Warum 1 so oft vorne steht –

– das eigenartige Gesetz von Newcomb und Benford

Hans Humenberger

universität wien

Fakultät für Mathematik

Geschichte

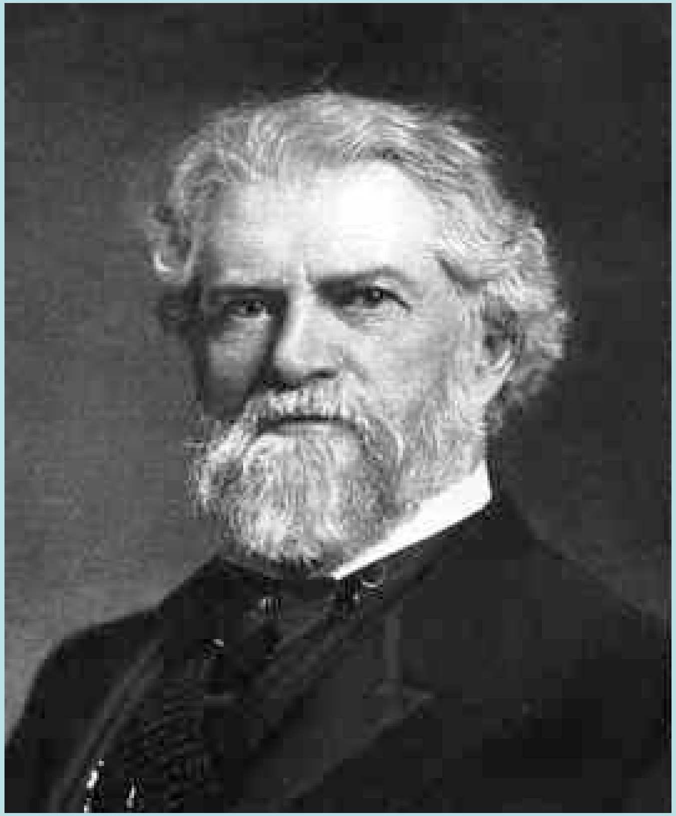
- **Simon Newcomb** (Amerikanischer Astronom und Mathematiker, 1835 – 1909) **1881:** Logarithmentafeln in Bibliotheken auf den ersten Seiten viel abgegriffener und dreckiger als auf den hinteren!
- D. h.: Log von Zahlen mit 1. Ziffer (nicht 0!) 1, 2, . . . häufiger gesucht als 9, 8, . . . !
- Aber warum? Kommen Zahlen mit niedrigen Anfangsziffern **in der Welt** häufiger vor? Bevorzugt die Natur die 1 als Anfangsziffer?

Newcomb gab bereits eine Formel an:

Relative Häufigkeit für $d = 1, \ldots, 9$ als Anfangsziffer:

$$\log_{10}(d+1) - \log_{10} d = \log_{10} \frac{d+1}{d}$$

Keine Erklärung, kein Beweis, nur als interessante Kuriosität, bald wieder vergessen!



Frank Benford (Amerikanischer Elektroingenieur und Physiker, 1883 – 1948)
1938: dieselbe Entdeckung!

Viel mehr fasziniert davon!

Sammelte mit Akribie eine Unmenge von „Daten“!

→ Gesetz primär nach ihm benannt!

Führende Ziffern:

1: ca. 30%, . . ., **9**: ca. 4%

Auch er: kein Beweis!

Benford hat u. a. untersucht:

		Relative Häufigkeit als Anfangsziffer								
Untersuchungs-Objekte	*Anzahl*	1	2	3	4	5	6	7	8	9
Seen (Fläche)	335	31,0	16,4	10,7	11,3	7,2	8,6	5,5	4,2	5,1
Bewohner der US-Counties	3259	33,9	20,4	14,2	8,1	7,2	6,2	4,1	3,7	2,2
Naturkonstanten	104	41,3	14,4	4,8	8,6	10,6	5,8	1,0	2,9	10,6
Zeitungen	100	30,0	18,0	12,0	10,0	8,0	6,0	6,0	5,0	5,0
Spezifische Wärme	1389	24,0	18,4	16,2	14,6	10,6	4,1	3,2	4,8	4,1
Molekulargewicht	1800	26,7	25,2	15,4	10,8	6,7	5,1	4,1	2,8	3,2
Atomgewicht	91	47,2	18,7	5,5	4,4	6,6	4,4	3,3	4,4	5,5
Readers Digest	308	33,4	18,5	12,4	7,5	7,1	6,5	5,5	4,9	4,2
Hausnummern										
American Men of Science	342	28,9	19,2	12,6	8,8	8,5	6,4	5,6	5,0	5,0

Auch: Halbwertszeiten, Energieverbrauchszahlen, Längen von Flüssen, Größe von Dateien auf Festplatte, Einwohnerzahlen von Städten, etc.

Spiel (Computer mit ungefälschten aber unvorhersehbaren Daten): Halbwertszeiten, Energieverbrauchszahlen, Längen von Flüssen, Einwohnerzahlen, . . .

Anfangsziffer 1, 2, 3: Verlust von 10 €

Anfangsziffer 4, 5, 6, 7, 8, 9 Gewinn von 10 €

Naive Rechnung: In 6 von 9 Fällen Gewinn!

→ Dieses Spiel lohnt sich auf lange Sicht (?!):

Durchschnittlicher Gewinn pro Spiel: $(-10)\cdot\frac{1}{3}+(+10)\cdot\frac{2}{3}=3\frac{1}{3}$

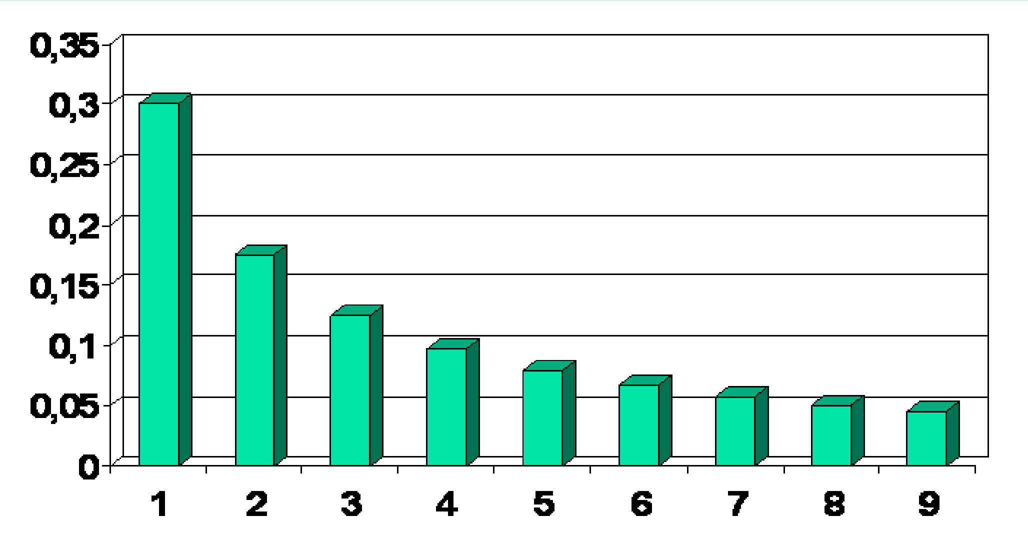
Dabei vorausgesetzt: die Ziffern 1, . . ., 9 sind als Anfangsziffern gleich wahrscheinlich!

Nach Benford/Newcomb sind aber 1, . . ., 9
nicht gleich wahrscheinlich als Anfangsziffer:

Newcomb-Benford-Gesetz: Für $d = 1, 2, \ldots, 9$ gilt

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

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



Wie kann man das verstehen, einsehen, plausibel machen, begründen, . . . ?

Warum ist 1 als Anfangsziffer bevorzugt?

1	2	3	4	5	6	7	8	9
30%	18%	12%	10%	8%	7%	6%	5%	4%

Wenn diese Zahlen stimmen:

60%	40%

Durchschnittlicher Gewinn pro Spiel:

$$(-10)\cdot 0,60+(+10)\cdot 0,40=-2$$

Spiel lohnt sich also doch nicht ?!?!?!

