

## Eine Ergänzung zum BENFORD-Gesetz — weitere mögliche schulrelevante Aspekte

von Hans Humenberger, Wien

**Zusammenfassung:** In Humenberger, 1996 wurde das Problem näher beleuchtet, warum die Auftrittswahrscheinlichkeit  $P(k)$  einer Ziffer  $k$ , als erste Ziffer einer Zufallszahl zu stehen, *nicht* für alle 9 möglichen Ziffern jeweils  $\frac{1}{9}$  beträgt, sondern von 1 bis 9 abnimmt nach einem logarithmischen Gesetz, das nach BENFORD benannt ist :

$$P(k) = \log_{10}(k+1) - \log_{10} k .$$

(Im Gegensatz zum erwähnten Aufsatz werden hier die Ziffern mit  $k$  statt mit  $p$  bezeichnet.) Dieses Gesetz gilt für beliebige „physikalische Konstanten“ als mögliche Zufallszahlen (positive reelle Zahlen), wie sich aus der Forderung nach *Skaleninvarianz* (denn die Wahrscheinlichkeiten der einzelnen Ziffern sollen nicht von den doch höchst subjektiven zugrundegelegten Maßeinheiten abhängen) ergibt. Statt reellen Zahlen kann man zunächst natürliche Zahlen als potentielle Zufallswerte nehmen und fragen, wie dann die Auftrittswahrscheinlichkeiten sich abschätzen lassen. Genau dies ist der Inhalt dieses Ergänzungsbeitrages, der ausschließlich elementarste Schulmathematik enthält und uns dadurch für selbständige Schüleraktivitäten besonders geeignet erscheint.

### Die Wahrscheinlichkeiten $P(k)$ 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 ( $k \neq 0$ ) genau einmal! Wir erkennen auch sofort, daß bei natürlichen Zahlen die Wahrscheinlichkeit  $\frac{1}{9}$  nur dann bei jeder Ziffer  $k = 1, 2, \dots, 9$  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  $(P_n(1))_{n \in \mathbb{N}}$ :

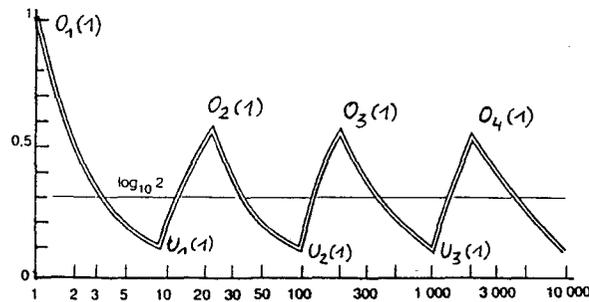
Bei  $n = 1$  ist  $P(1) = 1$ , bei  $n = 2$  ist  $P(1) = \frac{1}{2}$  usw., diese Wahrscheinlichkeit sinkt dann bis  $P(1) = \frac{1}{9}$  bei  $n = 9$ . Dann steigt die Wahrscheinlichkeit wieder bis  $n = 19$  (auf  $P(1) = \frac{11}{19}$ ), um dann wieder bis  $n = 99$  abzufallen (wieder  $P(1)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

$n$	1	9	19	99	199	999	1999	9999	19999
$P_n(1)$	$\frac{1}{1}$	$\frac{1}{9}$	$\frac{11}{19}$	$\frac{11}{99} = \frac{1}{9}$	$\frac{111}{199}$	$\frac{111}{999} = \frac{1}{9}$	$\frac{1111}{1999}$	$\frac{1111}{9999} = \frac{1}{9}$	$\frac{11111}{19999}$

Tabelle 1: Wahrscheinlichkeiten  $P_n(1)$ ; Obergrenzen  $O_m(1) = \frac{11\dots1}{19\dots9}$ , Untergrenze  $\frac{1}{9}$

