UNIVERSITÄT LEIPZIG

**FAKULTÄT FÜR MATHEMATIK UND INFORMATIK**

# DIPLOMARBEIT

Titel der Diplomarbeit

## Rundum das Benfordsche Gesetz

Verfasser

Nico Uhlig

angestrebter akademischer Grad

Diplom Mathematiker

Leipzig, Januar 2015

Studiengang: Mathematik
Betreuer: Prof. Dr. Tatjana Eisner

# Danksagung

Zunächst möchte ich mich an dieser Stelle bei all denjenigen bedanken, die mich während der Anfertigung dieser Arbeit unterstützt und motiviert haben.

Ganz besonders gilt dieser Dank Prof. Tatjana Eisner, welche mich während meiner Diplomarbeit betreut und umfangreich unterstützt hat. Ihre Vorlesungen haben mir erst die nötigen Kenntnisse zur Bearbeitung der Diplomarbeit vermittelt. Für Ihre Hinweise und die häufigen Besprechungen bin ich Ihnen sehr dankbar.

Daneben gilt mein Dank Prof. Arno Berger und Prof. Theodore Hill, welche nicht nur wesentliche Beiträge zu der Theorie des Benfordschen Gesetz geleistet haben, sondern mir auch mit Hinweisen in Bezug auf die vorliegende Arbeit weiterhalfen.

Weiterhin möchte ich Prof. Bernd Kirstein danken. Sie haben mir nicht nur mein Wissen in der Wahrscheinlichkeitstheorie und Maßtheorie vermittelt, sondern standen auch mit Rat und Tat für meine Fragen bezüglich der Diplomarbeit zur Seite.

Schließlich bedanke ich mich auch bei Prof. Andreas Lasarow, der bereit war meine Arbeit als Zweitgutachter zu bewerten.

# Einleitung

1881 entdeckte der Astronom Simon Newcomb ($*$12.03.1835; †11.07.1909) bei der Betrachtung der damals üblichen Logarithmentafeln, dass die Seiten, in denen die in den Tabellen auftretenden Zahlen mit der Ziffer 1 anfingen, deutlich verschmutzter und abgegriffener waren als andere Seiten. Er stellte als erster die Vermutung auf, dass die Anfangsziffern der Zahlen gewisser Datensätze *nicht* gleichwahrscheinlich auftreten. Er beobachtete, dass die Wahrscheinlichkeit, dass eine Zahl die (signifikante) Anfangsziffer $d \in \{1, 2, \ldots, 9\}$ besitzt, sich oft relativ gut durch

$$P\left(\text{Anfangsziffer von } Z \text{ ist } d\right) = \log\left(d+1\right) - \log d = \log\left(1 + \frac{1}{d}\right), \qquad \text{für } d = 1, 2, \ldots, 9$$

beschreiben lässt. Mit $\log\left(\cdot\right)$ ist im Weiteren stets der Logarithmus zur Basis 10 gemeint und die „signifikante Anfangsziffer einer Zahl" ist die erste von Null verschiedene Dezimalstelle einer Zahl. Seine Arbeit [16] wurde jedoch lange Zeit nicht beachtet und geriet in Vergessenheit bis der Physiker Frank Benford ($*$29.05.1883; †04.12.1948) im Jahre 1938 Newcombs Arbeit wieder entdeckte und Newcombs Vermutung an 20 Datensätzen testete – siehe [1]. So betrachtete er Atom- und Molekülgewichte, Einwohnerzahlen von Städten, Auflagenzahlen von Zeitungen und sogar mathematische und physikalische Konstanten. Er untersuchte zum

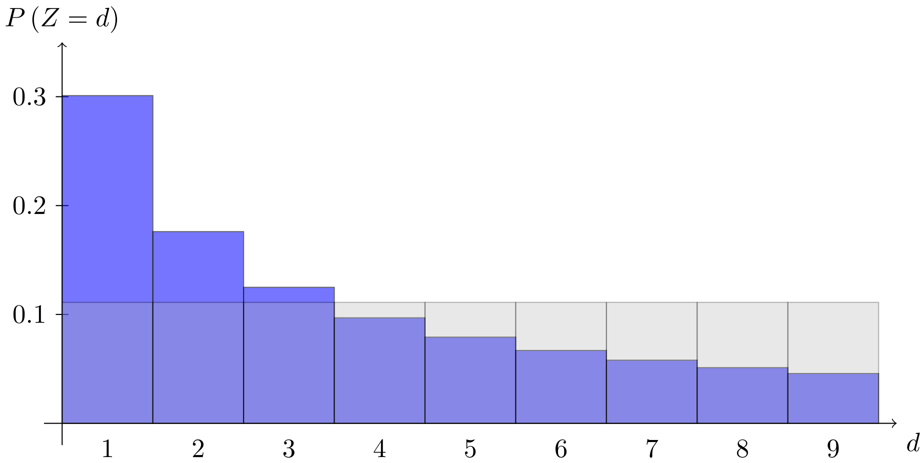


Abbildung 1: Die Verteilung der Anfangsziffern beim Benfordschen Gesetz $P\left(Z = d\right) = \log\left(1 + 1/d\right)$ (blau) in Gegenüberstellung mit der Gleichverteilung der Anfangsziffern $P\left(Z = d\right) = 1/9$ (grau)

Beispiel die Anfangsziffern der Einwohnerzahlen von 3259 amerikanischen Städten und erhielt Abbildung 2. Er stelle insgesamt fest, dass kein einzelner Datensatz genau dem Benfordschen Gesetz genügt. Benford bildete dann das arithmetische Mittel $\frac{1}{20}\sum_{k=1}^{20} P\left(Z_k = d\right)$ mit $d = 1, 2, \ldots, 9$ aller 20 Datensätze und stellte überraschenderweise fest, dass die so gewonnen Anteile am dichtesten am Benfordschen Gesetz lagen – vergleiche hierzu Abbildung 3 und [1].

