Die Eins von Planet Zob

Ein bizarres mathematisches Gesetz über die Häufigkeit von Zahlen hilft, Steuersündern und Spendengaunern auf die Spur zu kommen / Von Jörg Albrecht

Roger Pinkham
Theodore Hill, Georgia Institute of Tech.
vor ca. 4 Jahren

Welche Statistik man auch immer betrachtet, die Eins ist die häufigste Ziffer, fast siebenmal so häufig wie die Neun

Als Sir Arthur Conan Doyle auf dem Höhepunkt des Erfolgs stand, entschloss er sich zum Mord. Er hatte es satt, Detektivgeschichten über Sherlock Holmes zu schreiben. So trat dessen Gegenspieler auf den Plan, Professor Moriarty, der Napoleon des Verbrechens. Beim Showdown in den Alpen stürzten beide in den Abgrund. Der Fall wäre erledigt gewesen, hätten die Leser nicht laut protestiert und den Autor zur Wiedererweckung seines Helden gezwungen.

Für den Schurken hatte Conan Doyle ein lebendes Vorbild: Moriartys fiktive Karriere erinnert streckenweise an den – hoch angesehenen und kreuzbraven – amerikanischen Mathematiker und Astronomen Simon Newcomb. 1835 in Nova Scotia geboren, wuchs Newcomb zur wissenschaftlichen Autorität heran. Seine Berechnungen der Planetenbahnen galten als Standard, kleinste Abweichungen, etwa beim Merkur, ließen sich später nur durch Einsteins Relativitätstheorie erklären. Newcomb pries unermüdlich den Wert der wissenschaftlichen Methode; auch im politischen, ökonomischen und sozialen Alltag sah er Gesetzmäßigkeiten am Werk. Seine sonderbarste Entdeckung aber machte Newcomb in Logarithmentafeln.

Logarithmentafeln wurden früher für Kalkulationen aller Art benutzt. Nach Jahren des Gebrauchs waren sie entsprechend zerfleddert. Aber, wie Newcomb fand, keineswegs durch und durch: Die vorderen Seiten schienen schmutziger als die hinteren, als hätten sich die Benutzer bevorzugt für die niedrigen Zahlen interessiert. Newcomb, der die Welt für prinzipiell berechenbar hielt, zog daraus einen weit reichenden Schluss: Die Ziffer 1 am Beginn einer Zahl müsse häufiger vorkommen als die 2, die 2 wiederum häufiger als die 3 und so fort, am seltensten beginne eine Zahl mit einer 9. Das klang absurd. Warum sollte die Natur kleine Zahlen bevorzugen? Schließlich nutzen Ingenieure und Wissenschaftler alle erdenklichen Messgrößen. Aber Newcomb ging noch weiter, er stellte ein Gesetz auf: Die Wahrscheinlichkeit p(d), dass eine beliebige Zahl mit einer bestimmten Ziffer d beginne, sei exakt errechenbar (nach der Formel $p(d) = \log(1+1/d)$). Mit einer Logarithmentafel oder einem Taschenrechner lässt sich einfach prüfen: Die Wahrscheinlichkeit, dass irgendeine Zahl mit der Ziffer d = 1 beginnt, beträgt rund 0,3, was 30 Prozent entspricht. Die Ziffer 2 folgt mit 17, die 3 mit 12 Prozent, und die Ziffer 9 muss sich gar mit 4,5 Prozent begnügen.

1881 veröffentlichte Simon Newcomb seine Berechnungen im *American Journal of Mathematics* – und erntete Schweigen. Sei es, dass die Arbeit niemand las, sei es, dass niemand ihre Bedeutung erkannte, Newcombs merkwürdiges Gesetz geriet in Vergessenheit. Erst ein halbes Jahrhundert später wurde es erneut entdeckt; der Physiker Frank Benford stolperte in den Diensten der General Electric Company ebenfalls über abgegriffene Logarithmen. Nach ihm heißt die Formel heute Benfords Gesetz. Benford ließ es nicht bei Logarithmen, er stürzte sich auf alles, was die Statistik hergab: Luftdruckmessungen, die Ergebnisse der amerikanischen Baseball-Liga, Atomgewichte, Bevölkerungszahlen, die Höhe der Stromrechnungen auf den Solomon Islands, Artikel aus dem *Reader's Digest*, insgesamt über 20 000 Einzelbeobachtungen, die alle dasselbe ergaben. Die 1 lag immer vorn.

