

„Das Benfordsche Gesetz“

oder

„Wenn sich Zahlen an Gesetze halten“

Die Wette

Angenommen, jemand schlägt Ihnen die folgende Wette vor:

„Schlagen Sie die Zeitung von heute auf einer beliebigen Seite auf und suchen Sie die erste Zahl, die auf Ihr steht. Ich wette, dass die erste Ziffer dieser Zahl kleiner als 4 ist! Und damit alles mit rechten Dingen zugeht, schließen wir noch Jahreszahlen, Datumsangaben, Uhrzeiten und Telefonnummern aus (die ja in aller Regel mit einer der Ziffern von 1 bis 3 anfangen).“

Würden Sie die Wette annehmen und dagegen setzen?

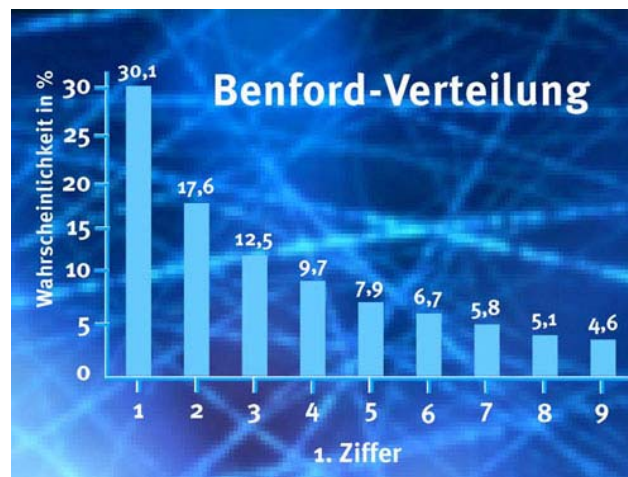
Klingt doch nach einem sicheren Geschäft für Sie? Schließlich gibt es sechs Möglichkeiten, dass eine Zahl mit 4, 5, 6, 7, 8 oder 9 beginnt im Gegensatz zu nur drei Möglichkeiten, dass sie mit 1, 2 oder 3 anfängt.

Ich rate Ihnen dennoch, die Wette nicht anzunehmen, da Ihre Chancen zu gewinnen nur bei 40% liegen!

Wie ist das zu erklären?

Das Erstaunliche

Jeder vernünftige Mensch würde doch zunächst annehmen, dass die Ziffern ,1' bis ,9' als Anfangsziffern von Zahlen gleich wahrscheinlich vorkommen. Dann könnte man vorige Wette getrost annehmen. Tatsächlich kann man aber feststellen, dass die meisten Zahlen mit der Ziffer ,1' anfangen, während die Ziffer ,9' nur relativ selten vorkommt. Genauer gesagt, die Ziffern von ,1' bis ,9' (die Ziffer ,0' kommt hier nicht in Frage) treten mit abnehmender Wahrscheinlichkeit als Anfangsziffern von Zahlen auf. Die Wahrscheinlichkeit ihres Auftretens lässt sich mit Hilfe einer konkreten Wahrscheinlichkeitsfunktion berechnen. Die zugrunde liegende Verteilung nennt sich Benford-Verteilung.



Man erkennt, je niedriger die Anfangsziffer einer Zahl ist, umso häufiger kommt diese Anfangsziffer vor: Zahlen des Zehnersystems mit der Anfangsziffer ,1' sind etwa 6,5-mal häufiger als solche mit der Anfangsziffer ,9'.

Dies ist die Aussage eines der universellen Verteilungsgesetze der Stochastik. Es beschreibt eine fundamentale Gesetzmäßigkeit der Verteilung von Zahlen in Zahlenlisten in Abhängigkeit von ihren ersten Ziffer, seien es Einwohnerzahlen von Städten oder Geldbeträge in der Buchhaltung, Listen von Naturkonstanten oder Ergebnisse wissenschaftlicher Beobachtungen.

Historisches

Alles begann mit einer kuriosen Entdeckung Ende des 19. Jahrhunderts: Damals nutzten Wissenschaftler Logarithmentafeln, um komplizierte Rechnungen zu vereinfachen. Nach einigen Jahren sah man ihnen ihren intensiven Gebrauch an: Die Ecken waren abgegriffen und auch die Seiten vom vielen Blättern deutlich dunkler gefärbt. Das Erstaunliche aber war: Die ersten Seiten waren viel schmutziger als die hinteren – so als ob jene Zahlen, die mit niedrigen Ziffern beginnen, häufiger nachgeschlagen wurden als jene, die mit höheren Ziffern anfangen

Diese Entdeckung veröffentlichte als Erster der Mathematiker **Simon Newcomb** (1835–1909): Im *American Journal of Mathematics* schrieb er im Jahr 1881, dass die ‚1‘ als erste Ziffer einer Zahl mit einer Wahrscheinlichkeit von 30,1% auftauche. Die ‚2‘ als erste Ziffer folge mit 17,6% und die ‚9‘ komme am seltensten vor: Lediglich 4,6% aller Zahlen begännen mit dieser Ziffer.