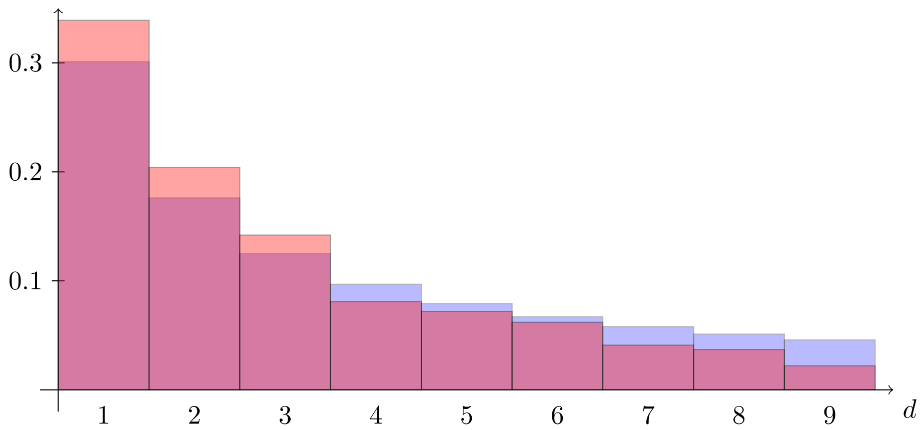


Abbildung 2: Vergleich der Anfangsziffern der Einwohnerzahlen (rot) von 3259 Städten mit dem Benfordschen Gesetz (blau).

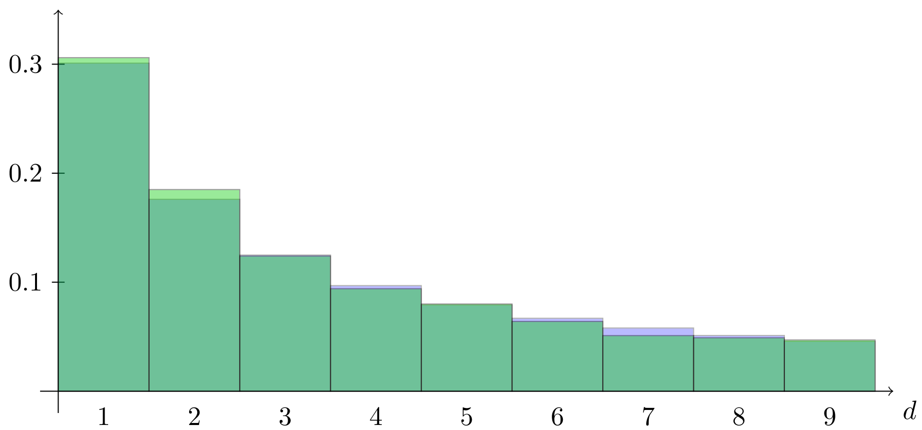


Abbildung 3: Vergleich des arithmetischen Mittels aller 20 Datensätze (grün) mit dem Benfordschen Gesetz (blau).

Nach seinen Untersuchungen stellte Benford 1938 das sogenannte *empirische Benfordsche Gesetz* auf:

**Empirisches Benfordsches Gesetz.** Die Verteilung der signifikanten Anfangsziffern gewisser Datensätze ist *keine* Gleichverteilung, sondern genügt der Gesetzmäßigkeit

$$P\left(\{\text{Anfangsziffer ist } d\}\right) = \log\left(1 + \frac{1}{d}\right), \qquad \text{für } d = 1, 2, \ldots, 9. \tag{1.1}$$

Aufgrund der Vorarbeiten von Newcomb wird es oft auch „*Newcomb-Benford-Law*" genannt. Weder Benford noch Newcomb spezifizierten diese Behauptung oder lieferten einen Beweis. Natürlich ist die Formulierung „gewisse Datensätze" alles andere als mathematisch exakt und es bleibt zu klären für welche Objekte diese Behauptung in welchem Zusammenhang gilt. Es ist zunächst nicht einmal klar, in welchem Sinne die Wahrscheinlichkeit $P$ zu verstehen ist. Die hier verfolgte Vorgehensweise wird sich weitesgehend nach [4] richten.

Ziel dieser Arbeit ist es nun das Benfordsche Gesetz (1.1) zu formalisieren. Zunächst muss der Begriff der signifikanten Dezimalziffer präzisiert werden. Danach wird eine exakte mathematische Formulierung der beobachteten Eigenschaft für verschiedene Objekte erfolgen, um schließlich verschiedenste Kriterien für die Gültigkeit der Gesetzmäßigkeit aufzustellen. Hierbei werden wir vor allem Methoden der Wahrscheinlichkeitstheorie, der Gleichverteilung modulo 1, der Fourier-Analysis, der Spieltheorie und der Ergodentheorie benötigen.

In Kapitel 1 werden wir zunächst die wichtigsten Definitionen für die weitere Vorgehensweise tätigen. Dabei werden wir den Begriff der Signifikanten und der signifikanten $\sigma$-Algebra kennenlernen. Zudem wird bereits die sogenannte Benford-Verteilung eingeführt.