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 Anstieges bzw. Abfalles klarerweise länger, die Folge  $(P_n(1))_{n \in \mathbb{N}}$  ist sicher divergent. Einfach durch Grenzwertbildung ( $\lim_{n \rightarrow \infty}$ ) 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 Obergrenze  $O_m(1) = \frac{11}{19}, \frac{111}{199}, \frac{1111}{1999}, \dots$ ; 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  $n = 10^\ell - 1$  ist  $P_n(1) = \frac{1}{9}$ ). Salopp formuliert: nach jeder neu dazukommenden Stelle wird die Ziffer 1 lange Zeit „bevorzugt“, bevor die anderen Ziffern der Reihe nach diesen Rückstand wieder aufholen. In Abb. 1 ist der Verlauf von  $P_n(1)$  als kontinuierliche Funktion in Abhängigkeit von  $n$  graphisch dargestellt, wobei die  $n$ -Achse logarithmisch skaliert wurde, damit die Abstände zwischen zwei aufeinanderfolgenden Zehnerpotenzen nicht immer größer werden, sondern gleich bleiben.

Die Obergrenzen  $O_m(1)$  könnten auch im Schulunterricht näher studiert werden. Was passiert mit diesen Obergrenzen? Existiert ein Grenzwert? Eine Frage, die durchaus von Schülern selbständig bearbeitet werden kann (Vernetzung: Grenzwert von Folgen), indem sie die Zahlen  $O_m(1) = \frac{11\dots1}{19\dots9}$  näher betrachten. Sie sollten z.B. in der Lage sein zu beweisen, daß die Folge dieser Zahlen  $O_m(1)$  (Obergrenzen,  $m$  ist die Anzahl der Stellen in Zähler und Nenner) monoton fallend und nach unten beschränkt ist (z.B. durch 0 oder sogar durch  $\frac{1}{2}$ ), wodurch die Konvergenz gesichert wäre. Sie

Abbildung 1: Graphische Darstellung der Werte  $P_n(1)$  als Funktion von  $n$ 

werden auch schnell (z.B. durch Probieren, i.e. Berechnen einiger Werte mit dem Taschenrechner oder Computer) auf die Vermutung kommen, daß es der Wert  $0.5 = \frac{5}{9}$  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 Differenz  $|\frac{11\dots 1}{19\dots 9} - \frac{5}{9}| < \varepsilon$  ist, was für  $19\dots 9 > \frac{4}{9\varepsilon}$  erfüllt ist. So könnte der Grenzwert *ohne* explizite Termdarstellung der Zahlen  $O_m(1) \stackrel{\text{def}}{=} \frac{11\dots 1}{19\dots 9}$  bestimmt werden.

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  $11\dots 1$  ( $m$  Stellen) können z.B. als  $\frac{10^m - 1}{9}$  und die Nenner  $19\dots 9$  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.

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 (vgl. Abb. 1). 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 reelle Zahl in Dezimalschreibweise bei Weglassen des Kommas einer natürlichen Zahl entsprechen. So argumentierte Flehinger, 1966 beim nach BENFORD benannten Gesetz des Logarithmus für alle denkbaren positiven reellen Zahlen (Konstanten).

## 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  $n = 1, \dots, 8$  ist  $P(9) = 0$ , bei  $n = 9$  ist  $P(9) = \frac{1}{9}$ , diese Wahrscheinlichkeit sinkt dann bis  $P(9) = \frac{1}{89}$  bei  $n = 89$ . Dann steigt die Wahrscheinlichkeit wieder bis  $n = 99$  auf den Wert  $P(9) = \frac{11}{99} = \frac{1}{9}$ , um dann wieder bis  $n = 899$  abzufallen auf  $P(9) = \frac{111}{899}$ , dann kommt wieder eine hundert Zahlen lange Aufholphase bis  $n = 999$  auf  $P(9) = \frac{1111}{999} = \frac{1}{9}$ , der wieder ein Abfall bis  $n = 8999$  folgt usw. (siehe Tab. 2).  $P_n(9)$  schwankt immer zwischen  $\frac{1}{9}$  (konstante Obergrenze!) und der Untergrenze  $U_m(9) = 0, \frac{1}{89}, \frac{11}{899}, \frac{111}{8999}, \frac{1111}{89999}, \dots$ . Es ist auch hier plausibel, daß  $P(9) < \frac{1}{9}$  ist, da für fast alle  $n \in \mathbb{N}$  der entsprechende Wert  $P_n(9) < \frac{1}{9}$  ist (nur bei  $n = 10^l - 1$  ist  $P_n(9) = \frac{1}{9}$ ). Salopp formuliert: nach jeder neu dazukommenden Stelle hat die Ziffer 9 lange Zeit „Rückstand“, bis sie unmittelbar vor der nächsten Stelle den Rückstand wieder aufholt. Die Ziffer 9 hat also nie einen „Vorteil“ gegenüber einer anderen, sie kann ihren Rückstand nur manchmal wettmachen!