Newcombs Entdeckung geriet in Vergessenheit. Viele Jahr später – in den 1920er Jahren – waren es wiederum Logarithmentafeln, an denen die charakteristisch



erscheinende Verteilung der Ziffern neu entdeckt wurde: Der Physiker Dr. **Frank Benford** (1883–1948) untersuchte daraufhin die Zahlen aus ganz unterschiedlichen Lebensbereichen. Egal ob er die Länge der Flüsse in den USA, die Fläche von Seen, die Hausnummern einer Stadt oder alle Zahlen in einer Ausgabe des *Reader's*

Digest untersuchte: Er fand überall, dass etwa 30% dieser Zahlen mit einer ‚1‘ beginnen. Benford veröffentlichte seine Beobachtungen schließlich 1938. Die verblüffende Verteilung der Anfangsziffern bei Zahlen ist schließlich auch nach ihrem zweiten Entdecker **Benfords Gesetz** genannt worden und war bis vor wenigen Jahren nicht einmal allen Statistikern bekannt. Da sie schon bei der Lösung zahlreicher praktischer Probleme hilfreich war, wächst ihr Bekanntheitsgrad rasch.

Die Formel und ihre empirische Relevanz

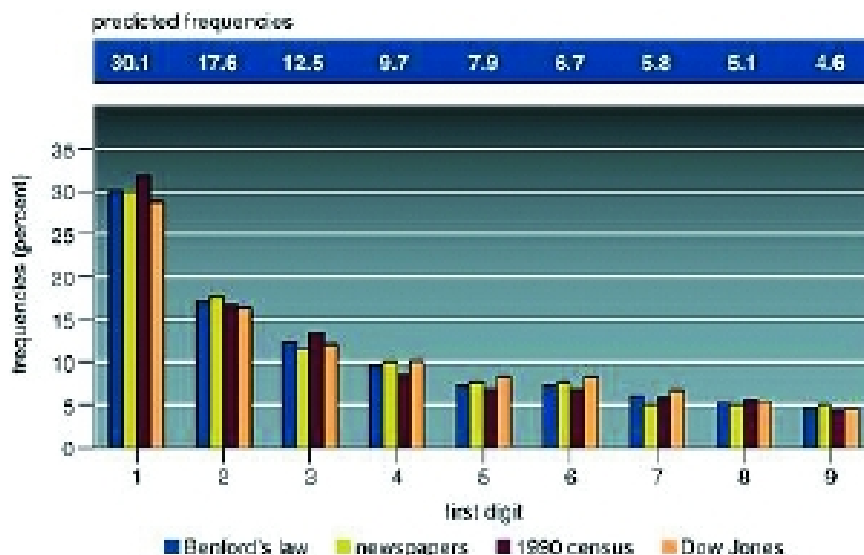
Die Wahrscheinlichkeitsfunktion der Benford-Verteilung stellt sich wie folgt dar. Ist d eine Ziffer von ,1' bis ,9', dann ist die Wahrscheinlichkeit ihres Auftretens als erste Ziffer einer Zahl

$$W(d) = \log_{10} \left(1 + \frac{1}{d} \right).$$

Mehrere Forscher haben die Gültigkeit von Benfords Gesetz inzwischen an größeren Zahlenmengen untersucht:

- Dr. Benford selbst hat Zahlen ausgewertet, die auf Titelseiten von Zeitungen vorkommen;
- Dr. Nigrini hat die Bevölkerungszahlen eines US-Census untersucht;
- Mr. Ley hat die Dow-Jones-Mittelwerte von 1990 bis 1993 betrachtet

... und andere mehr.



Immer wieder scheint sich die Behauptung zu bewahrheiten, mit gewissen geringen Abweichungen.