In Kapitel 2 wird zunächst die Definition der Benford-Eigenschaft für Folgen, Funktionen, Zufallsgrößen und Wahrscheinlichkeitsmaße getätigt. Eine erste Charakterisierung der Benford-Eigenschaft mittels der Theorie der Gleichverteilung modulo 1 erfolgt dann in Satz 2.2. Die zweite große Charakterisierung der Benford-Eigenschaft mittels skaleninvarianten signifikanten Dezimalziffern ist dann Satz 2.7. Eine interessante spieltheoretische Charakterisierung der Benford-Verteilung folgt in Satz 2.8. Abgeschlossen wird das Kapitel 2 dann mit Satz 2.9, welcher die Benford-Verteilung mit Basis-invarianten signifikanten Ziffern in Verbindung bringt und vor allem Techniken aus der Fourier-Analysis verwendet.

Kapitel 3 widmet sich dann asymptotischen Betrachtungen zum Benfordschen Gesetz für Folgen von Zufallsgrößen. Dabei interessieren wir uns insbesondere für das Produkt von Zufallsgrößen. Falls die Zufallsgröße $X$ eine Riemann-integrierbare Dichte besitzt, ist Satz 3.2 ein wichtiges asymptotisches Resultat für das Produkt $\prod_{k=1}^{n} X = X^n$ dieser Zufallsgröße. Hauptresultat dieses Kapitels ist dann Satz 3.9, welcher Auskunft über das asymptotische Verhalten des Produkts $\prod_{k=1}^{n} X_k$ einer Folge von unabhängigen Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ mit Blick auf das Benfordsche Gesetz gibt. Hierfür werden vor allem ergodentheoretische Konzepte wie der Birkhoffsche Ergodensatz benötigt.

# Inhaltsverzeichnis

# 1 Einführung

## 1.1 Die Signifikante

In der Formulierung des Benfordschen Gesetzes taucht immer wieder der Begriff „signifikante Anfangsziffer" auf. Dieser Abschnitt soll diesen Begriff formalisieren. Im Weiteren bezeichne stets $\mathbb{R}^+ := \{x \in \mathbb{R} \,|\, x > 0\}$ die Menge der positiven reellen Zahlen. Für $n \in \mathbb{N}$ bezeichne weiterhin $\mathbb{N}_{1,n} := \{1, 2, \ldots, n\}$ die Menge aller natürlichen Zahlen zwischen 1 und $n$ und $\mathbb{Z}_{0,n} := \{0, 1, 2, \ldots, n\}$ die Menge aller ganzen Zahlen zwischen 0 und $n$.

**Definition 1.1.** Für $x \in \mathbb{R}\setminus \{0\}$ heißt die *eindeutig* bestimmte Zahl $k \in \mathbb{N}_{1,9}$ mit

$$10^\ell \cdot k \leq |x| < 10^\ell \cdot (k+1)\,, \qquad \text{für ein gewisses } \ell \in \mathbb{Z},$$

die *erste signifikante Dezimalziffer* der Zahl $x$, welche wir mit $d_1(x)$ bezeichnen. Für $m \in \mathbb{N}$ mit $m > 1$ definieren wir weiterhin induktiv $d_m(x)$ als die *$m$-te signifikante Dezimalziffer* der Zahl $x$, das heißt jene ebenfalls eindeutig bestimmte Zahl $k \in \mathbb{Z}_{0,9}$ mit

$$10^\ell \left[\sum_{j=1}^{m-1} d_j(x)\, 10^{m-j} + k\right] \leq |x| < 10^\ell \left[\sum_{j=1}^{m-1} d_j(x)\, 10^{m-j} + (k+1)\right], \quad \text{für ein } \ell \in \mathbb{Z}.$$

Ferner setzen wir noch $d_m(0) := 0$ für alle $m \in \mathbb{N}$.

*Bemerkung.* Offenbar hat $d_1(x)$ die Eigenschaft, dass für alle $x \in \mathbb{R}^+$

$$d_1(x) = d_1(-x) = d_1(10 \cdot x)\,. \qquad \clubsuit$$

**Beispiel 1.1.** Es ist $d_1(\pi) = 3$, $d_2(\pi) = 1$, $d_1\left(\sqrt{2}\right) = d_1\left(-\sqrt{2}\right) = 1$ und $d_1\left(10 \cdot \sqrt{2}\right) = d_1\left(10^k\sqrt{2}\right) = 1$ für jedes $k \in \mathbb{Z}$.

Oftmals bietet es sich jedoch an, statt den einzelnen signifikanten Dezimalziffern $d_m$ eine andere Funktion, die sogenannte *Signifikante*, zu betrachten:

**Definition 1.2** (Signifikante)**.** Die Funktion $S\colon \mathbb{R} \longrightarrow [1, 10)$ definiert gemäß $S(x) := t$, wobei $t$ die für $x \neq 0$ eindeutig bestimmte reelle Zahl $1 \leq t < 10$, so dass $|x| = 10^k \cdot t$ für ein gewisses $k \in \mathbb{Z}$ ist, heißt *Signifikante*. Wir setzen zudem $S(0) := 0$.

Die Signifikante ist wohldefiniert, da für $x \neq 0$ das $t \in [1, 10)$ und das $k \in \mathbb{Z}$ *eindeutig* bestimmt sind. Angenommen es gäbe $s, t \in [1, 10)$ mit $|x| = 10^k \cdot t = 10^\ell \cdot s$. Ohne Einschränkung sei zudem $\ell \geq k$. Dann wäre

$$0 = |x| - |x| = 10^k \cdot t - 10^\ell \cdot s = 10^k \left(t - 10^{\ell-k} s\right)$$

bzw. $t = 10^{\ell-k}s$. Falls $\ell = k$ ist, gilt offenbar $t = s$. Ist andererseits aber $\ell \neq k$, so wäre für $s \in [1, 10)$

$$t = 10^{\ell-k}s \notin [1, 10)\,,$$

was einen Widerspruch darstellt. Damit sind die Zahlen $k \in \mathbb{Z}$ und $t \in [1, 10)$ aus Definition 1.2 *eindeutig* bestimmt.