Google-Experiment

EXCEL!

Erklärung bei natürlichen Zahlen: Wahrscheinlichkeiten $P_n(1)$

Zufallszahlen aus der Menge: $\{1, 2, \ldots, n\}$

$P_n(1)$ bei wachsendem n ?

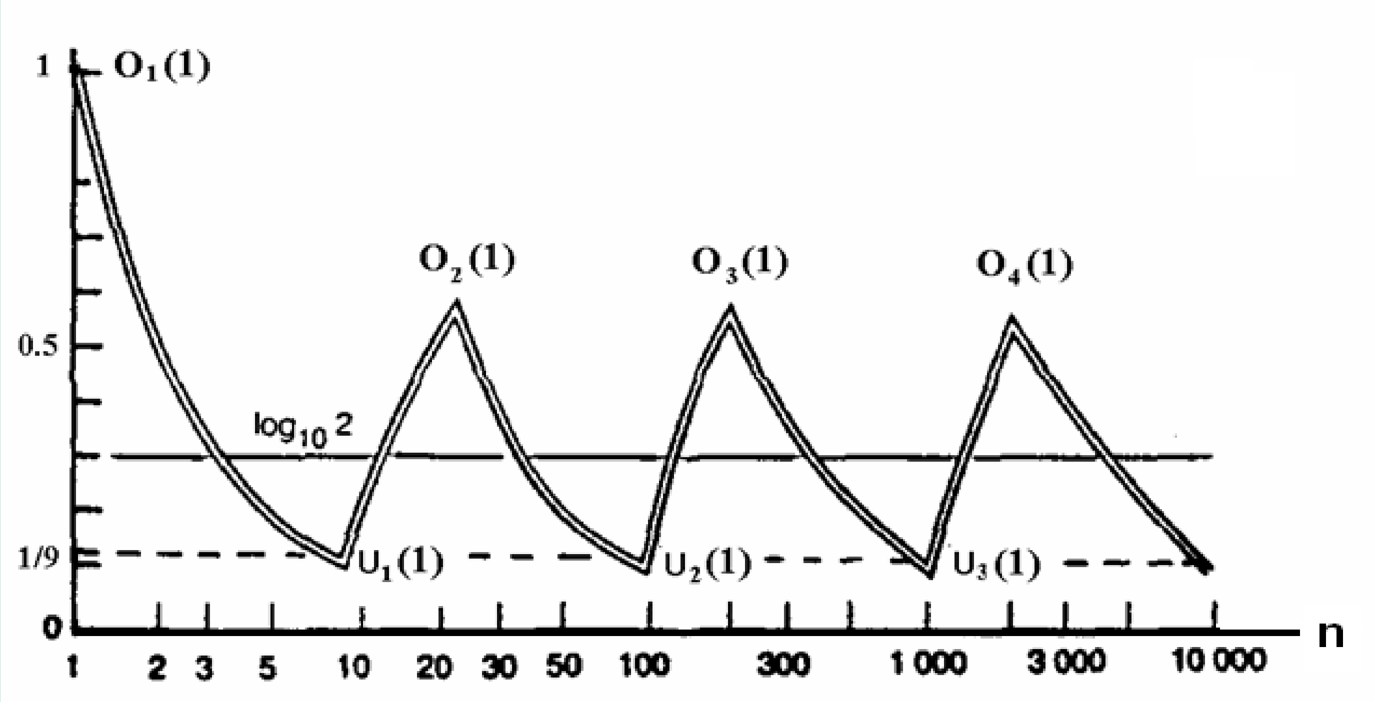
n	1	2	3	4	5	6	7	8	9
$P_n(1)$	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9

Aufholphase:

n	10	11	12	13	14	15	16	17	18	19
$P_n(1)$	2/10	3/11	4/12	5/13	6/14	7/15	8/16	9/17	10/18	11/19

Das Auf und Ab von P$_n$(1):

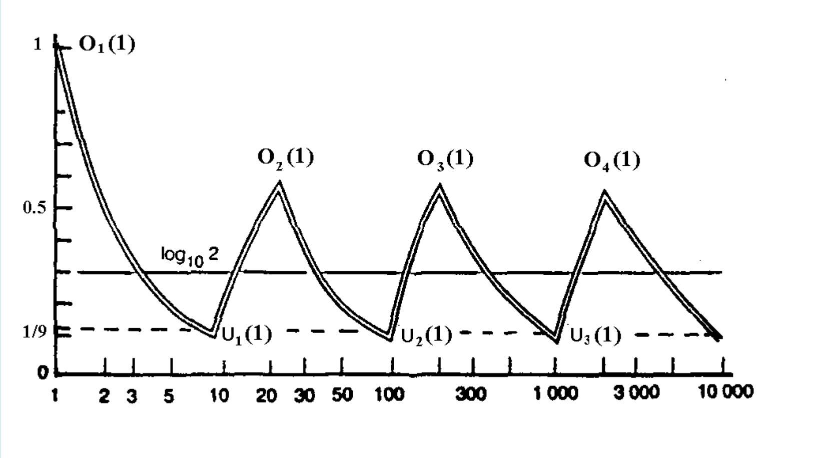
(als kontinuierliche Funktion, n-Achse logarithmisch skaliert)



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

Wahrscheinlichkeiten $P_n(1)$: Obergrenzen $O_m(1) = \frac{11\ldots1}{19\ldots9}$ und Untergrenze $\frac{1}{9}$

Schwankungen bleiben: $\nexists \lim\limits_{n\to\infty} \mathrm{P}_n(1)$



Aber: $U(1) = \lim\limits_{m\to\infty} U_m(1) = \frac{1}{9}$

$$O(1) = \lim_{m\to\infty} O_m(1)$$

Dann für erste Näherungszwecke: $\mathrm{P}(1) \approx \dfrac{O(1)+U(1)}{2}$

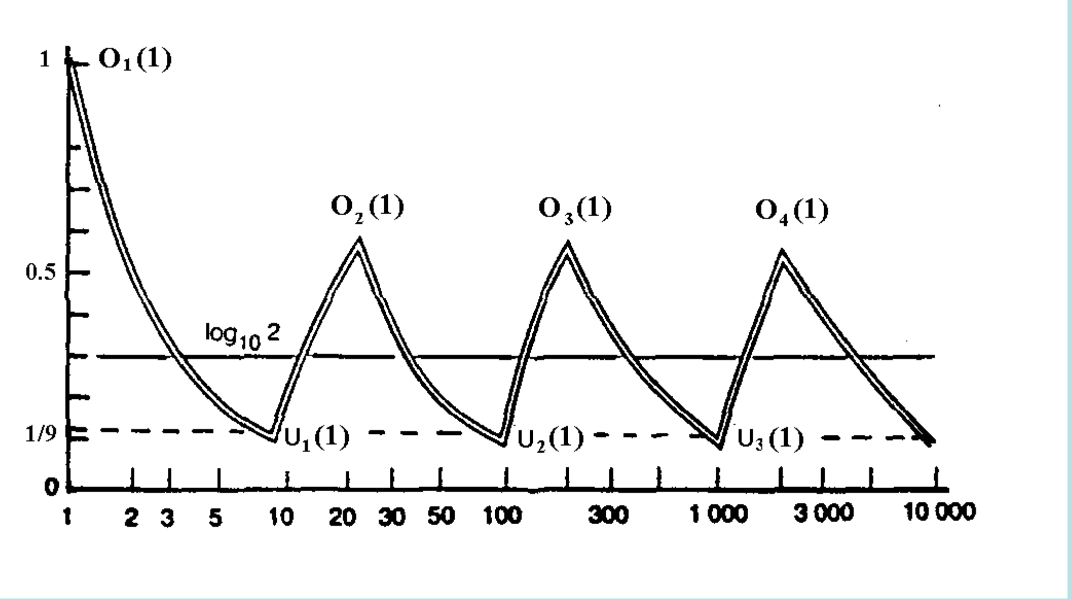
$O_m(1) = \dfrac{11\ldots1}{19\ldots9}$ jeweils m Ziffern in Zähler und Nenner!

