Gleichverteilung auf dem Kreis

Teilnehmer:

Thoralf Dietrich	Heinrich-Hertz-Oberschule, Berlin
Duc Son Le	Immanuel-Kant-Oberschule, Berlin
Markus Penner	Goethe-Oberschule, Berlin
Alexander Schirmer	Herder-Oberschule, Berlin
Tim Venzke	Immanuel-Kant-Oberschule, Berlin

Gruppenleiter:

Elke Warmuth — Humboldt-Universität zu Berlin
Mitglied im DFG-Forschungszentrum Matheon
„Mathematik für Schlüsseltechnologien“

Was bedeutet es, dass eine Zahlenfolge gleichverteilt mod 1 ist? Wo findet man die Gleichverteilung wieder? Was sind Benford-Folgen? Ist die 1 die bevorzugte führende Ziffer bei empirischen Daten? Wenn ja, würde man diese Daten bei auffallender Abweichung von der Benford-Verteilung anzweifeln und überprüfen müssen? Nach diesem Prinzip könnten Steuerangaben nach ihrem Wahrheitsgrad beurteilt werden. Das und auch die Frage des Nachweises dieser Eigenschaften sollen in der folgenden Arbeit genauer beantwortet werden.

1 Die Definition der Gleichverteilung

Wir führen zunächst den Begriff des gebrochenen Anteils einer reellen Zahl ein, der im Folgenden wichtig sein wird.
Für eine reelle Zahl x bezeichnet $[x]$ die größte ganze Zahl, die kleiner oder gleich x ist. Die Differenz aus dem ganzen Anteil und der Zahl ergibt den *gebrochenen Anteil* $\{x\}$.

$$x = [x] + \{x\}$$

$$\text{Beispiel: } -3,14 = [-3,14] + \{-3,14\} = -4 + 0,86$$

Wir beschäftigen uns mit Folgen reeller Zahlen (a_n). In der folgenden Definition zählt $K_{\alpha,\beta}(n)$ die Anzahl der ins Intervall $[\alpha, \beta)$ fallenden gebrochenen Anteile unter den ersten n Folgengliedern.

Definition 1.1. Eine Folge reeller Zahlen (a_n) heißt gleichverteilt mod 1 (u.d. mod 1), falls für alle $\alpha, \beta \in (0,1), \alpha < \beta$ gilt:

$$\lim_{n\to\infty} \frac{K_{\alpha,\beta}(n)}{n} = \beta - \alpha \tag{1.1}$$

Die Definition besagt, dass für eine Folge (a_n), die u.d. mod 1 ist, für beliebige Intervalle derselben Länge der Anteil der ins Intervall fallenden gebrochenen Anteile der Folgenglieder gleich ist, wenn die Anzahl der betrachteten Anfangsfolgenglieder gegen Unendlich strebt. Das muss auch bei beliebiger Intervallgröße gelten. Geometrisch betrachtet bedeutet dies, dass man die Folgenglieder auf einen Kreis mit dem Umfang $u = 1$ aufwickelt. Nach dem Aufwickeln der unendlich vielen Folgenglieder ist der Kreisumfang überall gleich dicht bedeckt.

Die in der Definition beschriebene Eigenschaft lässt sich nur schwer nachweisen, da man für jedes α und β zeigen muss, dass der Grenzwert der relativen Häufigkeiten der gebrochenen Anteile gleich $\beta - \alpha$ ist. Deshalb haben wir Kriterien untersucht, mit denen wir das einfacher nachprüfen können. Im Folgenden sollen zwei dieser Kriterien erläutert werden.

2 Die Weyl-Kriterien

Die beiden folgenden Kriterien haben wir uns aus der Arbeit [5] von Hermann Weyl (1885–1955) angeeignet.