**Lemma 1.1.** *Die Signifikante hat folgende Eigenschaften:*

(a) $S(S(x)) = S(x)$ *sowie* $S(x) = S(|x|)$.

(b) *Für jedes* $\ell \in \mathbb{Z}$ *gilt* $S\left(10^{\ell}x\right) = S(x)$.

(c) *Für* $x \neq 0$ *gilt* $S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor}$.

Beweis. Die Behauptungen *(a)* und *(b)* sind nach der Definition der Signifikanten klar. Wir beweisen also nur den Teil *(c)*: Für $x \neq 0$ gilt

$$\begin{aligned} |x| = 10^{\log|x|} &= 10^{\lfloor \log|x| \rfloor - \lfloor \log|x| \rfloor + \log|x|} \\ &= 10^{\lfloor \log|x| \rfloor} \cdot 10^{\log|x| - \lfloor \log|x| \rfloor}. \end{aligned}$$

Da $\lfloor \log|x| \rfloor \in \mathbb{Z}$ und $10^{\log|x| - \lfloor \log|x| \rfloor} \in [1, 10)$, muss nach Definition 1.2 also

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor}$$

sein. □

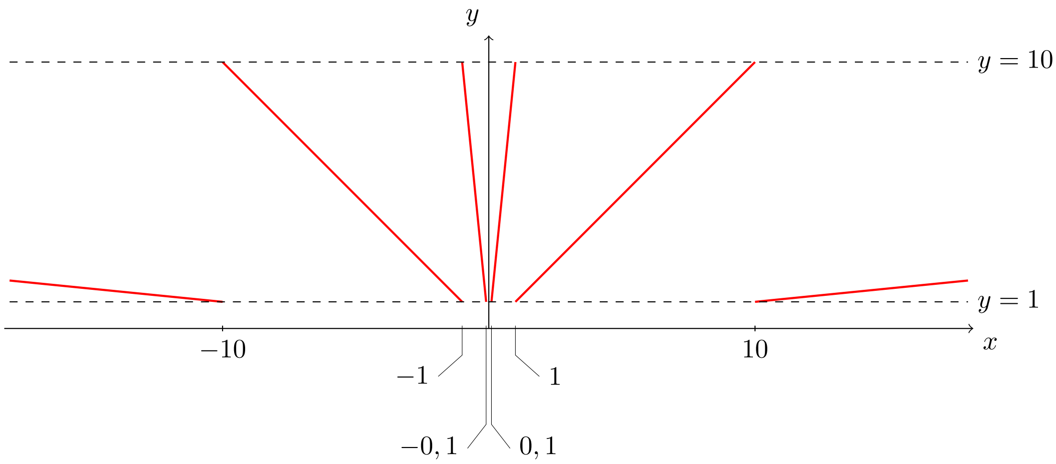


Abbildung 4: Der Graph der Signifikanten $S$: Man erkennt deutlich, die Eigenschaft *(b)* aus Lemma 1.1.

**Beispiel 1.2.** Man erkennt sofort aus der Definition der Signifikanten, dass zum Beispiel $S(\pi) = \pi$ gilt. Nach Lemma 1.1 jedoch können wir auch berechnen

$$S(\pi) = 10^{\log|\pi| - \lfloor \log|\pi| \rfloor} = 10^{\log \pi - 0} = \pi.$$

Genauso erhält man zum Beispiel

$$S(16) = S\left(4^2\right) = 10^{2 \cdot \log 4 - \lfloor 2 \cdot \log 4 \rfloor} = 10^{2 \cdot \log 4 - 1} = \left(10^{\log 4}\right)^2 \cdot 10^{-1} = \frac{4^2}{10} = 1,6.$$

Wir werden nun erkennen, dass die Signifikante $S$ und die signifikanten Dezimalstellen in natürlicher Art und Weise miteinander verbunden sind:

**Satz 1.1.** *Für jedes* $x \in \mathbb{R}^+$ *gilt:*

(a) $S(x) = \sum_{n=1}^{\infty} 10^{1-n} d_n(x)$.

(b) *Für jedes* $m \in \mathbb{N}$ *ist* $d_m(x) = \left\lfloor 10^{m-1} S(x) \right\rfloor - 10 \left\lfloor 10^{m-2} S(x) \right\rfloor$.

Beweis.

**Zu *(a)*:** Beachte zunächst, dass wegen $d_1(x) \in \mathbb{N}_{1,9}$ und $d_n(x) \in \mathbb{Z}_{0,9}$ für $n \geq 2$ dann

$$\sum_{n=1}^{\infty} 10^{1-n} d_n(x) = d_1(x) + \underbrace{\sum_{n=1}^{\infty} 10^{-n} d_{n+1}(x)}_{\in[0,1)} \in [1, 10).$$

Um nun zu zeigen, dass die Signifikante diese Darstellung besitzt, zeigen wir die Eigenschaft aus Definition 1.2

$$x = 10^k \cdot \sum_{n=1}^{\infty} 10^{1-n} d_n(x), \qquad \text{mit einem } k \in \mathbb{Z}. \tag{1.1}$$