Zähler: $\frac{1}{9}\underbrace{(10^m-1)}_{99\ldots9}$ Nenner: $\frac{1}{5}10^m - 1 = \frac{1}{5}(10^m-5)$

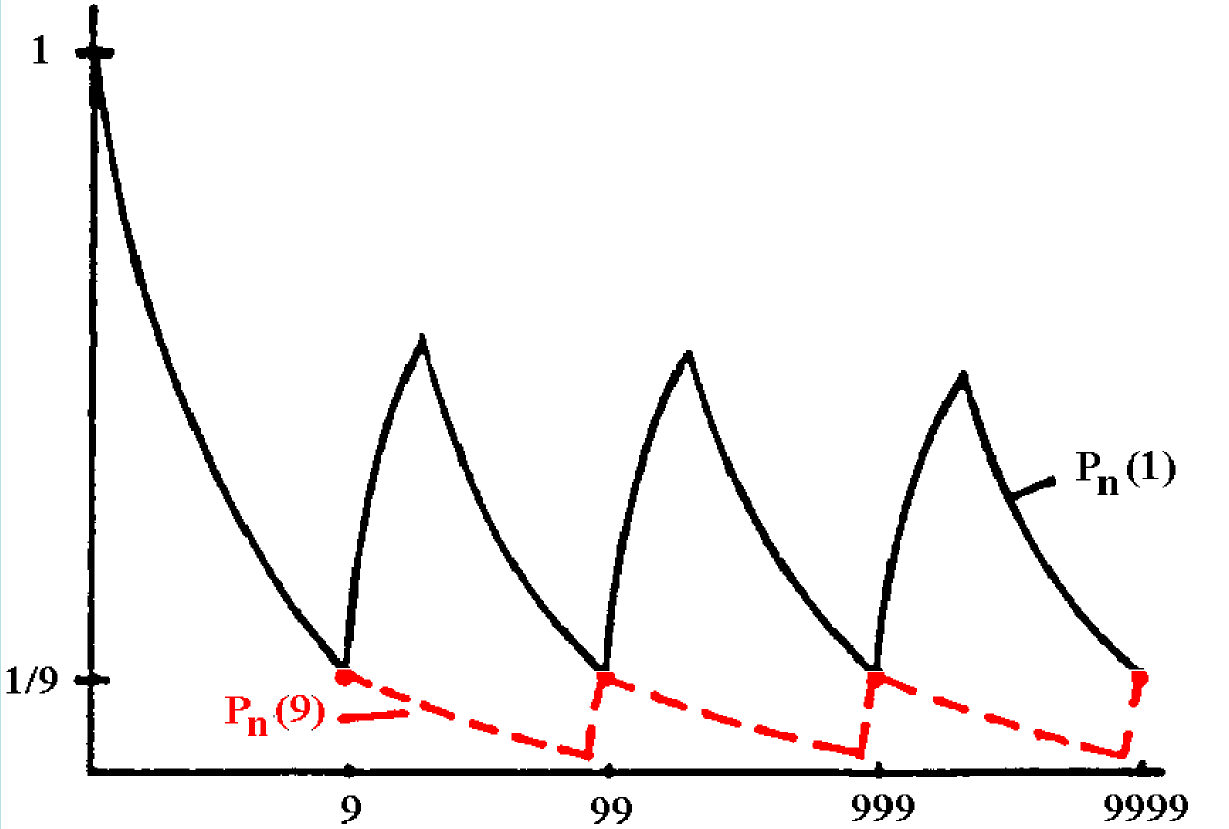
$$O_m(1) = \frac{11\ldots1}{19\ldots9} = \frac{\frac{1}{9}(10^m-1)}{\frac{1}{5}(10^m-5)} = \frac{5}{9}\cdot\frac{10^m-1}{10^m-5} \to \frac{5}{9} = O(1)$$

Näherungswert: $P(1) \approx \frac{3}{9} = \frac{1}{3} = 0,333$

Benford: $P(1) = \log_{10} 2 \approx 0,301$



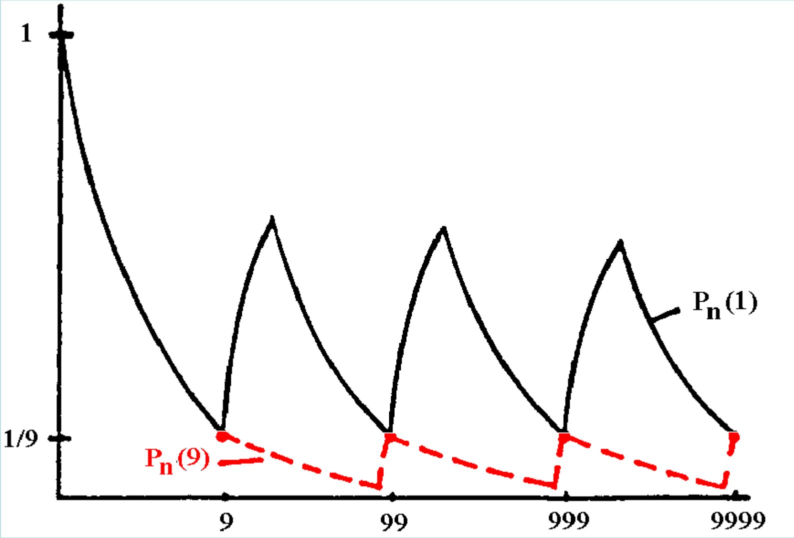
Nun dasselbe mit der Ziffer 9:



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

Wahrscheinlichkeiten $P_n(9)$: Obergrenze $\frac{1}{9}$ und Untergrenzen $U_m(9)=\frac{1\ldots1}{89\ldots9}$

Hier schon plausibel: P(9) < P(1) !



Nur bei den Zahlen 9, 99, 999, etc.:

$P_n(9) = P_n(1) = 1/9,$

sonst immer: $P_n(9) < P_n(1)$!

Salopp formuliert:

Nach jeder neu dazu kommenden Dezimalstelle wird die Ziffer 1 lange Zeit bevorzugt (zwischen 10…0 und 19…9), bevor die anderen Ziffern ihren Rückstand der Reihe nach wieder aufholen: zum Schluss erst die Ziffer 9!

Rechnerisch:

$$O_m(9) = \frac{1}{9} = O(9) \qquad U_m(9) = \frac{\overbrace{1\ldots1}^{m}}{8\underbrace{9\ldots9}_{m}} = \frac{\frac{1}{9}(10^m - 1)}{9 \cdot 10^m - 1} = \frac{1}{81} \cdot \frac{10^m - 1}{10^m - \frac{1}{9}} \rightarrow \frac{1}{81} = U(9)$$

Näherung: $\mathrm{P}(9) \approx \dfrac{O(9) + U(9)}{2} \approx 0{,}062$ Benford: 0,046

Überblick zu dieser Methode auch für die anderen Ziffern:

d	[U(d),O(d)]	(U(d)+O(d))/2	P(d) = log (d+1) – log d
1	$\left[\frac{1}{9},\frac{5}{9}\right]$	0,333	0,301
2	$\left[\frac{1}{18},\frac{10}{27}\right]$	0,213	0,176
3	$\left[\frac{1}{27},\frac{5}{18}\right]$	0,157	0,125
4	$\left[\frac{1}{36},\frac{2}{9}\right]$	0,125	0,097
5	$\left[\frac{1}{45},\frac{5}{27}\right]$	0,104	0,079
6	$\left[\frac{1}{54},\frac{10}{63}\right]$	0,089	0,067
7	$\left[\frac{1}{63},\frac{5}{36}\right]$	0,077	0,058
8	$\left[\frac{1}{72},\frac{10}{81}\right]$	0,069	0,051
9	$\left[\frac{1}{81},\frac{1}{9}\right]$	0,062	0,046