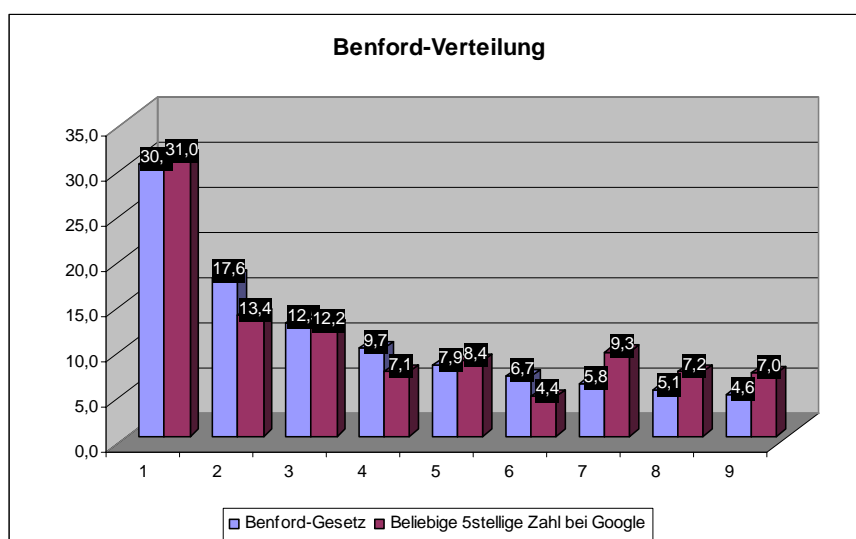
Eine mögliche Erklärung

Dass die ‚1‘ immer „die Nase vorn“ hat, kann man sich schnell plausibel machen, wenn man z.B. Zahlen aus dem Wirtschaftsleben betrachtet: Wenn etwa der Deutsche Aktien-Index DAX von 1000 auf 2000 Punkte steigt, muss er um 100 Prozent zulegen. Der gleiche 1000er-Sprung von 8000 auf 9000 Punkte hingegen ist schon mit einem Zuwachs von nur 12,5 Prozent geschafft. Die Werte verharren also im Bereich der führenden ‚1‘ deutlich länger als in den anderen.

Aber auch für andere Zahlen, wie etwa die Größe der Dateien auf der Festplatte eines Computers, die Einwohnerzahlen deutscher Städte oder die Geldbeträge im Rahmen des Rechnungs- bzw. Steuerwesens folgen approximativ der Benford-Verteilung.

Noch skeptisch?

Dann denken Sie sich irgendeine – etwa vierstellige – Zahl und schreiben eine ‚1‘ davor. Nach dieser (dann fünfstelligen) Zahl lassen Sie bei Google suchen. Machen Sie das auch mit einer ‚2‘ am Anfang usw. Sie werden Häufigkeiten erhalten, die in etwa denen des nachfolgend dargestellten Experiments ähneln.



Die Anzahl der Treffer nimmt immer weiter ab, das sollte auch die letzten Zweifler überzeugen.

Die praktische Bedeutung und Anwendung

In der Wirtschaft

Zur Aufdeckung von Betrug bei der Bilanzerstellung, der Fälschung in Abrechnungen, generell zum raschen Auffinden eklatanter Unregelmäßigkeiten im Rechnungswesen. Mit Hilfe des Benfordschen Gesetzes wurde das bemerkenswert "kreative" Rechnungswesen bei Enron und Worldcom aufgedeckt, durch welches das Management die Anleger um ihre Einlagen betrogen hatte. Heute benutzen Wirtschaftsprüfer und Steuerfahnder Methoden, die auf dem Benfordschen Gesetz beruhen. Diese Methoden stellen einen wichtigen Teil der mathematisch-statistischen Methoden dar, die seit mehreren Jahren zur Aufdeckung von Bilanzfälschung, Steuer- und Investorenbetrug und allgemein Datenbetrug in Verwendung sind.

In der Forschung

Das benfordsche Gesetz ermöglicht die Aufdeckung konsequenter Datenfälschung auch in der Wissenschaft. Es waren schließlich Messwerte aus der Natur, die zum Wissen über die Existenz des benfordschen Gesetzes führten. Dessen ungeachtet ist das benfordsche Gesetz nicht allen Wissenschaftlern bekannt, wie Wissenschaftsskandale mit gewisser Periodizität belegen.

Datenfälschung mit Benford?

Die Erstellung gefälschter Benford-adäquater, selbst großvolumiger, Datensätze ist an sich nicht schwierig; um sie jedoch gegen den Blick des geübten numerisch-statistischen Prüfers zu wappnen, muss man bedenken, dass für alle Daten das benfordsche Gesetz nicht nur für die Anfangsziffer, sondern auch für die Folgeziffern (nach modifizierten Regeln) gelten muss; ferner muss noch einer Reihe anderer stochastischer Gesetzmäßigkeiten Rechnung getragen werden. Trimmt der Datenfälscher die Daten allzu genau auf die theoretische Erwartung hin, besteht Gefahr, dass die Manipulationen eben daran erkannt werden.