$n$	8	9	89	99	899	999	8999	9999	89999
$P_n(9)$	0	$\frac{1}{9}$	$\frac{1}{89}$	$\frac{11}{99} = \frac{1}{9}$	$\frac{111}{899}$	$\frac{1111}{999} = \frac{1}{9}$	$\frac{11111}{8999}$	$\frac{111111}{9999} = \frac{1}{9}$	$\frac{1111111}{89999}$

Tabelle 2: Wahrscheinlichkeiten  $P_n(9)$ ; Obergrenze  $\frac{1}{9}$  und Untergrenzen  $U_m(9) = \frac{1\dots 1}{89\dots 9}$ 

Die Folge der Untergrenzen  $U_m(9) = \frac{1\dots 1}{89\dots 9}$  ( $m$  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 (als explizite Darstellung der Zahlen  $U_m(9)$  ergibt sich  $U_m(9) = \frac{1}{81} \cdot \frac{10^m - 1}{10^m - 179}$ ).

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  $k = 2, \dots, 8$  die Werte der jeweiligen Ober- bzw. Untergrenzen  $O_m(k)$  bzw.  $U_m(k)$  nicht mehr konstant. Tab. 3 gibt einen Überblick über die entsprechenden Werte bei der Ziffer 3.

$n$	2	3	29	39	299	399	2999	3999	29999
$P_n(3)$	0	$\frac{1}{3}$	$\frac{1}{29}$	$\frac{11}{39}$	$\frac{11}{299}$	$\frac{111}{399}$	$\frac{111}{2999}$	$\frac{1111}{3999}$	$\frac{1111}{29999}$

Tabelle 3: Obergrenzen  $O_m(3) = \frac{11\dots 1}{39\dots 9}$  und Untergrenzen  $U_m(3) = \frac{1\dots 1}{29\dots 9}$  der Werte  $P_n(3)$ 

Die Schüler können ganz selbständig die Form der Unter- bzw. Obergrenzen für jede beliebige Ziffer  $k$  erarbeiten: ganz allgemein haben diese für die Ziffer  $k$  die Gestalt:

$$U_m(k) = \frac{1\dots 1}{(k-1)9\dots 9} = \frac{1}{9k} \cdot \frac{10^m - 1}{10^m - \frac{1}{k}}$$

und

$$O_m(k) = \frac{11\dots 1}{k9\dots 9} = \frac{10}{9(k+1)} \cdot \frac{10^m - 1}{10^m - \frac{10}{k+1}}$$

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

$$U(k) \stackrel{\text{def}}{=} \lim_{m \rightarrow \infty} U_m(k) = \lim_{m \rightarrow \infty} \left( \frac{1}{9k} \cdot \frac{10^m - 1}{10^m - \frac{1}{k}} \right) = \frac{1}{9k},$$

und

$$O(k) \stackrel{\text{def}}{=} \lim_{m \rightarrow \infty} O_m(k) = \lim_{m \rightarrow \infty} \left( \frac{10}{9(k+1)} \cdot \frac{10^m - 1}{10^m - \frac{10}{k+1}} \right) = \frac{10}{9(k+1)}.$$