Summe: 1,229 1,000

Überblick (korrigiert):

d	[U(d),O(d)]	(U(d)+O(d))/2	P(d) = $\log_{10}(d+1) - \log_{10} d$
1	$\left[\frac{1}{9},\frac{5}{9}\right]$	0,333 **(0,271)**	0,301
2	$\left[\frac{1}{18},\frac{10}{27}\right]$	0,213 **(0,173)**	0,176
3	$\left[\frac{1}{27},\frac{5}{18}\right]$	0,157 **(0,128)**	0,125
4	$\left[\frac{1}{36},\frac{2}{9}\right]$	0,125 **(0,102)**	0,097
5	$\left[\frac{1}{45},\frac{5}{27}\right]$	0,104 **(0,084)**	0,079
6	$\left[\frac{1}{54},\frac{10}{63}\right]$	0,089 **(0,072)**	0,067
7	$\left[\frac{1}{63},\frac{5}{36}\right]$	0,077 **(0,063)**	0,058
8	$\left[\frac{1}{72},\frac{10}{81}\right]$	0,069 **(0,056)**	0,051
9	$\left[\frac{1}{81},\frac{1}{9}\right]$	0,062 **(0,050)**	0,046

Plausibilitätserklärungen bei nicht natürlichen Zahlen

- **Größen mit zeitlichen Veränderungen:** (Dworschak 1998, S.229)
  „Die 1 ist auf der Zahlenskala von der 2 nicht weiter entfernt als die 5 von der 6. Für die wirklichen Dinge allerdings, die gezählt, gemessen oder gewogen werden, kann der Weg von der 1 zur 2 sehr lang sein:
  Um ihn zurückzulegen, müssen sie auf das Doppelte wachsen.
  Einer 5 fehlt dagegen nur ein Fünftel, um zur 6 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 1 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 5 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 1.“

  Analog: Kontostände, Regenmengen, . . .

Plausibilitätserklärung bei Größen, die sich nicht zeitlich ändern:

- Es gibt vermutlich insgesamt mehr kleine Dinge als große:

  (Grundstücke; Kieselsteine – Felsbrocken),

  d. h. es gibt wahrscheinlich mehr Objekte

  - zwischen 10 E und 20 E als zwischen 20 E und 30 E
  - zwischen 100 E und 200 E als zwischen 200 E und 300 E

  etc.

  Analog: Lacken – Teiche – Seen – Ozeane

  WARUM das so ist, dadurch nicht erklärt!

Mathematische Erklärungen bei positiv-reellen Messwerten

Bis jetzt:

- Natürliche Zahlen
- Plausibilitätserklärungen

Wie groß ist die „Wahrscheinlichkeit", dass ein positiv-reeller Messwert Anfangsziffer 1 hat?

Wenn man *Gleichverteilung*
(d. h. keine Bevorzugung) voraussetzt:
Wahrscheinlichkeiten =
„relative Anteile" von Mengen

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|\text{günstig}|}{|\text{möglich}|}$$

Beispiel: Tunnel – Lüftung

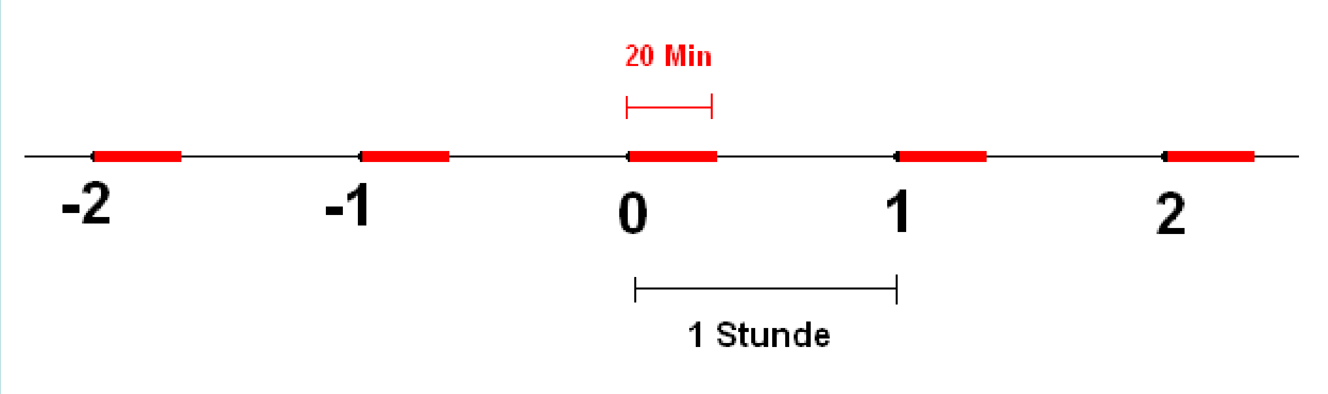
- Die Lüftung eines Tunnels ist von jeder vollen Stunde an 20 Minuten lang in Betrieb (z. B. 00.00 – 00.20 Uhr).

  Wie groß ist die Wahrscheinlichkeit, dass ein zu einem „zufälligen Zeitpunkt“ in den Tunnel einfahrendes Auto die Lüftung in Betrieb vorfindet?

  „zufällig“ . . . kein Zeitpunkt bevorzugt („Gleichverteilung“)

Innerhalb jeder vollen Stunde ist
die Wahrscheinlichkeit abzulesen:

$$\text{Gleichverteilung}: \frac{|\text{günstig}|}{|\text{möglich}|} = \frac{20}{60} = \frac{1}{3}$$

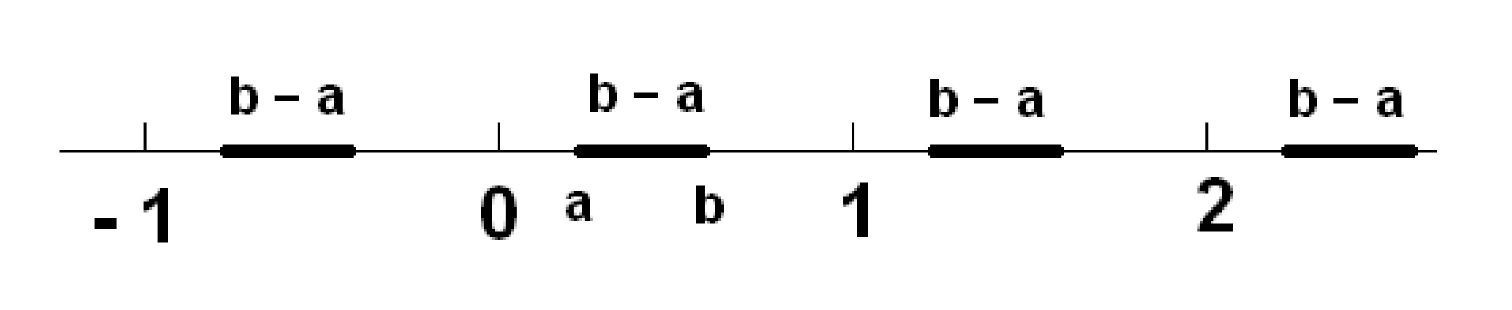


Dies gilt auch insgesamt, d. h. bei einer (rein theoretisch!) unbeschränkten Menge: Die rote Menge macht 1/3 der gesamten Zahlengeraden aus!

Im Allgemeinen (bei unbeschränkten Mengen) aber gar nicht so leicht („Maßtheorie“):
Welchen relativen Anteil hat eine bestimmte Menge???

Verallgemeinerung des Tunnel-Beispiels:
Vereinigung unendlich vieler Intervalle:

$$K_{a,b} := \bigcup_{n=-\infty}^{\infty} [n+a\ ,\ n+b[\qquad \text{mit} \qquad 0 \le a \le b \le 1$$



Wie groß ist der „relative Anteil“ von $K_{a,b}$ in Bezug auf ganz **R** ? Wie groß ist die W‘, dass eine reelle Zufallszahl Z in $K_{a,b}$ liegt (bei „Gleichverteilung“)?