Satz 2.1. *Die Folge (x_n) reeller Zahlen ist genau dann gleichverteilt mod 1, wenn für alle reellwertigen stetigen Funktionen f mit $D_f \supseteq [0;1]$ gilt:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(\{x_i\}) = \int_0^1 f(x)dx. \tag{2.1}$$

Beweis. Sei $c_{[a;b)}(x)$ die Funktion, die definiert ist als

$$c_{[a;b)}(x) = \begin{cases} 1 & \text{falls} \quad x \in [a;b) \\ 0 & \text{falls} \quad x \notin [a;b). \end{cases}$$

Dann kann die Gleichung (1.1) geschrieben werden als

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} c_{[a;b)}(\{x_i\}) = \int_0^1 c_{[a;b)}(x)dx \tag{2.2}$$

Für das Folgende geben wir nur eine *Beweisskizze* an, die die wesentliche Beweisidee verdeutlicht. Auf die Ausführung der einzelnen Schritte verzichten wir.
Jede stetige Funktion im Intervall $[0;1]$ kann mit Hilfe von Treppenfunktionen gleichmäßig beliebig gut angenähert werden. Treppenfunktionen kann man als endliche Linearkombinationen von Funktionen des Typs $c_{[a;b)}(x)$ auffassen. Deshalb gilt (2.2) auch für alle Treppen- und somit auch für alle stetigen Funktionen. Damit haben wir bewiesen, dass Satz 2.1 eine notwendige Bedingung liefert.
Um zu sehen, dass die Bedingung auch hinreichend ist, reicht es aus zu zeigen, dass die Gleichung 2.2 gilt. Das erreicht man, weil man eine Funktion vom Typ $c_{[a;b)}$ gleichmäßig beliebig gut durch eine stetige Funktion approximieren kann. □

Satz 2.2. *Die Folge (x_n) ist genau dann gleichverteilt mod 1, wenn*

$$\lim_{n\to\infty} \tfrac{1}{n} \sum_{k=1}^{n} e^{2\pi m i x_k} = 0 \quad \textit{für alle} \quad m \neq 0, m \in \mathbb{Z}. \tag{2.3}$$

Beweis. Sowohl anschaulich als auch rechnerisch ist klar:

$$\int_0^1 e^{2\pi m i x} dx = 0 \text{ für alle } m \neq 0, m \in \mathbb{Z}.$$

Begründung: Anschaulich wird um den Kreis integriert, sodass sich alle Terme wegheben und rechnerisch wird dies mit der EULERschen Formel klar:

$$e^{i\varphi} = \cos(\varphi) + i\sin(\varphi).$$

Damit ist bewiesen, dass Satz 2.2 eine notwendige Bedingung beinhaltet, denn $e^{2\pi mix_k}$ ist eine stetige Funktion und somit folgt nach Satz 2.1 die Gleichung (2.3).

Wir zeigen nun, dass Satz 2.2 auch eine hinreichende Bedingung liefert, indem wir zeigen, dass für jede stetige Funktion die Bedingung aus Satz 2.1 erfüllt ist. Nach einem Satz von Weierstrass kann jede beliebige stetige Funktion mit Periode 1 mit Hilfe von trigonometrischen Polynomen $\Psi(x)$ angenähert werden, wobei $\Psi(x)$ als Linearkombination von Funktionen des Typs $e^{2\pi mix}, m \in \mathbb{Z}$, ausgedrückt werden kann. Es gilt also: Für alle positiven ε existiert ein trigonometrisches Polynomen $\Psi(x)$, so dass

$$\sup_{0\leq x\leq 1} |f(x) - \Psi(x)| \leq \varepsilon. \tag{2.4}$$

Mit Hilfe der Dreiecksungleichung und der Eigenschaft des Betrages im Zusammenhang mit Integralen haben wir:

$$\left| \int_0^1 f(x)dx - \frac{1}{n}\sum_{i=1}^{n} f(\{x_i\}) \right| \leq$$

$$\leq \left| \int_0^1 (f(x) - \Psi(x))dx \right| + \left| \int_0^1 \Psi(x)dx - \frac{1}{n}\sum_{i=1}^{n} \Psi(\{x_i\}) \right| + \left| \frac{1}{n}\sum_{i=1}^{n} (f(\{x_i\}) - \Psi(\{x_i\}) \right|$$

$$\leq \int_0^1 |f(x) - \Psi(x)|\, dx + \left| \int_0^1 \Psi(x)dx - \frac{1}{n}\sum_{i=1}^{n} \Psi(\{x_i\}) \right| + \frac{1}{n}\sum_{i=1}^{n} |f(\{x_i\}) - \Psi(\{x_i\}|\,.$$

Der erste und der dritte Term auf der rechten Seite sind kleiner als ε wegen (2.4), der zweite Term wird für genügend großes n auch kleiner als ε wegen der Voraussetzung von Satz 2.2. □

Wir wollen nun das zweite Weyl-Kriterium anwenden.

