45 Das Benfordsche Gesetz – Wie Mathematik Zahlenfälscher entlarvt

Ein einfaches mathematisches Gesetz hilft Fälschungen in verschiedenen Bereichen aufzudecken.

Die Geschichte von Benfords Gesetz beginnt mit einem Artikel von Frank Albert Benford[37] aus dem Jahre 1938 und zeigt wie auch theoretische Ergebnisse konkrete praktische Anwendungen, wie etwa Aufdeckung von Steuerbetrug, haben können.

Frank Benford interessierte sich für Muster in Zahlendaten und beobachtete eine Ziffernverzerrung bei der Verteilung der ersten oder der führenden Ziffern und wollte dies mathematisch erklären. Gerade in einem Zeitalter, in dem wir ständig mit riesigen Datenmengen bombardiert werden, ist diese Entdeckung sehr wichtig. Man nennt man diese Gesetzmäßigkeit auch „Newcomb-Benford Gesetz" (NBG), da es genaugenommen bereits 1881 vom Mathematiker Simon Newcomb entdeckt wurde[38] nachdem er bemerkt hatte, dass die in Logarithmentabellen diejenigen Seiten mit der Eins als erster Ziffer schmutziger als andere Seiten waren, also häufiger benutzt worden waren. Und so schrieb Benford in seinem Artikel:

> *There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.*
>
> Frank Benford, 1938

Bevor wir das Gesetz beschreiben können, müssen wir zunächst die Notation festlegen. Jede positive Zahl x kann im Dezimalsystem als $M(x) \cdot 10^E$ geschrieben werden, wobei $M(x) \in [1, 10)$ die sogenannte Mantisse und E der Exponent ist. Der ganzzahlige Anteil der Mantisse (die Zahl vor dem Komma) wird als führende Stelle oder erste Stelle bezeichnet. Die Zahl 14568,79 wird in wissenschaftlicher Notation als $1.456879 \cdot 10^4$ geschrieben. Die Mantisse ist 1,456879, der Exponent ist 4, die führende Ziffer ist 1.

Die naheliegendste Vermutung wäre, zu behaupten, dass bei einem allgemeinen Datensatz alle Zahlen mit gleicher Wahrscheinlichkeit die führende Ziffer sind. Wir würden dann also annehmen, dass wir in etwa 11 % der Fälle eine führende Ziffer von 1,2,...,9 beobachten sollten. Hierbei sind wir davon ausgegangen, dass jede Zahl 1/9 der Zeit und nicht 1/10 der Zeit vorkommt, da 0 nur die führende Ziffer für die Zahl 0 ist.). Der Inhalt des Benfordschen Gesetzes ist, dass dies häufig nicht der Fall ist; insbesondere erwarten wir in vielen Situationen, dass die *führende Ziffer* d mit einer Wahrscheinlichkeit von etwa $P(d) = \log_{10}\left(\frac{d+1}{d}\right)$ bzw. $\log_{10}(d+1) - \log_{10}(d)$ ist, was bedeutet, dass

[37] Frank Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society, 78, No. 4 (1938), 551–572.

[38] Simon Newcomb, *Note on the frequency of use of the different digits in natural numbers*, American Journal of Mathematics, 4 (1881), 39–40.

die Wahrscheinlichkeit einer ersten Ziffer von 1 etwa $\log_{10} 2 \approx 30,1\,\%$ beträgt, während eine erste Ziffer von 9 etwa $\log_{10}\left(\frac{10}{9}\right) \approx 4,58\,\%$ der Zeit vorkommt.

Eine Menge von Zahlen erfüllt das Benfordsche Gesetz für die führende Ziffer, wenn die Wahrscheinlichkeit eine erste Ziffer von d zu beobachten ungefähr gleich $P(d) = \log_{10}\left(\frac{d+1}{d}\right)$ ist. Schließlich könnten wir, anstatt nur die führende Ziffer zu untersuchen, die gesamte Mantisse untersuchen. Anstatt also nach der Wahrscheinlichkeit für eine erste Ziffer von 1 oder 2 oder 3 zu fragen, fragen wir nun nach der Wahrscheinlichkeit für die Beobachtung einer Mantisse z.B. zwischen 1 und 3. Diese Verallgemeinerung wird häufig als das Starke Benfordsche Gesetz bezeichnet: Wir sagen, dass ein Datensatz das Starke Benfordsche Gesetz erfüllt, wenn die Wahrscheinlichkeit von einer Mantisse in $[1, m)$ zu beobachten, $\log_{10} m$ ist. Beachten Sie, dass das starke Benfordsche Gesetz das Benfordsche Gesetz impliziert: die Wahrscheinlichkeit einer ersten Stelle von d ist einfach die Wahrscheinlichkeit, dass die Mantisse in $[d, d+1)$ liegt, also $\log_{10}(d+1) - \log_{10}(d)$.

Das Benfordsche Gesetz kann man vielfältig einsetzen, etwa von Wirtschaftsprüfern bei der Aufdeckung von Steuerbetrug, Fälschung von Bilanzen oder generell bei dem Erkennen von Unregelmäßigkeiten im Rechnungswesen. Ebenso kann es in der Wissenschaft dazu dienen Datenfälschungen aufzudecken. Bekannt geworden sind auch Untersuchungen von signifikanten Unregelmäßigkeiten bei Wahlen, z.B. bei den Präsidentschaftswahlen 2009 im Iran.