Warum das so ist, leuchtet auf den ersten Blick nicht ein. Selbst für Statistiker ist Benfords Gesetz eine harte Nuss. So häufig es gilt, so häufig gilt es scheinbar auch wieder nicht. Der Preis aller Biermarken etwa, die im Supermarkt in der Dose angeboten werden, bewegt sich aus Konkurrenzgründen innerhalb einer bestimmten Spanne; durchaus möglich, dass sie alle zwischen 90 und 99 Pfennige kosten. Die Temperaturschwankungen zwischen Tag und Nacht folgen einer Wellenlinie, keineswegs Benfords Gesetz. Andererseits gibt es Verteilungen, die vollständig dem Zufall unterworfen sind – Lotteriezahlen sind ein bekanntes Beispiel, ihre Ziehung lässt sich nicht vorhersagen. Zwischen schierem Zufall und strikter Notwendigkeit liegt ein dritter Bereich, der sich mit der Gaußschen Glockenkurve beschreiben lässt. Unter dieser so genannten Normalverteilung streuen die Werte um eine Mittelachse, Abweichungen nach oben oder unten sind entsprechend seltener, was ebenfalls nicht der Benfordschen Häufigkeit entspricht.

Wie lässt sich Benfords Gesetz einordnen? Welche Daten folgen der Logarithmusformel, welche nicht? Lange Zeit stocherten die Zahlenforscher im Nebel. Das alles schien obskur, eine Laune der Natur, die sich dem mathematischen Verständnis entzog. Seriös wurde die Sache erst, als Roger Pinkham von der Rutgers University in New Brunswick zeigen konnte, dass Benfords Gesetz sogar universelle Gültigkeit besitzt. Er demonstrierte das am Beispiel der geografischen Fläche verschiedener Flüsse, die alle der Benfordschen Regel gehorchen, und zwar egal, ob man die Fläche nun in Quadratkilometern, Hektar, Morgen oder Ruten bemisst; selbst das fiktive Flächenmaß Zinkolis vom Planeten Zob, erläuterte Pinkham, führe zum selben Ergebnis. Zwar würden sich die einzelnen Ziffern bei der Umrechnung ändern, doch an der Gesamtverteilung ändere das nichts – sie ist, wie der Mathematiker sagen würde, skaleninvariant. Benfords Gesetz ist, wie Pinkham außerdem bewies, zugleich das einzig mögliche Gesetz über Ziffernhäufigkeiten, das diese Bedingung erfüllt.

Wie der Geist über dem Wasser schwebt Benfords Gesetz über dem Meer der Statistik. Letzte Zweifel konnte vor vier Jahren der Mathematiker Theodore Hill vom Georgia Institute of Technology ausräumen. Benfords Gesetz, fand er heraus, ist gewissermaßen die Mutter aller Verteilungshäufigkeiten. Hill hat das am Beispiel der Titelseite einer Tageszeitung erläutert: Eine Meldung betrifft die aktuellen Gewinnzahlen, eine zweite die Entwicklung des Benzinpreises, eine dritte das Wetter, eine vierte den Haushalt des Finanzministers. Jede Angabe für sich beruht auf einer anders gearteten Statistik – aber alle zusammen folgen Benfords Gesetz. Wer mag, kann das schnell überprüfen: Knapp jede dritte Zahlenangabe einer Titelseite sollte mit der Ziffer 1 beginnen, knapp jede fünfte mit der 2 und so weiter. Wo das nicht so ist, darf man misstrauisch werden.

