben. Skalierte Maße sind also nichts anderes als **skaleninvariante Wahrscheinlichkeitsmaße auf** $\mathbb{R}^+$ – ideal für unsere Verhältnisse! Das beantwortet schon teilweise die folgende Frage, die man stellen könnte:

Was hat das mit dem Problem der ersten Ziffer bzw. mit dem Gesetz von BENFORD zu tun?

Diese Frage ist nun sehr leicht beantwortet! Sei nämlich analog zu oben $D_p \subset \mathbb{R}^+$ die Menge aller positiven reellen Zahlen, deren erste Ziffer in Dezimalschreibweise $\leq p$ ist. Dann läßt sich D_p [siehe oben] als folgende abzählbare Vereinigung paarweise disjunkter Mengen schreiben:

$$D_p = \bigcup_{n=-\infty}^{\infty} [10^n, (p+1)\,10^n) \qquad p = 1, \ldots, 9.$$

Das skalierte Maß Ψ dieser Menge D_p gibt die Wahrscheinlichkeit an, daß eine zufällig gewählte positive reelle Zahl in D_p liegt, also eine erste Ziffer $\leq p$ hat (mit der im Maß geforderten Skaleninvarianz, analog zu PINKHAM). Dieser Wert wird sich für *jedes* skalierte Maß als $\log(p+1)$ herausstellen, *unabhängig* davon, *welches* BANACH-Maß zugrunde gelegt wird! Betrachten wir das schon erwähnte skalierte Maß, das über ein BANACH-Maß definiert ist: $\Psi(D_p) = \Theta(\log D_p)$. Nun gilt aber

$$\log D_p = \bigcup_{n=-\infty}^{\infty} [n, n + \log(p+1))\ ,$$

d. h. die Menge $\log D_p$ ist nichts anderes als in vorheriger Diktion $K_{\log(p+1)}$. Bestimmen wir $\Psi(D_p) = \Theta(\log D_p)$, so erhalten wir – tatsächlich unabhängig von der speziellen Wahl des BANACH-Maßes Θ –

$$\begin{aligned} P(x \in D_p) &= \Psi(D_p) = \Theta(\log D_p) \\ &= \Theta\left(K_{\log(p+1)}\right) = \log(p+1)\ , \end{aligned}$$

wie wir oben für jedes beliebige BANACH-Maß ausführlich dargestellt haben! D. h. wir haben uns zu Recht nicht den Kopf über ein konkretes BANACH-Maß zerbrochen! Mit $\Psi(D_p) = P(x \in D_p) = P(1.\ \text{Ziffer} \leq p) = \log(p+1)$ ergibt sich natürlich

$$\boxed{P(1.\ \text{Ziffer} = p) = \log(p+1) - \log p\ .}$$

Zusammenfassende Bemerkungen:

- Das „dubiose" Intervall $(0,T]$, dem im Modell von PINKHAM Maß (Wahrscheinlichkeit) $\frac{1}{2}$ zukommt, ist in diesem Modell einfach *nicht meßbar*, denn $\Psi((0,T]) = \Theta((-\infty, \log T])$ und solche Mengen sind laut oben *nicht* BANACH-meßbar!

- Jedes beschränkte Intervall $[x,y) \subset \mathbb{R}^+$ hat 0 als skaliertes Maß, denn $\Psi([x,y)) = \Theta([\log x, \log y)) = 0$. (Jedes BANACH-Maß Θ eines beschränkten Intervalles $[a,b) \subset \mathbb{R}$ muß natürlich 0 sein, weil sonst aufgrund der Translationsinvarianz ein Widerspruch zu $\Theta(\mathbb{R}) = 1$ entstünde!) Dies scheint auch philosophisch haltbar zu sein, denn wenn *alle* denkbaren physikalischen Einheiten „gleichberechtigt" sind, dann werden „fast alle" Eintragungen in den resultierenden Tabellen nahe 0 oder ∞ liegen. Salopp formuliert: Es gibt viel mehr Einheiten (und daher Zahlenwerte in Listen) außerhalb jedes konkret vorgegebenen beschränkten Intervalls als innerhalb desselben!