Die so definierten (Grenz-)Werte  $U(k)$  bzw.  $O(k)$  sind natürlich Schranken für die Wahrscheinlichkeiten  $P(k)$ , falls diese überhaupt existieren. Man sieht (vgl. auch Tab. 4), daß die Intervalle  $[U(k), O(k)]$  für alle  $k = 1, \dots, 9$  den Wert  $\frac{1}{9}$  enthalten, m.a.W. daß die Wahrscheinlichkeiten  $P_n(k)$  jeweils um diesen Wert „schwanken“ (mit den beiden Extremfällen: bei  $k = 1$  als konstante Untergrenze, bei  $k = 9$  als konstante Obergrenze) — bei  $n = 9, 99, 999 \dots$  ist ja  $P_n(k) = \frac{1}{9}$  für alle  $k = 1, \dots, 9$  (die Wahrscheinlichkeiten  $P_n(k)$  „kehren mit wachsendem  $n$  immer wieder zum Wert  $\frac{1}{9}$  zurück“).

Als Schätzwert für  $P(k)$  böte sich in erster Näherung z.B. der jeweilige Mittelwert  $P(k) \approx \frac{U(k)+O(k)}{2}$  an, eine Schätzung, die — wie wir sehen werden

— gar nicht so schlecht ist. Tab. 4 gibt einen Überblick über die jeweiligen Werte von  $U(k)$  und  $O(k)$  (als Intervallgrenzen) in exakter Form und in Dezimaldarstellung (auf drei Dezimalen gerundet); weiters ist der jeweilige Intervallmittelpunkt  $\frac{U(k)+O(k)}{2}$  angegeben und der wirkliche Wert  $P(k)$  (auf drei Dezimalen gerundet). Ein Vergleich dieser beiden Werte zeigt, daß der jeweilige Schätzwert durch den Intervallmittelpunkt eine Differenz zum jeweils exakten Wert von nur 0.016 (bei  $k = 9$ ) bis 0.037 (bei  $k = 2$ ) aufweist.

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

Tabelle 4: Überblick über die Werte von  $U(k)$  bzw.  $O(k)$ 

Eine elementare Möglichkeit zur Verbesserung der Genauigkeit der durch die Intervallmittelpunkte gegebenen Schätzwerte könnte folgende Idee sein: Da die Summe dieser Schätzwerte (vierte Spalte in Tab. 4) nicht 1, sondern 1.229 ergibt, so ist es besonders naheliegend, alle diese Werte durch 1.229 zu dividieren, was bei  $k = 2, \dots, 9$  die Genauigkeit beträchtlich und bei  $k = 1$  auch ein wenig erhöhte (diese Werte sind die Werte in Klammer in der vierten Spalte von Tab. 4). Die Differenz zum exakten Wert beträgt dann nur 0.003 bis 0.006 bei  $k = 2, \dots, 9$  und 0.030 bei  $k = 1$ .

So könnten plausible und auch passable Schätzwerte gewonnen werden für die Wahrscheinlichkeiten  $P(k)$  — und zwar *a priori*, d.h. ohne die Überlegungen von z.B. Humenberger, 1996 (oder anderen), die den mathematischen Hintergrund erst etwas erleuchten.

Uns scheint es sinnvoll zu sein, die hier aufgezeigten Ideen *vor* einer allfälligen Bearbeitung im Sinne von z.B. Humenberger, 1996 den Studenten bzw. Schülern nahezubringen.

### Literatur

Flehinger, B.J. (1966): On the probability that a random integer has initial digit  $a$ . In: American Mathematical Monthly **73**, 1056–1061.

Humenberger, H. (1996): Das BENFORD-Gesetz über die Verteilung der ersten Ziffer von Zahlen. In: Stochastik in der Schule **16**, 3, 2–17. Vgl. auch die dort angegebene Literaturliste.

### Adresse des Autors:

Hans Humenberger  
Institut für Mathematik und Angewandte Statistik  
Fachbereich Mathematik  
Universität für Bodenkultur  
Gregor Mendel-Straße 33  
A – 1180 Wien

E-mail: [hans@edv1.boku.ac.at](mailto:hans@edv1.boku.ac.at)