Sei zunächst $0 < x < 1$. Dann besitzt $x$ die Dezimaldarstellung $x = \sum_{n=1}^{\infty} 10^{-n} \cdot a_n$ mit $a_n \in \mathbb{Z}_{0,9}$. Da $x \neq 0$ gibt es ein kleinstes $n_0 \in \mathbb{N}$ mit $a_{n_0} =: a \neq 0$. Sei ohne Beschränkung der Allgemeinheit nun $n_0 = 1$, dann ist

$$10^{-1} a \leq \sum_{n=1}^{\infty} 10^{-n} a_n \leq 10^{-1} (a+1),$$

was per Definition 1.1 bedeutet, dass $a = d_1(x)$. Induktiv zeigt man analog, dass

$$a_n = d_n(x), \qquad \text{für alle } n \in \mathbb{N}.$$

Damit ist also

$$x = \sum_{n=1}^{\infty} 10^{-n} a_n = \sum_{n=1}^{\infty} 10^{-n} d_n(x) = \sum_{m=1}^{\infty} 10^{-1+1-n} d_n(x) = 10^{-1} \sum_{n=1}^{\infty} 10^{1-n} d_n(x),$$

was gerade (1.1) mit $k = -1$ für $0 < x < 1$ entspricht. Ist $x \geq 1$, dann gibt es ein (kleinstes) $\ell \in \mathbb{N}$ mit $0 < 10^{-\ell} x < 1$. Aus dem bereits Gezeigten folgt dann die Existenz eines $k \in \mathbb{Z}$ mit $10^{-\ell} x = 10^k \sum_{n=1}^{\infty} 10^{1-n} d_n(x)$, was gerade

$$x = 10^{k+\ell} \sum_{n=1}^{\infty} 10^{1-n} d_n(x),$$

bedeutet und (1.1) entspricht.

**Zu *(b)*:** Ähnlich wie in *(a)* genügt es die Behauptung für $0 < x < 1$ zu zeigen, da sich der Fall $x \geq 1$ auf diesen zurückführen lässt. Sei also $0 < x < 1$. Wir werden die Aussage nun mittels Induktion über $m \in \mathbb{N}$ beweisen. Sei also $m = 1$. Dann ist klar, dass wegen $10^{-1} S(x) \in [0, 1)$ und damit $\lfloor 10^{-1} S(x) \rfloor = 0$ für $m = 1$ gilt

$$d_1(x) = \lfloor S(x) \rfloor = \lfloor 10^{1-1} S(x) \rfloor - 10 \cdot 0 = \lfloor 10^{1-1} S(x) \rfloor - 10 \cdot \lfloor 10^{1-2} S(x) \rfloor.$$

Damit ist die Behauptung für $m = 1$ richtig. Es gibt also ein $m \in \mathbb{N}$, so dass die obige Behauptung für alle $j \in \mathbb{N}_{1,m}$ richtig ist. Sei für $j \in \mathbb{N}$ nun $b_j := \lfloor 10^{j-2} S(x) \rfloor$. Dann gilt nach Induktionsvoraussetzung für $j \in \mathbb{N}_{1,m}$

$$d_j(x) = \lfloor 10^{j-1} S(x) \rfloor - 10 \cdot \lfloor 10^{j-2} S(x) \rfloor = b_{j+1} - 10 \cdot b_j,$$

wobei $b_1 = \lfloor 10^{-1} S(x) \rfloor = 0$ wie gerade gezeigt. Für $m + 1$ ist also

$$\begin{aligned} \sum_{j=1}^{m} d_j(x) 10^{m+1-j} &= 10^{m+1} \sum_{j=1}^{m} [b_{j+1} - 10 \cdot b_j] 10^{-j} \\ &= 10^{m+1} \sum_{j=1}^{m} \left[ 10^{-j} b_{j+1} - 10^{-(j-1)} b_j \right] \\ &= 10^{m+1} \cdot \left( -b_1 + 10^{-m} b_{m+1} \right) = 10 \cdot b_{m+1}. \end{aligned}$$

Sei nun $x = \sum_{n=1}^{\infty} 10^{-n} a_n$ die Dezimaldarstellung von $x$. Da $x \neq 0$ gibt es ein kleinstes $n_0 \in \mathbb{N}$ mit $a_{n_0} =: a \neq 0$. Sei wieder ohne Beschränkung der Allgemeinheit $n_0 = 1$. Wie in *(a)* gilt dann

$$x = 10^{-1} \sum_{j=1}^{\infty} 10^{1-j} d_j(x).$$

Es gilt nun

$$\begin{aligned} & 10^{-(1+m)} \left[ \sum_{j=1}^{m} d_j(x) \, 10^{m+1-j} + (b_{m+2} - 10 b_{m+1}) \right] \\ =& 10^{-(1+m)} \left[ 10 b_{m+1} + b_{m+2} - 10 b_{m+1} \right] \\ =& 10^{-(1+m)} \left\lfloor 10^m S(x) \right\rfloor . \end{aligned}$$

Wegen *(a)* gilt zudem

$$\begin{aligned} \lfloor 10^m S(x) \rfloor &= \left\lfloor 10^m \sum_{j=1}^{\infty} 10^{1-j} d_j(x) \right\rfloor = \left\lfloor \sum_{j=1}^{m+1} 10^{m+1-j} d_j(x) + \sum_{j=m+2}^{\infty} 10^{m+1-j} d_j(x) \right\rfloor \\ &= \sum_{j=1}^{m+1} 10^{m+1-j} d_j(x) = 10^m \sum_{j=1}^{m+1} 10^{1-j} d_j(x) \leq 10^m \underbrace{\sum_{j=1}^{\infty} 10^{1-j} d_j(x)}_{=10x} \\ &= 10^m \cdot 10x = 10^{1+m} x, \end{aligned}$$

das heißt insgesamt

$$\begin{aligned} 10^{-(1+m)} \left[ \sum_{j=1}^{m} d_j(x) \, 10^{m+1-j} + (b_{m+2} - 10 b_{m+1}) \right] &= 10^{-(1+m)} \left\lfloor 10^m S(x) \right\rfloor \\ &\leq 10^{-(1+m)} \cdot 10^{1+m} x = x. \end{aligned}$$