- Am Ergebnis und an der prinzipiellen Vorgangsweise ändert sich nichts, wenn man realistischererweise nicht $\mathbb{R}^+$, sondern $\mathbb{Q}^+$ (oder gar nur die *endlichen* Dezimalzahlen) als das mögliche Universum aller physikalischen Konstanten ansieht (bei allen in Dezimalschreibweise angegebenen Werten aller möglichen Tabellen können ja nur endlich viele Stellen berücksichtigt werden) – ohne Beweis.

- Das BENFORD-Gesetz hat mit der Darstellung im *Dezimalsystem* **nichts** zu tun. Auch bei Darstellungen mit jeder anderen natürlichen Zahl $a > 2$ als Basis ergäbe sich ein analoges Gesetz für die Wahrscheinlichkeit, daß eine Zahl mit einer Ziffer $\leq p$ beginnt: $P(x \in D_p) = \log_a(p+1)$ für $p = 1, 2, 3, \ldots, a-1$.

- *Wenn* es überhaupt eine Wahrscheinlichkeit für das Auftreten von Ziffern als erste Ziffer gibt (wenn also D_p ein Wahrscheinlichkeitsmaß $\Phi(D_p)$ besitzt) – und die empirischen Beobachtungen unterstützen diese These – , und wenn man vernünftigerweise zusätzlich die Unabhängigkeit von zugrunde gelegten Skalen fordert, dann muß $\Phi(D_p) = \log(p+1)$ sein! Dies erklärt also, warum die Daten von Energieverbrauchszahlen, die Oberflächen von Seen in Quadratmeilen oder in Quadratmillimeter und alle anderen physikalischen Größen, bei denen es Einheiten gibt, das BENFORD-Gesetz „befolgen". Da aber der Begriff der Skaleninvarianz bei *Hausnummern* sicher nichts zu bedeuten hat, so bleibt die dadurch unbeantwortete Frage, warum sich auch Hausnummern in den Privatadressen z. B. aller deutschsprachigen Universitätsangestellten an dieses Gesetz halten sollten?

Literatur:

BENFORD, F. (1938): The law of anomalous numbers. In: Proceedings of the American Philosophical Society **78**, 551–572.

DWORSCHAK, M. (1998): Weiter Weg zur Zwei – ein kurioses Gesetz der Wahrscheinlichkeitstheorie kann Finanzbeamten helfen, Steuersünder aufzuspüren. In: Der Spiegel 47/1998, 228–229.

HILL, T.P. (1995): The significant digit phenomenon. In: American Mathematical Monthly **102**, 322–327.

HILL, T.P. (1995b): Base-invariance implies Benford's Law. In: Proceedings of the American Mathematical Society **123**, 3, 887–895.

HUMENBERGER, H. (1996): Das BENFORD-Gesetz über die Verteilung der ersten Ziffer von Zahlen. In: Stochastik in der Schule **16**, 3, 2–17. Kurzfassung: Beiträge zum Mathematikunterricht 1997, 251–254.

HUMENBERGER, H. (1997): Eine Ergänzung zum BENFORD-Gesetz — weitere mögliche schulrelevante Aspekte. In: Stochastik in der Schule **17**, 3, 42–48.

KRÄMER, W. (1990): Das Gesetz der abnormalen Zahl. In: Stochastik in der Schule **10**, 3, 48–52.

MATTHEWS R. (1999): The power of one. In: New Scientist No2194 (10 July 1999), 27–30.

PINKHAM, R.S. (1961): On the distribution of first significant digits. In: Annals of Mathematical Statistics **32**, 1223–1230.

RAIMI, R.A. (1969): On the distribution of first significant figures. In: American Mathematical Monthly **76**, 342–348.

RAIMI, R.A. (1969b): The peculiar distribution of first digits. In: Scientific American **221** (Dezember 1969), 109–120.

RAIMI, R.A. (1976): The first digit problem. In: American Mathematical Monthly **83**, 521–538.

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.

E-mail: hans@edv1.boku.ac.at