Klar:

$$\mathrm{P}(Z \in K_{a,b}) = \frac{b-a}{1} = b-a$$

Eine zunächst „vordergründige" Argumentation

$Z_d :=$ Menge aller **positiven** reellen Messwerte, die als Dezimalzahl mit d $(1,\ldots,9)$ beginnen

$$Z_1 = \bigcup_{n=-\infty}^{\infty} [1\cdot 10^n\ ,\ 2\cdot 10^n[\qquad\qquad Z_2 = \bigcup_{n=-\infty}^{\infty} [2\cdot 10^n\ ,\ 3\cdot 10^n[$$

Allgemein:

$$Z_d = \bigcup_{n=-\infty}^{\infty} [d\cdot 10^n\ ,\ (d+1)\cdot 10^n[\qquad d = 1,\ldots,9$$

Vereinigung unendlich vieler Intervalle $\subset \mathbb{R}^+$;
leider hier die W' bzw. der „relative Anteil" nicht so einfach:

$$\mathrm{P}(Z \in Z_d) = ?$$

Wir wollen zeigen: $P(Z \in Z_d) = P((1.\text{Ziffer von } Z) = d) =$

$$= \log_{10}(d+1) - \log_{10} d \quad = \quad \log_{10} \frac{d+1}{d}$$

$\Rightarrow$ 10-Logarithmus liegt nahe: $\lg := \log_{10}$

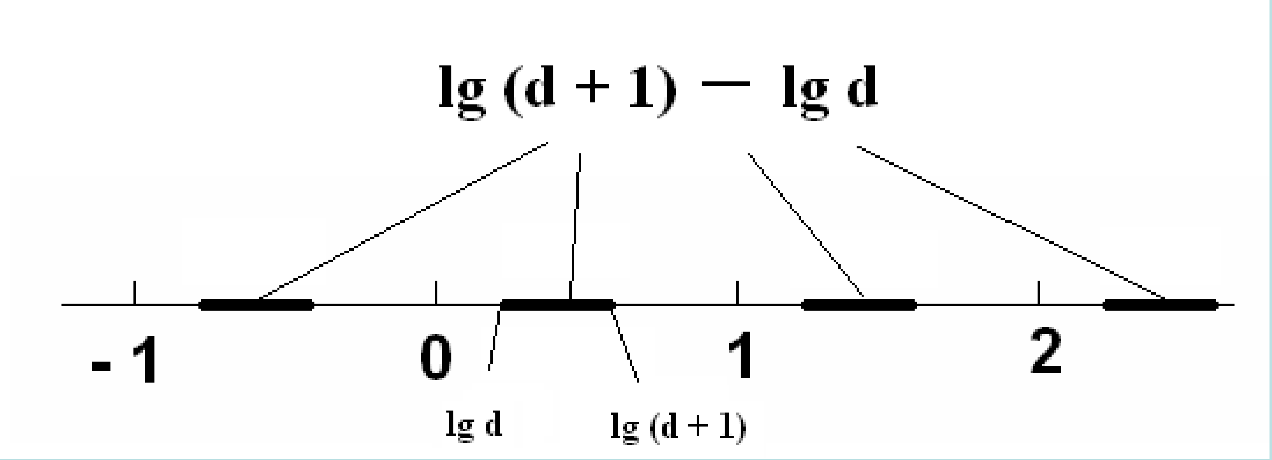
$$Z_d = \bigcup_{n=-\infty}^{\infty} [d \cdot 10^n \ , (d+1) \cdot 10^n[\qquad d = 1, \ldots, 9$$

$$\lg Z_d = \bigcup_{n=-\infty}^{\infty} [n + \lg d \ , n + \lg(d+1)[\qquad d = 1, \ldots, 9$$

Mit Hilfe des Log:

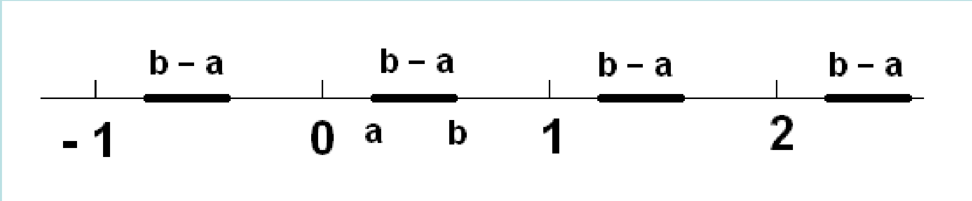
Mult. → Addition

(hier viel leichter!)



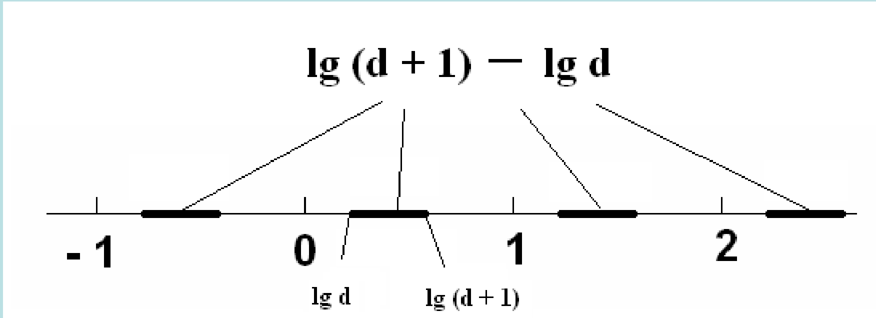
Wie oben bei

$$P(Z \in K_{a,b}) = \frac{b-a}{1} = b-a$$



Nun analog:

$$P(\lg Z \in \lg Z_d) = \lg(d+1) - \lg d$$



Wegen $\lg Z \in \lg Z_d \;\Leftrightarrow\; Z \in Z_d$

damit auch das Benford-Gesetz plausibel:

$$\boxed{P(Z \in Z_d)} = P(\lg Z \in \lg Z_d) \stackrel{?}{=} \boxed{\lg(d+1) - \lg d}$$

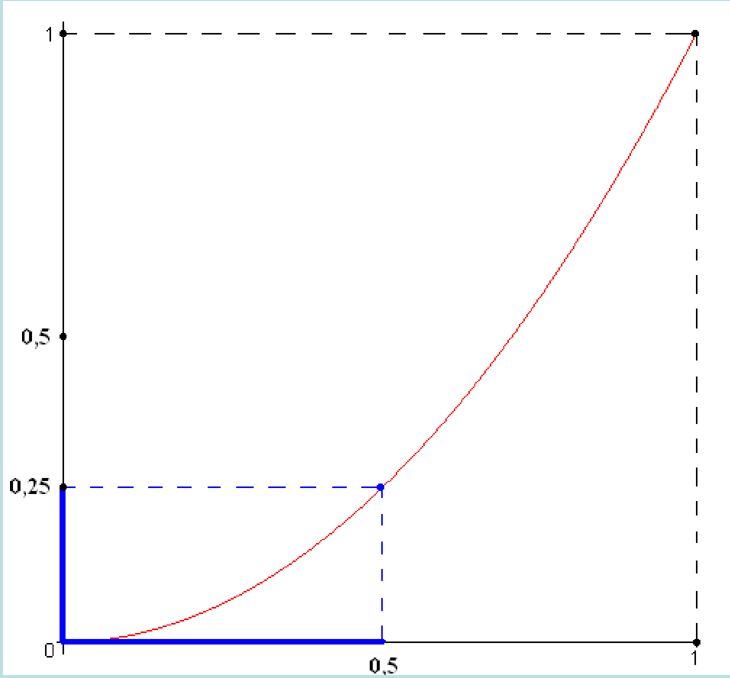
? Warum herrscht „Gleichverteilung“ bei $\lg Z$?
Warum darf/muss man bei der W‘ in Logarithmen denken?
(einstweilen „vordergründig“)

Wenn man auf eine Zufallsvariable eine Funktion anwendet („Transformation“), so ändert sich dabei i. A. auch das Verteilungsgesetz!

Beispiel: Zufallszahl Z aus [0 ; 1[– Gleichverteilung:

$$I := \left[0; \frac{1}{2}\right[\qquad \mathrm{P}(Z \in I) = \frac{1}{2}$$

$$f : x \mapsto x^2 \qquad f([0;1[) = [0;1[\qquad f(I) = \left[0; \frac{1}{4}\right[$$



Ohne Beachtung einer Änderung des Verteilungsgesetzes (wieder „naiv“ Gleichverteilung vorausgesetzt):

$$\mathrm{P}(f(Z) \in f(I)) = \frac{1}{4}$$

Widerspruch:

$$Z \in I \iff f(Z) \in f(I)$$

Anwendungen – reale Bedeutung?

- Lottospielen? Nein!
- **Steuerprüfung** bei großen Firmen:
  Datenmanipulation → 1, 2 zu selten vorne!
  Marc NIGRINI:
  Software Mitte 90er
  (Bei positivem Test: Verdacht!
  Gesetz seither erst so richtig bekannt!)