Analog zeigt man noch die andere Ungleichung

$$x \leq 10^{-(1+m)} \left[ \sum_{j=1}^{m} d_j(x) \, 10^{m+1-j} + (b_{m+2} - 10 b_{m+1} + 1) \right].$$

Nach Definition 1.1 gilt dann

$$d_{m+1}(x) = b_{m+2} - 10 b_{m+1} = \left\lfloor 10^{(m+1)-1} S(x) \right\rfloor - 10 \left\lfloor 10^{(m+1)-2} S(x) \right\rfloor ,$$

also ist die Behauptung für $m+1$ gezeigt. □

## 1.2 Die signifikante $\sigma$-Algebra

Im Weiteren bezeichne $\mathfrak{B}$ stets die Borelsche $\sigma$-Algebra auf $\mathbb{R}$ und $\mathfrak{B}I$ die Borelsche $\sigma$-Algebra auf einem Intervall $I \subseteq \mathbb{R}$. Für eine Funktion $f \colon \Omega \longrightarrow \mathbb{R}$ definieren wir

$$\sigma(f) := \left\{ f^{-1}(B) \,\middle|\, B \in \mathfrak{B} \right\}$$

als die von $f$ *erzeugte $\sigma$-Algebra* auf $\Omega$. Sie ist die kleinste $\sigma$-Algebra auf $\Omega$, so dass $f$ messbar ist. Da das System $\{[a,b) \subseteq \mathfrak{B} \mid a, b \in \mathbb{R}\}$ der halboffenen Intervalle ein Erzeuger von $\mathfrak{B}$ ist, gilt insbesondere

$$\sigma(f) = \sigma\left( \left\{ f^{-1}([a,b)) \,\middle|\, a, b \in \mathbb{R} \right\} \right).$$

Für eine Familie von Funktionen $(f_\iota)_{\iota \in J}$ mit $f_\iota \colon \Omega \longrightarrow \mathbb{R}$ und einer Indexmenge $J$ definieren wir noch

$$\sigma\left(f_\iota\right)_{\iota \in J} := \sigma\left(\bigcup_{\iota \in J} \sigma\left(f_\iota\right)\right),$$

als die von der Familie $(f_\iota)_{\iota \in J}$ *erzeugte $\sigma$-Algebra* auf $\Omega$. Sie die kleinste $\sigma$-Algebra auf $\Omega$, so dass alle $f_\iota$ messbar sind.

Wir interessieren uns nun für eine nähere Untersuchung der Signifikanten $S \colon \mathbb{R} \longrightarrow [1, 10)$. Durch die Symmetrie der Signifikanten, werden wir uns jedoch auf die positiven reellen Zahlen beschränken. Deshalb ist es sinnvoll die folgende Definition zu treffen:

**Definition 1.3** (signifikante $\sigma$-Algebra)**.** Die von $S$ erzeugte $\sigma$-Algebra

$$\mathcal{S} := \mathbb{R}^+ \cap \sigma\left(S\right) = \mathbb{R}^+ \cap \sigma\left(\left\{S^{-1}\left([1,t]\right) \,\middle|\, 1 \leq t < 10\right\}\right)$$

heißt *signifikante $\sigma$-Algebra* auf $\mathbb{R}^+$.

Die signifikante $\sigma$-Algebra ist also für die Untersuchung der signifikanten Stellen einer positiven reellen Zahl geeignet. Der folgende Satz (siehe [11]) gewährt einen Einblick in die Struktur der signifikanten $\sigma$-Algebra und gibt eine spezielle Charakterisierung der Mengen $A \in \mathcal{S}$:

**Satz 1.2.** *Für jedes $A \in \mathcal{S}$ gilt*

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k \cdot S\left(A\right) = \bigcup_{k=-\infty}^{+\infty} \left\{10^k \cdot S\left(x\right) \,\middle|\, x \in A\right\}.$$

*und überdies lässt sich die signifikante $\sigma$-Algebra darstellen als*

$$\mathcal{S} = \mathbb{R}^+ \cap \sigma\left(d_m\right)_{m \in \mathbb{N}} = \left\{\bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\left[1, 10\right)\right\}.$$

Beweis. Da $\mathcal{S} = \mathbb{R}^+ \cap \sigma\left(S\right) = \mathbb{R}^+ \cap \left\{S^{-1}\left(B\right) \,\middle|\, B \in \mathfrak{B}\left[1, 10\right)\right\}$ gibt es für jedes $A \in \mathcal{S}$ ein $B \in \mathfrak{B}\left[1, 10\right)$ mit

$$A = \mathbb{R}^+ \cap S^{-1}\left(B\right) = \left\{x > 0 \,|\, S\left(x\right) \in B\right\}.$$