Beispiel. Für welche c ist (x_n) mit $x_n = c \cdot n$ gleichverteilt mod 1?

1. Für rationale c lässt sich c als $\frac{p}{q}$ darstellen. Dann ist für $m = q$ beim Weyl-Kriterium der Term $e^{2\pi mix_n}$ immer gleich eins. Somit ist auch

   $$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} e^{2\pi mix_k} = 1 \neq 0.$$

   Damit haben wir gezeigt, dass für rationale Zahlen c die Folge $(c \cdot n)$ nicht gleichverteilt mod 1 ist. Anschaulich entstehen beim Aufwickeln auf den

Kreis maximal q Häufungspunkte. Als Beispiel betrachten wir einen Stammbruch $c = \frac{1}{q}$. Dann werden nur die Werte

$$\frac{0}{q}, \frac{1}{q}, \frac{2}{q}, \frac{3}{q}, \dots, \frac{q}{q} = \frac{0}{q}$$

angenommen.

2. Bleibt noch der Fall für irrationale c. Es gilt für geometrische Reihen folgende Formel, welche mit der Dreiecksungleichung abgeschätzt werden kann:

$$\left| \frac{1}{n} \sum_{k=1}^{n} e^{2\pi imck} \right| = \frac{1}{n} \left| \frac{e^{2\pi im(n+1)c} - 1}{e^{2\pi imc} - 1} \right| \leq \frac{1}{n} \frac{2}{|e^{2\pi imc} - 1|}$$

Da der rechte Term mit wachsendem n gegen Null geht, konvergiert auch das arithmetische Mittel gegen Null. Damit ist das zweite Weyl-Kriterium erfüllt und die Folge (x_n) gleichverteilt mod 1.

3 Benfordsche Folgen

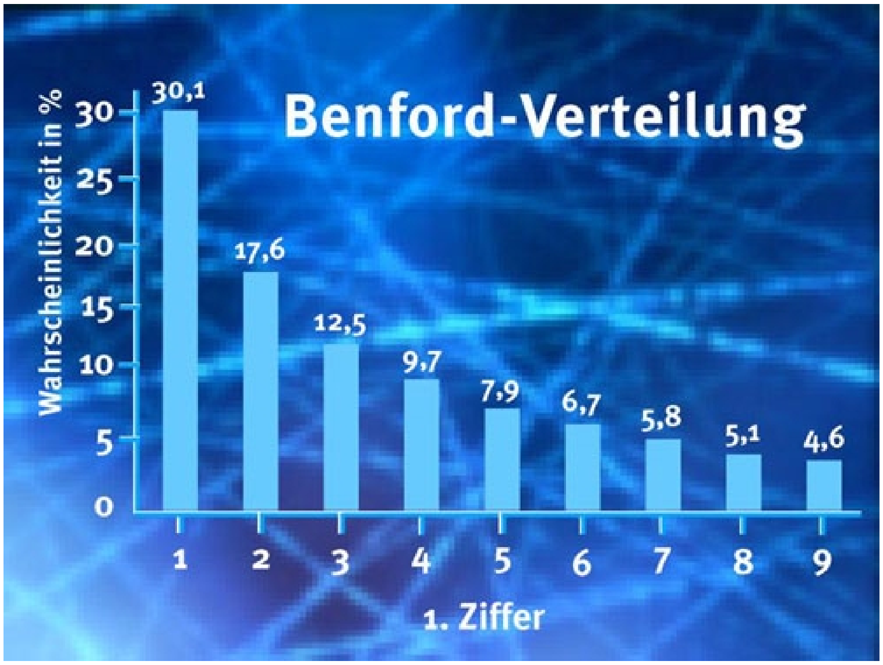
Mit dem Weyl-Kriterium sind wir nun in der Lage nachzuprüfen, wann eine Folge gleichverteilt mod 1 ist. Nun ist es Zeit, mit diesem Wissen die Verbindung zu einem der hauptsächlichen Aspekte unseres Themas herzustellen, dem Benford'schen Gesetz. Dieses macht eine Aussage darüber, dass bei bestimmten Zahlenfolgen die unterschiedlichen Anfangsziffern verschieden häufig auftreten. Diejenigen Zahlenfolgen, die dieser Verteilung der Anfangsziffern genügen, nennen wir *Benford-Folgen*:

Definition 3.1. Eine Folge (a_n) reeller Zahlen heißt *Benford-Folge*, wenn

$$\lim_{n\to\infty} \frac{L_i(n)}{n} = \log\left(1 + \frac{1}{i}\right) \ \forall i = 1, 2, \dots, 9, \tag{3.1}$$

$$\text{wobei } L_i(n) = \#\{a_k | k \leq n \text{ und die erste Ziffer von } a_k \text{ ist } i\}$$

Mit anderen Worten: $L_i(n)$ zählt die Anzahl der Folgenglieder mit führender Ziffer i unter den ersten n Folgengliedern. Die Bedingung der Definition bedeutet somit, dass sich die relativen Häufigkeiten der Anfangsziffern im Grenzwert wie $\log\left(1 + \frac{1}{i}\right)$ verhalten.
Die folgende Grafik stellt die *Benford-Verteilung* dar.



Quelle: www.wdr.de/tv/quarks/sendungsbeitraege/2006/1017/009_zahlen.jsp

Die entscheidende Verbindung zur Gleichverteilung mod 1 wird durch den folgenden Satz hergestellt:

Satz 3.1. *Falls die Folge* $(\log(a_n))$ *gleichverteilt mod 1 ist, so ist die Folge* (a_n) *eine Benford-Folge.*

Beweis. Zum Beweis des Satzes überlegen wir uns zunächst, was es bedeutet, die erste Ziffer i zu besitzen. Letztendlich bedeutet das nichts anderes, als dass die entsprechende Zahl zwischen dem i-fachen und dem $i+1$-fachen einer Zehnerpotenz liegt. In Formeln gesprochen, a_n hat die erste Ziffer i genau dann, wenn:

$$\begin{array}{llcll} \exists j \in \mathbb{Z} & i \cdot 10^j & \leq & a_n & < & (i+1) \cdot 10^j \\ \Leftrightarrow \exists j \in \mathbb{Z} & j + \log i & \leq & \log a_n & < & j + \log(i+1) \\ \Leftrightarrow & \log i & \leq & \{\log a_n\} & < & \log(i+1) \end{array}$$

Nun kann man die Voraussetzung der Gleichverteilung mod 1 der Folge $(\log(a_n))$ anwenden und schlussfolgern, dass der Grenzwert des Anteils der Folgenglieder von (a_n) im Intervall $[\log i, \log(i+1))$ genau

$$\log(i+1) - \log(i) = \log\left(\frac{i+1}{i}\right) = \log\left(1 + \frac{1}{n}\right)$$

ist. Da wir oben gesehen haben, dass dieser Anteil genau dem Anteil $\frac{L_i(n)}{n}$ in der Definition 3.1. entspricht, haben wir damit nachgewiesen, dass (a_n) eine Benford-Folge ist. $\square$

Um in diesem Satz sogar eine Äquivalenz zu erhalten, müssten wir die Bedingung an Benford-Folgen verstärken, damit wir den Schritt im Beweis auch umkehren können. Hierfür reichen die Bedingungen an die diskreten Intervalle zwischen $\log(i)$ und $\log(i+1)$ nicht aus, man muss stattdessen nicht nur die Bedingung an die Anfangsziffer fordern, sondern dieselbe Bedingung für beliebige Folgen von Anfangsziffern. Folgen, die diese Bedingung erfüllen, heißen starke Benford-Folgen, sollen hier aber nicht genauer betrachtet werden.

Stattdessen wollen wir uns ein praktisches Beispiel einer Benford-Folge ansehen:

Satz 3.2. *Die Fibonacci-Folge* (f_n) *definiert durch die Rekursion* $f_{n+2} = f_{n+1} + f_n$ *für* $n \geq 1$ *sowie die Startwerte* $f_1 = f_2 = 1$ *ist eine Benford-Folge.*