- Amerika, Deutschland, Schweiz

Anwendungen

- Angeblich: Bei der Entdeckung der Bilanzfälschungsskandale von ENRON (2001) und WORLDCOM (2002) war diese Software beteiligt
- Angeblich: Bill Clinton und Bill Gates wurden „überprüft“ → kein Verdacht
- Missbrauch bei „genehmigungspflichtigen Grenzen“ (Reise-, Anschaffungskosten)

Weitere Anwendungen

- Freiburg i. B. (D):
  Verwaltung von Speicherplatz auf Festplatten optimieren
- Belgien: Ungereimtheiten bei Krankenhausabrechnungen aufstöbern
- Wissenschaft: Fälschung von Daten, um ein angestrebtes „signifikantes“ Ergebnis zu erhalten

„Richtiges“ Datenfälschen ist gar nicht so leicht: entsprechendes Gesetz auch für die folgenden Ziffern:

- Häufigkeiten von 1 (0) bis 9 immer abnehmend,

  aber die Unterschiede werden immer kleiner:

  Ziffern folgen umso mehr einer „Gleichverteilung“, je kleiner ihr Stellenwert ist.

- Allzu genau ist auch verdächtig!

- auch noch andere Gesetzmäßigkeiten . . .

Position of digit in number

Digit	First	Second	Third	Fourth
0		.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.10882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982

Welche Zahlen (Daten) gehorchen sicher nicht dem Benford-Gesetz?

- Alter von Studierenden (1,2,3)
  oder Politikern (3,4,5,6)
- Zeiten von Marathon- oder 400m-Läufen
- etc.

D. h.: Daten, die von vornherein auf einen bestimmten relativ engen Bereich eingeschränkt sind!

Mathematische Begründung – „Skaleninvarianz“

- Unsere „Einheiten“ sind höchst willkürlich (m, kg, €, Liter, m², . . .); je nach EH: verschiedene Maßzahlen

- Wenn es überhaupt ein Verteilungsgesetz für die 1. Ziffer gibt → „universelles“ Gesetz, das nicht von den EH abhängt!

- EH-Wechsel bedeutet i. A.: Multiplikation mit einer positiven Zahl („Umrechnungsfaktor“) $s \in \mathbb{R}^+$:

  z. B. Meilen → km: $s \approx 1,6$

D. h. ein „vernünftiges“ Verteilungsgesetz für die 1. Ziffer darf sich nicht ändern, wenn alle denkmöglichen Werte mit $s \in \mathbb{R}^+$ multipliziert werden:

„skaleninvariant“

Werden sehen: Forderung nach Skaleninvarianz
→ Benford-Gesetz ist eine notwendige Folgerung!

Welche Verteilungsgesetze kommen dafür in Frage?

Die Gleichverteilung? (alle Ziffern gleich wahrscheinlich)

Z. B. Währungsänderung: € → DM (Beträge verdoppelt!)

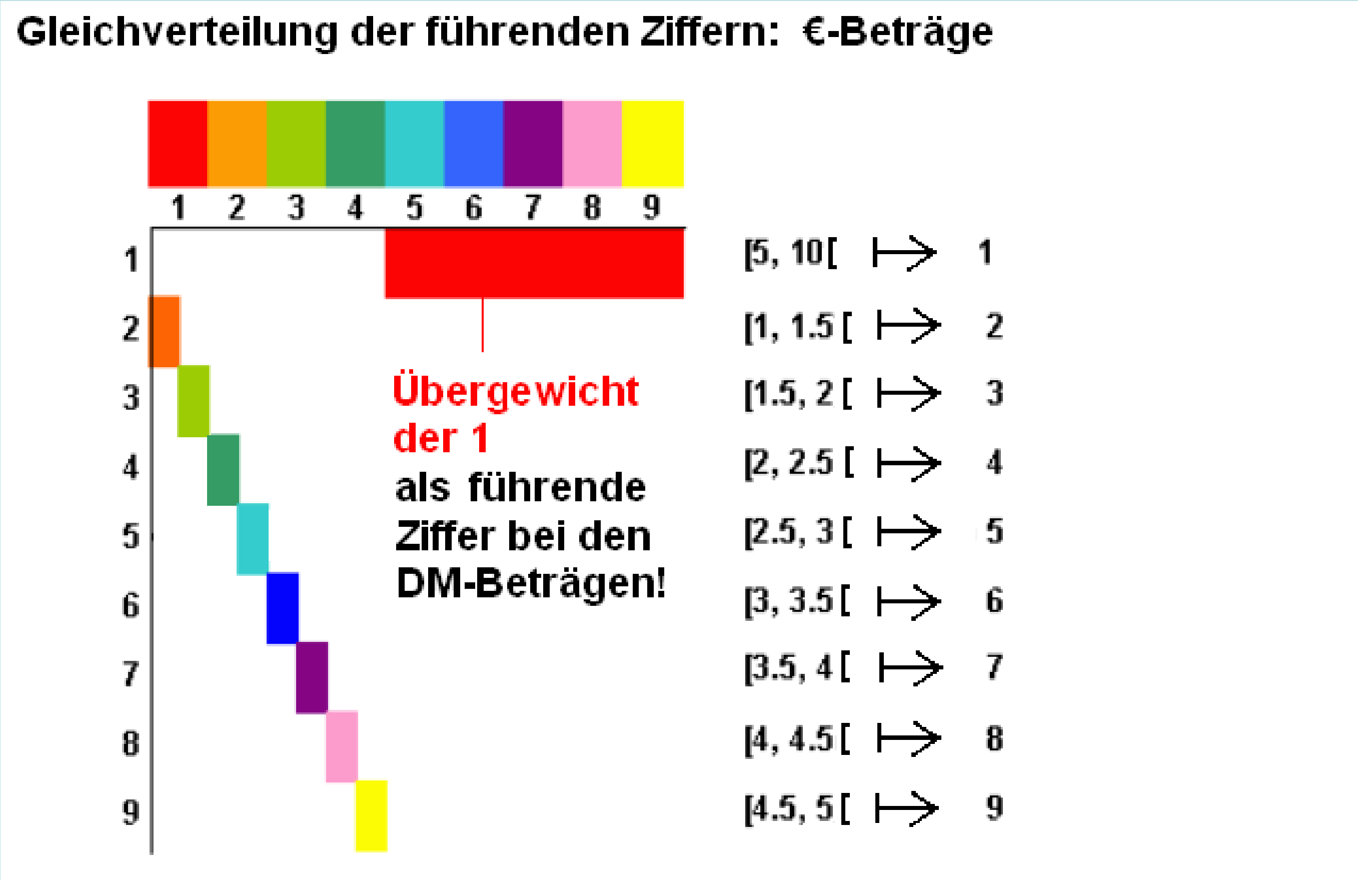
Gleichverteilung

kommt nicht

in Frage,

sie ist nicht

skaleninvariant!



Nur „Anfangsziffer“ interessant →
„wissenschaftliche“ Schreibweise (Zehnerpotenzen):

$$Z = M \cdot 10^n \qquad (1 \leq M < 10:\ \text{"Mantisse"})$$

Beschränkung auf M statt Z !

Suche: skaleninvariantes Verteilungsgesetz für Z

$\overset{\text{klar!}}{\Longleftrightarrow}$ skaleninvariantes Verteilungsgesetz für M
(d. h. M und $M \cdot s$ haben dasselbe Verteilungsgesetz!)