Wir führen nun folgende Überlegung durch: Ist $S(x) \in B$, so gibt es ein $b \in B$ mit $S(x) = b$, das heißt es gibt ein $k \in \mathbb{Z}$, so dass gilt $x = 10^k \cdot b$, das heißt $x \in 10^k B$. Falls umgekehrt $x \in 10^k B$ für ein $k \in \mathbb{Z}$, so gilt $x = 10^k b$ für ein $b \in B$, so dass wegen $1 \leq b < 10$ per Definition folgt $S(x) = b$. Aus dieser kurzen Überlegung erhalten wir

$$A = \left\{x > 0 \,|\, S\left(x\right) \in B\right\} = \left\{x > 0 \,\middle|\, \exists k \in \mathbb{Z} \colon x \in 10^k B\right\} = \bigcup_{k=-\infty}^{+\infty} 10^k B.$$

Da zudem $S(A) = S\left(\left\{x > 0 \,|\, S(x) \in B\right\}\right) = B$, ergibt sich

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k S\left(A\right).$$

Da $S$ nach Teil *(a)* von Satz 1.1 auf $\mathbb{R}^+$ komplett durch $d_1, d_2, \ldots$ bestimmt ist, gilt $\mathbb{R}^+ \cap \sigma(S) \subseteq \mathbb{R}^+ \cap \sigma(d_m)_{m \in \mathbb{N}}$. Da umgekehrt nach Teil *(b)* von Satz 1.1 auch $d_1, d_2, \ldots$ vollständig auf $\mathbb{R}^+$ durch $S$ bestimmt sind, gilt ebenso $\mathbb{R}^+ \cap \sigma(d_m)_{m \in \mathbb{N}} \subseteq \mathbb{R}^+ \cap \sigma(S)$, das heißt insgesamt

$$\mathcal{S} = \mathbb{R}^+ \cap \sigma\left(S\right) = \mathbb{R}^+ \cap \sigma\left(d_m\right)_{m \in \mathbb{N}}.$$

Es bleibt also nur noch
$$\mathcal{S} = \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\}$$
zu zeigen. Sei dazu $A \in \mathcal{S}$. Wegen dem schon Gezeigten und wegen $S\,(A) \in \mathfrak{B}\,[1,10)$ ist dann
$$A = \bigcup_{k=-\infty}^{+\infty} 10^k S\,(A) \in \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\},$$
das heißt $\mathcal{S} \subseteq \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\}$. Sei nun umgekehrt
$$\bigcup_{k=-\infty}^{+\infty} 10^k B_0 \in \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\}.$$
Wie gerade gezeigt, gilt dann
$$\bigcup_{k=-\infty}^{+\infty} 10^k B_0 = \mathbb{R}^+ \cap S^{-1}\,(B_0) \in \mathbb{R}^+ \cap \sigma\,(S) = \mathcal{S},$$
so dass auch $\left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\} \subseteq \mathcal{S}$ gilt, womit die Gleichheit bewiesen ist. □

Die Darstellung $\mathcal{S} = \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k B \,\middle|\, B \in \mathfrak{B}\,[1,10) \right\}$ aus Satz 1.2 impliziert, dass es für jedes $A \in \mathcal{S}$ ein $B \in \mathfrak{B}\,[1,10)$ gibt mit
$$A = \bigcup_{k=-\infty}^{+\infty} 10^k B.$$
In der Tat zeigt der Beweis von Satz 1.2 sogar, dass dieses $B$ *eindeutig* durch $B \coloneqq S\,(A)$ gegeben ist. Aufgrund Strukturaussage aus Satz 1.2, können wir an einfachen Beispielen feststellen, ob eine Menge zu $\mathcal{S}$ gehört:

**Beispiel 1.3.**

*(a)* $A_1 = \left\{10^k \,\middle|\, k \in \mathbb{Z}\right\} = \{\ldots,\ 10,\ 1,\ 0.1,\ 0.01,\ 0.001,\ \ldots\} = \{x > 0 \,|\, S\,(x) = 1\}$ gehört zu $\mathcal{S}$, denn offenbar ist $A_1 = \bigcup_{k=-\infty}^{+\infty} 10^k \{1\}$ und $\{1\} \in \mathfrak{B}\,[1,10)$. Jedoch gehört die Menge $\{1\}$ selbst *nicht* zu $\mathcal{S}$, denn
$$\{1\} \subsetneq \bigcup_{k=-\infty}^{+\infty} 10^k S\,(\{1\}) = \bigcup_{k=-\infty}^{+\infty} 10^k \{1\} = A_1.$$
Würde $\{1\}$ zu $\mathcal{S}$ gehören so müsste nach Satz 1.2 die Gleichheit gelten.

*(b)* $A_2 = \{x > 0 \,|\, d_2\,(x) = 1,\ d_3\,(x) = 7\} = \{217,\ 91.70583,\ 0.09175,\ 6.17012,\ \ldots\}$ gehört zu $\mathcal{S}$, da
$$A_2 = \mathbb{R}^+ \cap d_2^{-1}\,(\{1\}) \cap d_3^{-1}\,(\{7\}),$$
und nach Satz 1.2 gilt $\mathcal{S} = \mathbb{R}^+ \cap \sigma\,(d_m)_{m \in \mathbb{N}}$.

*(c)* Das Intervall $[2,3]$ gehört *nicht* zu $\mathcal{S}$, denn
$$[2,3] \neq \bigcup_{k=-\infty}^{+\infty} 10^k S\,([2,3]) = \bigcup_{k=-\infty}^{+\infty} 10^k\,[2,3] = \bigcup_{k=-\infty}^{+\infty} \left[10^k 2, 10^k 3\right].$$
Allgemein gilt für jedes Intervall der Form $[a,b]$ mit $1 \leq a \leq b < 10$, das $S\,([a,b]) = [a,b]$ und daher
$$[a,b] \neq \bigcup_{k=-\infty}^{+\infty} 10^k S\,([a,b]) = \bigcup_{k=-\infty}^{+\infty} \left[10^k a, 10^k b\right],$$
was wiederum zeigt $[a,b] \notin \mathcal{S}$.