Beweis. Zum Beweis wollen wir zeigen, dass die Folge der Logarithmen der Fibonacci-Zahlen gleichverteilt mod 1 ist, woraus sich mit dem vorigen Satz die Behauptung ergibt. Nun kann man zunächst aus der rekursiven Darstellung eine explizite Formel herleiten. Im Falle der Fibonacci-Folge ergibt sich

$$f_n = A \cdot x^n + B \cdot y^n,$$

wobei $x = \frac{\sqrt{5}+1}{2}$ und $y = \frac{1-\sqrt{5}}{2}$ die Nullstellen des Polynoms $z^2 - z - 1$ und A und B Konstanten sind, die so gewählt sind, dass $Ax + By = 1$ und $A \cdot x^2 + B \cdot y^2 = 1$, also die Anfangsbedingungen erfüllt sind.
Der Trick ist nun, zu erkennen, dass $|y| < 1$ gilt und somit $\lim_{n\to\infty} y^n = 0$ folgt. Somit ist für große n das Folgenglied f_n sehr gut approximiert durch $A \cdot x^n$. Wir benötigen für den weiteren Beweis folgendes Lemma:

Lemma. Wenn die Folge (x_n) gleichverteilt mod 1 ist, und (y_n) eine Folge mit

$$\lim_{n\to\infty} (x_n - y_n) = 0, \tag{3.2}$$

dann ist auch (y_n) gleichverteilt mod 1.

Beweis. Mit der Stetigkeit der Exponentialfunktion folgt aus der Bedingung (3.2) auch

$$\lim_{n\to\infty} (e^{2\pi i x_n} - e^{2\pi i y_n}) = 0. \tag{3.3}$$

Jetzt muss man noch einen anschaulich klaren Sachverhalt benutzen (der hier nicht streng bewiesen werden soll):
Wenn eine Folge gegen einen Grenzwert konvergiert, so konvergiert die Folge der arithmetischen Mittel gegen denselben Grenzwert.
Damit gewinnen wir aus (3.3) auch

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (e^{2\pi i x_k} - e^{2\pi i y_k}) = 0.$$

Da aus dem Weyl-Kriterium für (x_n) bereits

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i x_k} = 0$$

folgt, gilt somit auch

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i y_k} = 0,$$

woraus wiederum mit dem Weyl-Kriterium folgt, dass auch (y_n) gleichverteilt ist. Damit ist das Lemma bewiesen. □

Wir hatten oben bereits

$$\lim_{n\to\infty} (f_n - A \cdot x^n) = \lim_{n\to\infty} (B \cdot y_n) = 0$$

gezeigt. Nun muss man daraus folgern, dass auch

$$\lim_{n\to\infty} (\log(f_n) - \log(A \cdot x^n)) = 0,$$

um eine Aussage über $(\log(f_n))$ treffen zu können. Das geht z.B. so:

$$\frac{A \cdot x^n}{f_n} = 1 - \frac{f_n - A \cdot x^n}{f_n} \to 1 - 0 = 1 \text{ für } n \to \infty.$$

Daraus folgt

$$\lim_{n\to\infty} \log\left(\frac{A \cdot x^n}{f_n}\right) = 0$$

Mit dem 2. Logarithmengesetz erhalten wir die Aussage, die wir oben haben wollten:

$$\lim_{n\to\infty} (\log(A \cdot x^n) - \log(f_n)) = 0.$$

Es ist $\log(A \cdot x^n) = \log A + n \cdot \log(x)$. Da $\log(x)$ irrational ist (was hier nicht gezeigt werden soll), ist $n \cdot \log(x)$ gleichverteilt mod 1. Da ein konstanter Summand $\log A$ an der Gleichverteilung nichts ändert (anschaulich wird der Kreis nur gedreht), folgt mit unserem Lemma, dass $(\log f_n)$ gleichverteilt ist und damit die Behauptung des Satzes. □

Literatur

[1] P. Diaconis, The distribution of leading digits and uniform distribution mod 1, Ann. Prob. **5**, 1977, 72–81.

[2] J. L. Brown and R. L. Duncan, Modulo One Uniform Distribution of Certain Fibonacci-Related Sequences, The Fibonacci Quarterly **10**, 1973, 277–294.

[3] N. Hungerbühler, Benfords Gesetz über führende Ziffern: Wie die Mathematik Steuersündern das Fürchten lehrt, ETH Zürich, EducETH, 2007.

[4] L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences, Wiley, 1971.

[5] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, Math. Ann. **77**, 1916, 313–352.