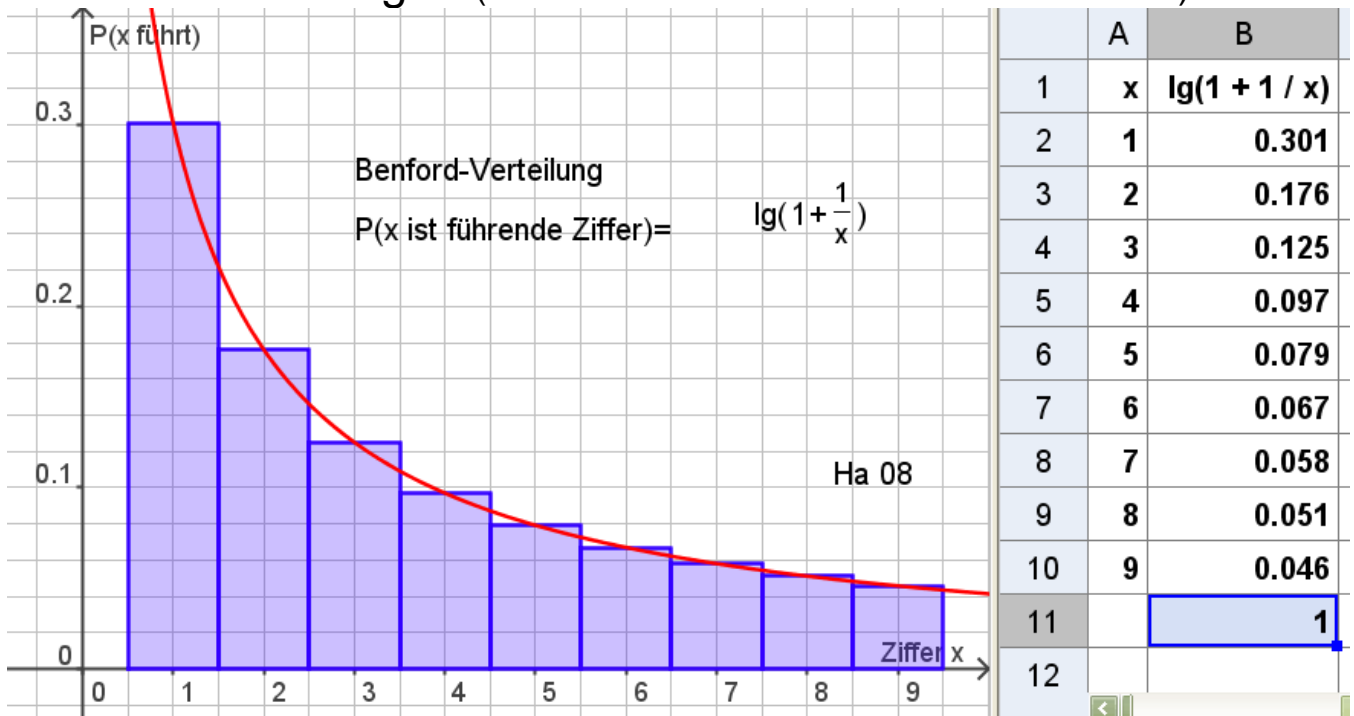


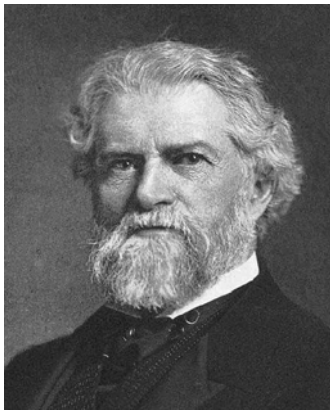
Benford-Verteilung (Beweis des Benford-Gesetzes auf Seite 2)



Historisches

Im Jahre 1881 entdeckte der Mathematiker Simon Newcomb, dass die Seiten einer fünfstelligen Logarithmentafel für die kleine führende Ziffernfolgen wesentlich stärker abgegriffen waren als für große.

1. Für die Zahl 132,6 bestimmt man die zugehörige Zehnerpotenz, hier die 2.
2. Dann sieht man in der Tafel die sogenannte „Mantisse“ für die Ziffernfolge 1326 nach. 1326 heißt „Numerus“ der gegebenen Zahl.
3. Man findet 12254. Dieses nimmt man als Nachkommastellen hinter der 2.
4. Damit hat man $\lg(132,6)=2.12254$ bestimmt.