Denn mithilfe der Benfordschen Regel lässt sich zwar kein Blumentopf gewinnen, erst recht kein Sechser im Lotto. Nützlich ist sie aber da, wo Menschen Zahlen frisieren. Also bei den Spesen etwas schummeln, bei der Steuer korrigieren oder bei der Spende ein paar Mark in die eigene Tasche stecken. Um solchen Praktiken auf die Schliche zu kommen, hat Mark Nigrini, Professor für Buchhaltungswesen an der Southern Methodist University in Texas, ein Programm entwickelt namens *Digital Analyzer*. Spaßeshalber überprüfte er damit die Steuererklärungen der Clintons. Ergebnis: Der Präsident habe über 13 Jahre halbwegs ehrliche Angaben gemacht und wohl nur im Detail ein bisschen nachgebessert. Vor kurzem hat Mark Nigrini ein Buch veröffentlicht (*Digital Analysis Using Benford's Law*), in dem er seine Methode und ein paar ernste Fälle vorstellt. Im Auftrag einer Hotelkette kam er einem millionenschweren Versicherungsbetrug auf die Spur. Eine Angestellte hatte Schecks der firmeneigenen Krankenversicherung gefälscht und Herzoperationen abgerechnet, die nicht stattgefunden hatten, zu jeweils 6500 Dollar. Die ersten beiden Ziffern 6 und 5 stachen aus Nigrinis Analyse hervor wie ein Daumenabdruck.

Zahlenfälscher kennen das Gesetz nicht – und fliegen auf

„Ein Werkzeug, das eines Sherlock Holmes würdig wäre", lobte der Finanzfahnder der Staatsanwaltschaft von Brooklyn, New York. Als die Zeitschrift *New Scientist* davon berichtete, gingen Berge von Leserbriefen ein. Nigrini erhält inzwischen Anfragen aus aller Welt. Australische Zollbeamte wollen wissen, wie man die Plausibilität von Zollerklärungen prüfen kann, ein Mathematiker aus der Ukraine hoffte, systematischen Urnenschwindel bei den jüngsten ukrainischen Parlamentswahlen aufzudecken.

Zu Nigrinis prominentesten Kunden gehört das internationale Wirtschaftsprüfungsunternehmen Ernst & Young. Weltweit setzt es den *Digital Analyzer* auf rund 4000 Rechnern ein, gewissermaßen als Krücke. Im Fall des Verdachts einer Unterschlagung kann Benfords angewandtes Gesetz erste Hinweise liefern; die Psychologie des gewöhnlichen Betrügers, sagt Mark Nigrini, führt im Allgemeinen zu auffälligen Zahlenmustern, die von der erwarteten Regel abweichen. Auch wenn Genehmigungsgrenzen systematisch umgangen werden, fällt das auf.

Hätte man mithilfe von Benfords Gesetz auch Licht in den aktuellen Parteispendenskandal bringen können? Bei Ernst & Young Deutschland, die den Rechenschaftsbericht der CDU unter die Lupe genommen haben, löst die Frage Nachdenken aus. Nein, vermutet Andreas Nutz von der Grundsatzabteilung Prüfungstechnik. Die Schwierigkeit liege in der Skaleninvarianz des Gesetzes. Eine Spende von 100 000 oder gar 1 000 000 Mark wäre in der Flut kleinerer Beträge von 10 beziehungsweise 100 oder 1000 Mark einfach untergegangen. Was man allerdings hätte finden können, seien ungewöhnlich häufige Spenden im Bereich von knapp unter 20 000 Mark.

Dort setzt das deutsche Parteiengesetz die Publizitätsgrenze: Oberhalb von 20 000 Mark sind Namen und Anschrift des Spenders im Rechenschaftsbericht zu nennen, unterhalb dieser Grenze kann er anonym bleiben. „Ich bedaure es fast", sagt Nutz, „dass wir das nicht geprüft haben."