DAS GESETZ DER ABNORMALEN ZAHL

von **Walter Krämer**, Dortmund

1. Einleitung und Zusammenfassung

Nichts fördert so sehr das Interesse an einer Sache wie ein paradoxes Resultat. Der vorliegende Beitrag diskutiert ein solches auf den ersten Blick verblüffendes stochastisches Phänomen. In Anlehnung an F. Benford (1938) nenne ich es einmal "Das Gesetz der abnormalen Zahl". Grob gesagt geht es dabei um folgendes:

Wir beobachten eine positive Zufallsvariable mit großem Wertebereich (etwa in die Milliarden hinein). Beispiel: Wir wählen aus der Tageszeitung eine Zahl zufällig aus (etwa die zweite positive Zahl in der dritten Spalte auf Seite 4). Mit welcher Wahrscheinlichkeit ist deren erste Ziffer eine 1?

Spontan antworten viele hier mit 1/9, mit dem folgenden Argument: Für das Favorisieren einer bestimmten Anfangsziffer gibt es keinen Grund - eine ist so wahrscheinlich wie die andere. Bei 9 möglichen Anfangsziffern 1,2,...,9 muß daher die Wahrscheinlichkeit für jede gleich 1/9 sein.

Dieser Logik widerspricht die Empirie jedoch ganz eklatant. Der nächste Abschnitt referiert ein Experiment dazu, das sich auch im schulischen Statistikunterricht leicht wiederholen läßt. Eine theoretische Begründung des Ergebnisses folgt in Abschnitt 3.

2. Zahlen in Zeitungen

Die folgende Tabelle zeigt die Häufigkeit der Anfangsziffern 1,2,...,9 unter 954 zufällig ausgewählten Zahlen aus englischen Tageszeitungen (die Times vom 7. und 16. Juli 1990 und der Daily Telegraph vom 9. Juli 1990). Neben den empirischen sind auch schon die im dritten Abschnitt noch abzuleitenden theoretischen Werte eingefügt:

Tabelle 1: Ziffernverteilung

Ziffer	1	2	3	4	5	6	7	8	9	Σ
empir. Häufigk.	301	157	102	93	84	77	46	49	45	954
theor. Häufigk.	287	168	119	92	76	64	55	49	44	954

Wir sehen, kleine Anfangsziffern kommen weit häufiger als große vor, die 1 etwa achtmal so häufig wie die 9. Außerdem korrespondieren die empirischen Häufigkeiten gut mit dem, was laut Theorie zu erwarten war: Der X^2-Anpassungstest liefert als Prüfgröße

$$\Sigma_{i=1}^{9}(\text{empir. - theor. Häufigk.})^2/\text{theor. Häufigk.} = 6.95,$$

weit weniger als die kritische Grenze (bei einem Signifikanzniveau von 5 Prozent und einer X^2-Null-Verteilung mit 8 Freiheitsgraden) von 15.51.

Die obige Tabelle berücksichtigt sowohl den redaktionellen wie den Werbeteil. Pro zufällig ausgewählter Seite gingen *alle* auf dieser Seite in welchem Kontext auch immer abgedruckten Zahlen in die Untersuchung ein. Eine Anzeige für den neuen Mazda 626, mit drei Jahren Garantie und Preis 10549 Pfund, lieferte also die Anfangsziffern 6, 3 und 1. Die 2.5 Milliarden Pfund Regierungsunterstützung für die englische Kommunalverwaltung brachte eine 2, ein 45-jähriger knapp einem Bombenattentat entgangener Taxifahrer eine 4, die bei der Tochter Breschnews beschlagnahmten 65000 Rubel eine 6 und so fort. Unberücksichtigt blieben allein die Kopfleisten der Seiten mit Datum und Seitenzahl, da diese Zahlen nur mit großer Mühe als "zufällig" zu interpretieren sind. Zugleich wird so der Einwand ausgeräumt, die große Häufigkeit der "1" beruhe nur auf der Dominanz der 1 als erster Ziffer bei Jahreszahlen. Bei Ermüdung meiner Auszähl-Helfer wurde das Experiment beendet.

Mögliche Stolpersteine bei einer Wiederholung dieses Experimentes im Schulunterricht sind Telefonnummern mit einer 0 am Anfang (hier wurde die erste von 0 verschiedene Ziffer als erste Ziffer übernommen) oder Sportergebnisse. Hier wurde etwa bei einem Tennisresultat von 6:4 nur eine 6 vermerkt. Auch die Auswahl der Zahlen selbst kann natürlich nach einem anderen als dem obigen Schema geschehen.

3. Der mathematische Hintergrund

Die zunächst so verblüffende Tabelle 1 geht auf ein allgemeines Resultat zu stetigen Zufallsvariablen modulo 1 zurück. Damit ist in normalem Deutsch einfach der Nachkommateil des zufälligen Wertes gemeint. Hat etwa die Zufallsvariable X den Wert 2.75, so hat X modulo 1 (im weiteren ^{o}X)

den Wert 0.75. Derart reduzierte Zufallsvariablen haben als Wertebereich also das halboffene Intervall [0,1).