→ Logarithmieren: $0 \leq \lg M < 1$
$\lg M$ und $\lg(M \cdot s)$ haben dasselbe Verteilungsgesetz:

Dieses muss also unverändert bleiben,
wenn man eine beliebige Konstante addiert:

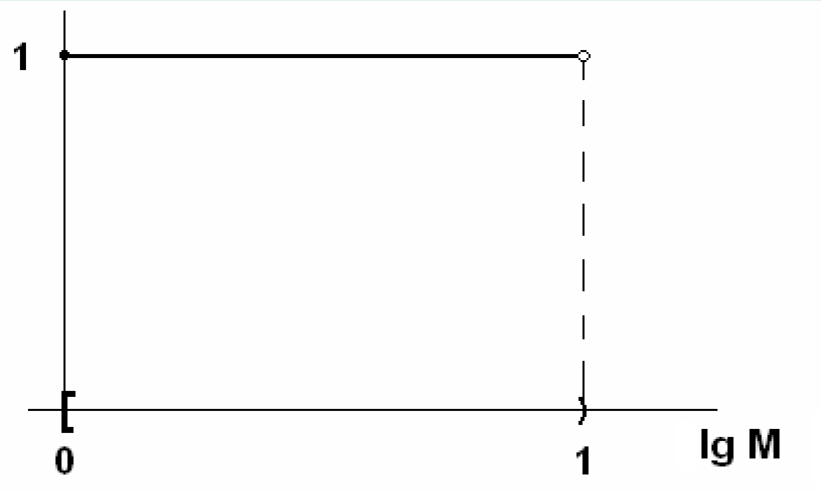
$$\lg(M \cdot s) = \lg M + \lg s$$

Leichter: welche Verteilungsgesetze bleiben durch **beliebige Additionen** unverändert?

Verteilungsgesetze durch **Dichtefunktionen** beschreibbar; *W'en* dann einfach: *Flächeninhalte unter Dichtefunktionen*

Die konstante Funktion ist die einzige Dichtefunktion, die durch beliebige Additionen (d. h. beliebiges horizontales Verschieben) unverändert bleibt („Gleichverteilung"):

lg M muss **gleichverteilt** sein!

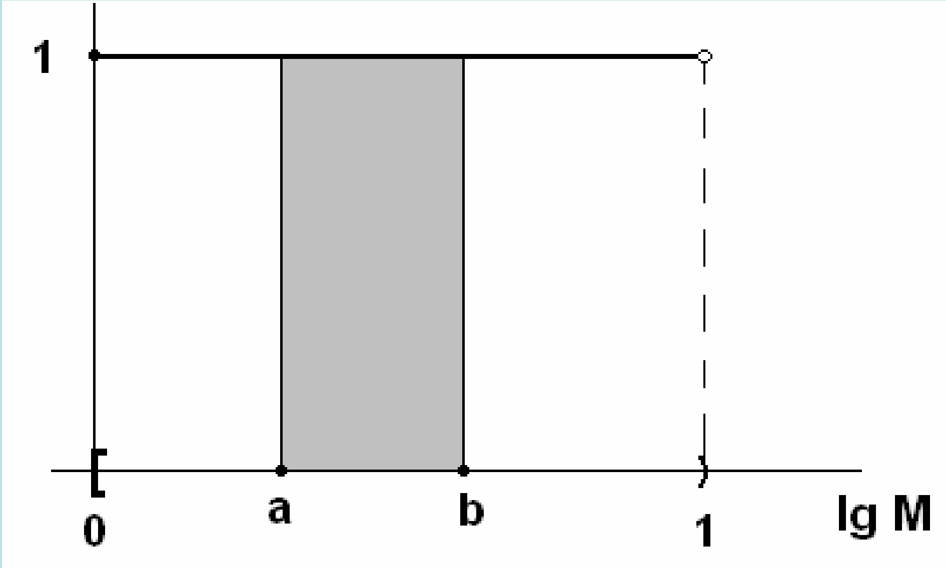


$$1 \leq M < 10$$

$$0 \leq \lg M < 1$$

Wahrscheinlichkeiten (Flächeninhalte)

bei Gleichverteilung ganz leicht zu bestimmen:



$$P(a \le \lg M < b) = (b-a)\cdot 1 = b-a$$

Dies ist **ideal** für die interessierenden Wahrscheinlichkeiten:

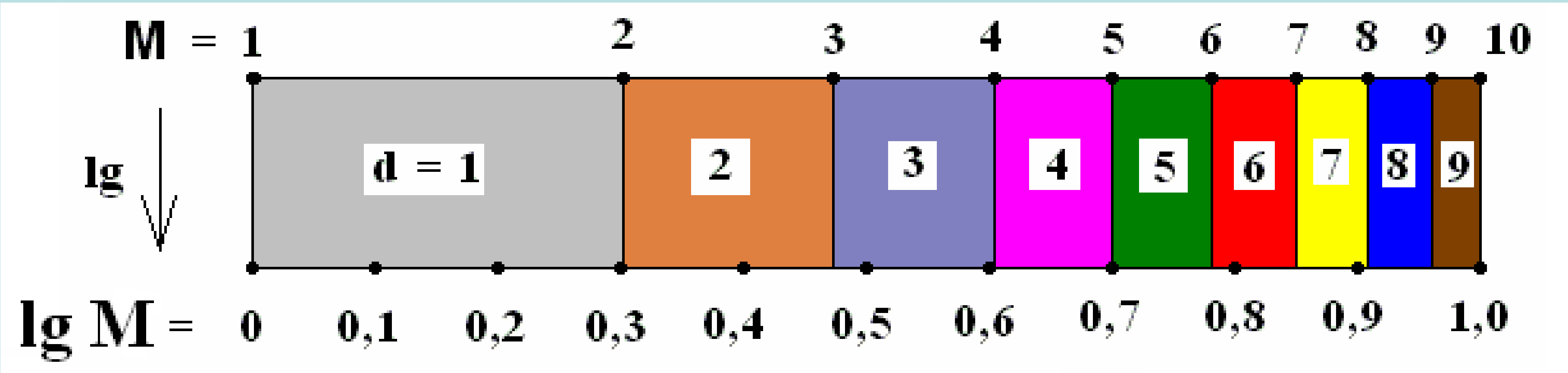
$$\boxed{P(1.\text{Ziffer von } Z = d)} = P(1.\text{Ziffer von } M = d) = P(d \le M < d+1)$$

$$= P(\lg d \le \lg M < \lg(d+1)) \underset{\substack{\text{Gleichverteilung}\\ \text{von } \lg M\,!}}{\overset{?}{=}} \boxed{\lg(d+1) - \lg d = \lg \frac{d+1}{d}}$$

– das Benford-Gesetz!

Gleichverteilung von $\lg M$ auf [0 ; 1[

→ „logarithmische“ Verteilung
der Mantisse M auf [1 ; 10[:



Andere Zahlensysteme

- **Dual**-system (Ziffern 0 und 1):
  1 ist immer die führende Ziffer!

- Andere natürliche Zahl $a > 2$ als „Basis“ des Zahlensystems:

$$\mathrm{P}(Z \in Z_d) = \log_a(d+1) - \log_a d \qquad (d = 1, 2, \ldots, a-1)$$

- Gesetz gilt auch, wenn man realistischerweise bei „Messwerten“ $\mathbb{Q}_0^+$ statt $\mathbb{R}_0^+$ voraussetzt.

Fachdidaktisches Potential

- Interessantes Phänomen (Motivation)
- Historische Bezüge
- Schüler/innen können selbständig Experimente machen (empirisch überprüfen: Google)
- Mögliche Anwendungen: Nigrini (Steuer)
- Auch für Schüler/innen lesbare Literatur (auch populärwissenschaftlich)
- Ausbau und Vertiefung möglich