Newcomb



Logarithmentafel



Benford

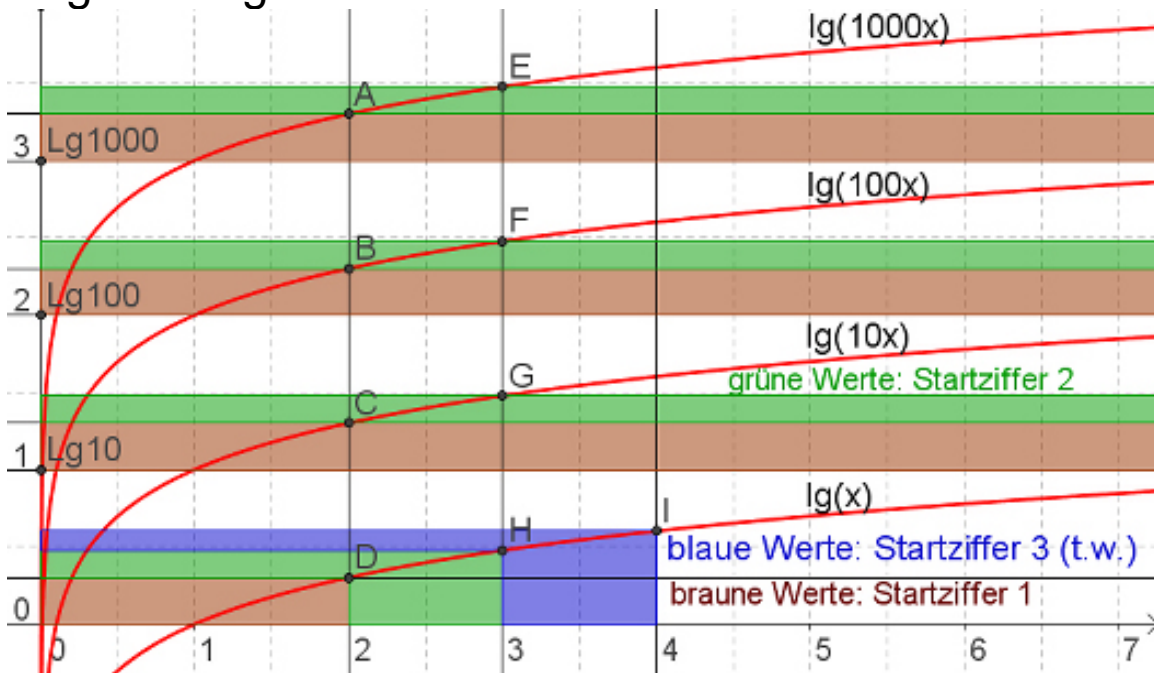
Bilder und Informationen aus: Norbert Hünigerbühler, EthZürich <http://www.educ.ethz.ch>

Newcomb veröffentlichte seine Beobachtung, stellte auch schon eine logarithmische Formel auf, aber seine Arbeit wurde nicht beachtet.

Im Jahr 1938 entdeckte der Physiker Frank Benford das Gesetz neu und untermauerte es mit Daten. Er bewies es aber nicht.

Erst 1995 deckte der Mathematiker Theodore Hill genaueres auf und bewies auch noch weitere Zusammenhänge. Diese setzte der Mathematiker Mark Negrini in einem Analyseprogramm um, mit dem man die Echtheit von Daten prüfen kann, die „Benford-verteilt“ sein müssten. Dazu gehören vor allem Daten aus exponentiellen Zusammenhängen, aber aggregierte Daten, die selbst nicht benford-verteilt sind, folgen der Benford-Verteilung. Auf diese Weise kann man Wirtschafts- und Bankdaten, wissenschaftliche Messdaten u.a. prüfen und Betrug aufdecken.

Begründung für das Benford-Gesetz



Wegen der Logarithmengesetze gilt: $\lg(10x) = \lg(x) + \lg(10) = \lg(x) + 1$

Darum sind die hier gezeichneten Logarithmusfunktionen alle um den Wert 1 gegeneinander verschoben.

Betrachten wir den grünen Streifen zwischen 2 und 3 auf der x-Achse. D und H haben dann die Ordinaten $\lg(2)$ und $\lg(3)$. Der grüne waagerechte Streifen färbt den Ordinaten-Unterschied. Wir berechnen diesen Unterschied gleich allgemein:

$$\Delta(x) = \lg(a \cdot (x+1)) - \lg(a \cdot x) = \lg\left(\frac{a \cdot (x+1)}{a \cdot x}\right) = \lg\left(\frac{x+1}{x}\right) = \lg\left(1 + \frac{1}{x}\right)$$

Setzt man $x=2$, so erhält man mit $\Delta(2) = \lg\left(1 + \frac{1}{2}\right) = 0,176\dots$ die Höhe **jedes** grünen

Balkens, denn unter dem a , das hier herausfällt, kann man sich 10, 100, 1000 u.s.w. vorstellen.

Die Wahrscheinlichkeit, an eine grüne Stelle zu treffen, wenn man irgendwo an die Ordinatenachse tippt, ist also $P(2 \text{ führt}) = (\text{Summe aller grünen Stücke}) / \text{Gesamtlänge}$

$$= \frac{k \Delta(2)}{1} = \Delta(2) = 0,1763\dots$$

Die zugehörigen Abszissen sind genau alle Zahlen,

die die führende tragende Ziffer 2 haben. Beispiele sind 29870 // 25,7 // 0,0234 // 0,00002019.

Also ist die Wahrscheinlichkeit für führende Ziffer 2: $P(2 \text{ führt}) = \Delta(2) = 0,176\dots$

Genauso kann man für die anderen führenden Ziffern argumentieren.

Braune Streifen entsprechen der führenden Ziffer 1, also $P(1 \text{ führt}) = \Delta(1) = 0,3010\dots$

Blaue Streifen entsprechen der führenden Ziffer 3: $P(3 \text{ führt}) = \Delta(3) = 0,1249\dots$

Da die Logarithmen die Umkehrfunktionen des natürlichen exponentiellen Wachstums sind, gilt für "natürlich" wachsende Zahlen, also etwa Maßzahlen von Größen, allgemein das

Benford-Gesetz Die Wahrscheinlichkeit, dass die führende

Ziffer x ist, wird bestimmt durch $P(x \text{ führt}) = \lg\left(1 + \frac{1}{x}\right)$