Lemma (siehe Feller, 1971, S. 62):

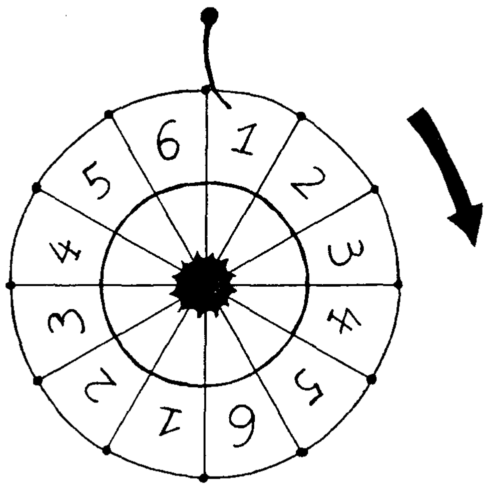
Hat eine stetige Zufallsvariable X die Dichte f(x), so hat die reduzierte Fassung ^{o}X von X die Dichte

$$g(x) = \Sigma_{n=-\infty}^{\infty} f(x+n).$$

Der Punkt ist nun: Für viele stetige Zufallsvariablen X mit großem Wertebereich (etwa solche mit unimodaler Dichte und großem Erwartungswert wie großer Varianz) ist die Funktion g unabhängig von der Gestalt der Ausgangsdichte f nahezu konstant, d.h. die Modulo 1 reduzierte Zufallsvariable ist annähernd gleichverteilt. Der Beweis für den einfachen Fall, daß die Dichte f von X bis zu einer Stelle a monoton wächst, in a ihr (sehr kleines) Maximum erreicht und für x>a monoton fällt, findet sich bei Feller. Jedoch ist dieser Beweis für den Schulunterricht wohl zu kompliziert. Stattdessen kann man das Resultat sehr schön durch ein physikalisches Experiment, etwa das sogenannte "Glücksrad", das man früher oft auf Jahrmärkten sah, veranschaulichen: Dieses Glücksrad ist eine große bunte Scheibe, in verschiedene Sektoren unterteilt, die der Schausteller in Schwung versetzt und deren Scheitelpunkt bei Stillstand den Gewinn bestimmt.

Schaubild 1: Glücksrad

Die Ziffern 1 - 6 sind annähernd gleichwahrscheinlich.
Die Approximation der Gleichverteilung ist umso besser, je öfter sich das Glücksrad dreht.



Denken wir uns nun den Umfang des Glücksrades von der Länge 1, so ist der insgesamt vom ursprünglichen Scheitelpunkt pro Versuch zurückgelegte Weg eine Zufallsvariable X, und die Entfernung bei Stillstand des ursprünglichen Scheitelpunktes vom Zenit bei Stillstand der Scheibe deren reduzierte Version ^{o}X. Nun ist X sicher nicht gleichverteilt, wohl aber, zumindest annähernd, ^{o}X. Jedenfalls gehen sowohl Jahrmarktsgäste wie Schausteller ohne nachzudenken immer davon aus. Alle Punkte des Scheibenrandes liegen bei Stillstand mit gleicher Wahrscheinlichkeit obenauf, oder anders ausgedrückt: die Endposition eines gegebenen Randpunktes ist gleichverteilt.

Genauso zeigt auch das Glücksspiel "Roulette" das obige Lemma in Aktion. Im Unterschied zum Jahrmarkts-Glücksrad dreht sich hier jedoch der Feststellfinger, wenn auch in entgegengesetzter Richtung, sozusagen mit. Mit anderen Worten, die zwischen diesem und dem ursprünglichen Scheitelpunkt erzeugte Distanz vergrößert sich ganz instinktiv, und ohne das obige Lemma zu kennen, haben also die Erfinder des Roulettes erkannt, daß mit steigendem Wertebereich von X die Approximation der Verteilung von ^{o}X durch die Gleichverteilung immer besser wird.

Der Zusammenhang mit unserem Ausgangsproblem ist nun wie folgt:

Sei X eine zufällig aus der Zeitung ausgewählte Zahl. Die erste Ziffer von X ist k genau dann, wenn

$$k*10^{n} \leq X < (k+1)10^{n}$$

für ein geeignetes n. Diese Ungleichung ist aber äquivalent zu

$$n + \log(k) \leq \log X < n + \log(k+1),$$

mit "log" gleich Logarithmus zur Basis 10. Aus dieser Fassung sehen wir jedoch, daß die erste Ziffer von X gleich k ist genau dann, wenn für die reduzierte Fassung ^{o}Z von Z=logX gilt:

$$\log(k) \leq {}^{o}Z \leq \log(k+1). \qquad (1)$$

Die Zufallsvariable ^{o}Z ist aber im Lichte unseres Lemmas annähernd gleichverteilt. Daher beträgt die Wahrscheinlichkeit für das Ereignis (1) also log(k+1)-log(k), d.h.:

$$
\begin{aligned}
P(k=1) &= \log(2) - \log(1) = 0.301,\\
P(k=2) &= \log(3) - \log(2) = 0.176,\\
P(k=3) &= \log(4) - \log(3) = 0.125,\\
P(k=4) &= \log(5) - \log(4) = 0.097,\\
P(k=5) &= \log(6) - \log(5) = 0.079,\\
P(k=6) &= \log(7) - \log(6) = 0.067,\\
P(k=7) &= \log(8) - \log(7) = 0.058,\\
P(k=8) &= \log(9) - \log(8) = 0.051,\\
P(k=9) &= \log(10) - \log(9) = 0.046.
\end{aligned}
$$

Durch Multiplikation dieser Wahrscheinlichkeiten mit 954 erhalten wir dann (bis auf die Rundungsfehler) die letzte Zeile von Tabelle 1. Wie wir dort sehen, stimmen diese theoretischen Werte verblüffend genau mit den empirischen Beobachtungen überein.

LITERATUR

BENFORD, F. (1938): *The law of anomalous numbers*, Proceedings of Philosophical Society 78, 551-572.

FELLER, W. (1971): *An Introduction to the Probability Theory and its Applications*, Band 2, New York (John Wiley).

Ich danke zwei Gutachtern für zahlreiche nützliche Hinweise sowie Denis und Eva Krämer und Florian und Sebastian Eberstaller (Alter 7-13), meinen Auszählhelfern, für ihren Eifer im empirischen Vorfeld dieses Manuskripts.