Fachdidaktisches Potential

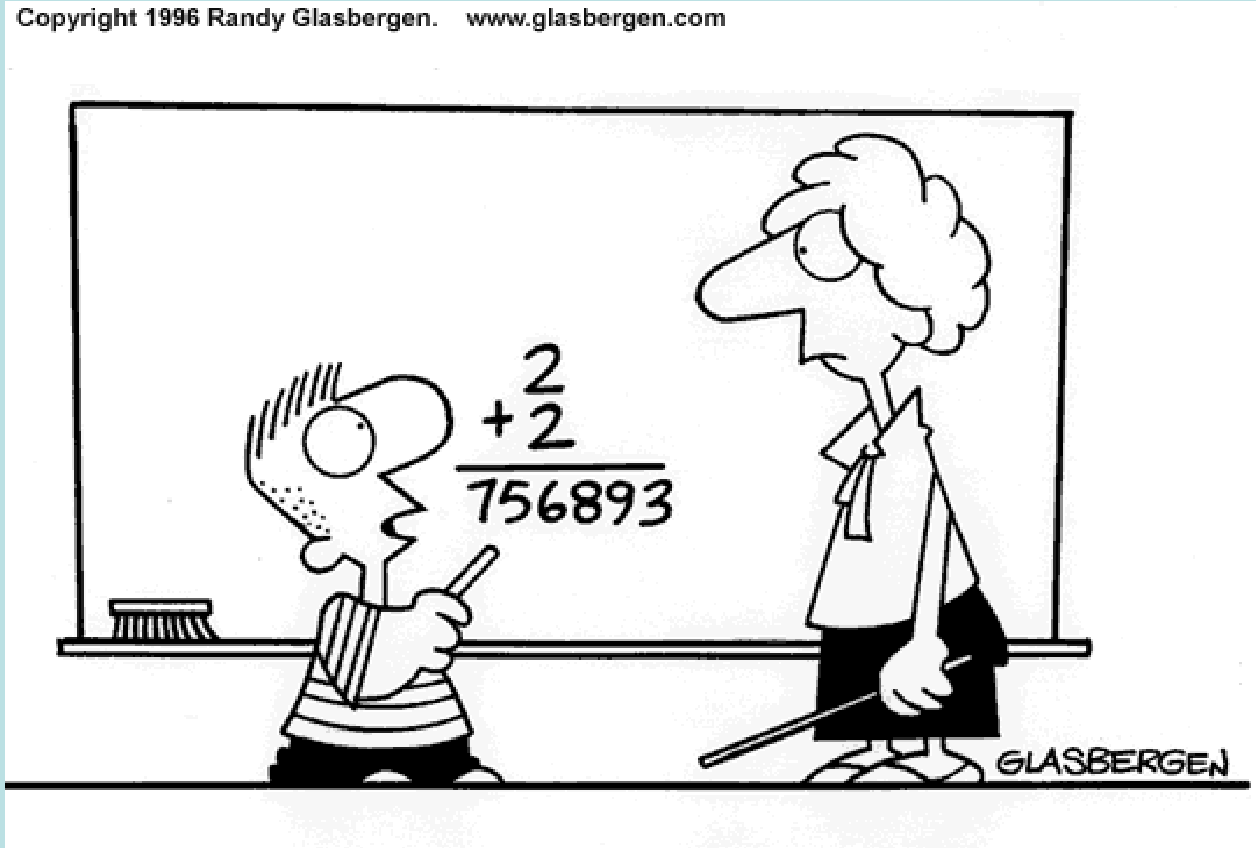
- Verschiedene Erklärungen (Stufung der Strenge): Plausibilitätsbetrachtungen und mathematischer Beweis
- Vernetzung verschiedener math. Inhalte: Wahrscheinlichkeit, Zahlendarstellungen-Grenzwerte, Logarithmus, Dichtefunktion, „Nicht immer Gleichverteilung“, „Skaleninvarianz“

Danke für Ihre Aufmerksamkeit !

Literaturauswahl

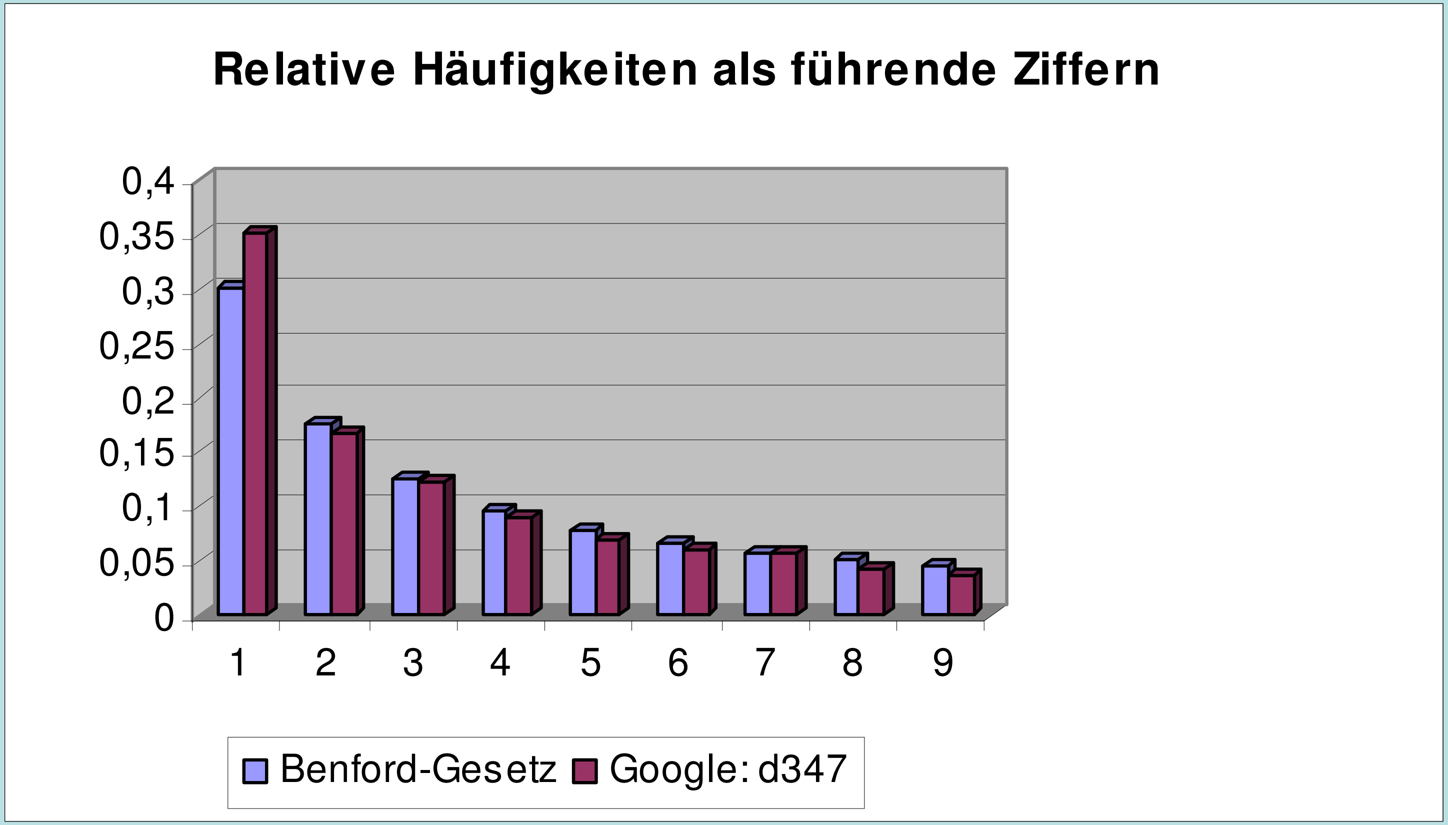
- **Albrecht, J. (2000):** Die Eins von Planet Zob.
  Die Zeit (40, 28. 09. 2000), 35.
- **Dworschak, M. (1998):** Weiter Weg zur Zwei.
  In: Der Spiegel 47/1998, 228 – 229.
- **Humenberger, H. (2000):** Das BENFORD-Gesetz –
  warum ist die Eins als führende Ziffer von Zahlen bevorzugt?
  In: ISTRON-Band 6, 138–150, Franzbecker, Hildesheim.
- **Humenberger, H. (2008):** Eine elementarmathematische
  Begründung des Benford-Gesetzes.
  In: Der Mathematikunterricht 54,1, 24 – 34.
  Leicht überarbeitet:
  http://www.oemg.ac.at/DK/Didaktikhefte/2008%20Band%2041/VortragHumenberger.pdf
- **Matthews R. (1999):** The power of one.
  In: New Scientist 2194 (July 10), 27– 30.
- **Walthoe,J. / Hunt R. / Pearson, M.:**
  Looking out for number one.
  http://plus.maths.org/issue9/features/benford/

Danke für Ihre Aufmerksamkeit !



“In an increasingly complex world, sometimes old questions require new answers.”

d	Google-Häufigkeit von: „d347“
1	35.700.000
2	17.000.000
3	12.400.000
4	9.110.000
5	7.170.000
6	6.250.000
7	5.720.000
8	4.370.000
9	3.680.000



Google-Versuch zeigt auch:

1 am häufigsten „vorne“!

9 am seltensten „vorne“!