Insbesondere zeigt Teil *(c)*, dass es „einfache” Mengen gibt, die überraschenderweise *nicht* zu $\mathcal{S}$ gehören. Dies ist ein fundamentaler Unterschied zur Borelschen $\sigma$-Algebra auf $\mathbb{R}^+$ bzw. $\mathbb{R}$. Jedoch gehören „komplizierte” Mengen wie $A_2$ in Teil *(b)* zu $\mathcal{S}$, eben gerade weil solche Mengen eindeutig durch die signifikanten Dezimalziffern bestimmt sind. Dies verdeutlicht gerade die Wahl von $\mathcal{S}$ als passende $\sigma$-Algebra um die signifikanten Stellen einer positiven reellen Zahl zu untersuchen.

Das folgende Lemma liefert weitere interessante Eigenschaften der signifikanten $\sigma$-Algebra und verdeutlicht gewisse Invarianzeigenschaften:

**Lemma 1.2.** *Für die signifikante $\sigma$-Algebra gelten die folgenden Aussagen:*

(a) *$\mathcal{S}$ ist* selbstähnlich *bezüglich der Multiplikation mit Zehnerpotenzen, das heißt*

$$A = 10^k A, \qquad \textit{für jedes } k \in \mathbb{Z} \textit{ und jedes } A \in \mathcal{S}.$$

(b) *$\mathcal{S}$ ist abgeschlossen bezüglich der Multiplikation mit positiven Zahlen, das heißt*

$$\alpha A \in \mathcal{S}, \qquad \textit{für jedes } \alpha > 0 \textit{ und jedes } A \in \mathcal{S}.$$

(c) *$\mathcal{S}$ ist abgeschlossen unter dem Wurzelziehen, das heißt*

$$\sqrt[n]{A} := \{x > 0 \,|\, x^n \in A\} \in \mathcal{S}, \qquad \textit{für jedes } n \in \mathbb{N} \textit{ und jedes } A \in \mathcal{S}.$$

Beweis.

**Zu *(a)*:** Nach Satz 1.2 gilt

$$A = \bigcup_{\ell=-\infty}^{+\infty} 10^\ell S\,(A)\,.$$

Jedoch impliziert Lemma 1.1, dass für jedes $k \in \mathbb{Z}$ gilt $S\left(10^k A\right) = S\,(A)$, das heißt umgekehrt nach Satz 1.2

$$A = \bigcup_{\ell=-\infty}^{+\infty} 10^\ell S\left(10^k A\right) = 10^k A.$$

**Zu *(b)*:** Wegen *(a)* können wir ohne Beschränkung der Allgemeinheit annehmen, dass $1 \leq \alpha < 10$. Denn falls $0 < \alpha < 1$, so gibt es ein $k \in \mathbb{Z}$ mit $1 < \alpha' := 10^k\alpha < 10$ und analog für $\alpha \geq 10$. Sei nun $A = \bigcup_{k=-\infty}^{+\infty} 10^k B \in \mathcal{S}$, für ein $B \in \mathfrak{B}\,[1, 10)$. Dann ist

$$\begin{aligned} \alpha A &= \bigcup_{k=-\infty}^{+\infty} 10^k \alpha B \\ &= \bigcup_{k=-\infty}^{+\infty} 10^k \left[(\alpha B \cap [\alpha, 10)) \cup (\alpha B \cap [10, 10\alpha))\right] \\ &= \bigcup_{k=-\infty}^{+\infty} 10^k \left[(\alpha B \cap [\alpha, 10)) \cup \left(\frac{\alpha}{10} B \cap [1, \alpha)\right)\right]. \end{aligned}$$

Nun ist $(\alpha B \cap [\alpha, 10)) \cup \left(\frac{\alpha}{10} B \cap [1, \alpha)\right) \in \mathfrak{B}\,[1, 10)$, das heißt $\alpha A$ besitzt die Darstellung

$$\alpha A = \bigcup_{k=-\infty}^{+\infty} 10^k B'$$

mit einem $B' \in \mathfrak{B}\,[1, 10)$, was wegen Satz 1.2 bedeutet, dass $\alpha A \in \mathcal{S}$.

**Zu *(c)*:** Da die Intervalle der Form $[1,t]$ mit $1<t<10$ die $\sigma$-Algebra $\mathfrak{B}\,[1,10)$ erzeugen, reicht es die Behauptung für

$$A=\bigcup_{k=-\infty}^{+\infty} 10^k\,[1,10^s]$$

mit $0<s<1$ zu zeigen. Es gilt nun für $n\in\mathbb{N}$

$$\sqrt[n]{A}=\sqrt[n]{\bigcup_{k=-\infty}^{+\infty} 10^k\,[1,10^s]}=\bigcup_{k=-\infty}^{+\infty} 10^{k/n}\left[1,10^{s/n}\right]=\bigcup_{k=-\infty}^{+\infty} 10^k\bigcup_{m=0}^{n-1}\left[10^{m/n},10^{(m+s)/n}\right],$$

so dass mit $B:=\bigcup_{m=0}^{n-1}\left[10^{m/n},10^{(m+s)/n}\right]\in\mathfrak{B}\,[1,10)$ die Darstellung

$$\sqrt[n]{A}=\bigcup_{k=-\infty}^{+\infty} 10^k B$$

gilt, so dass nach Satz 1.2 gilt $\sqrt[n]{A}\in\mathcal{S}$. □

*Bemerkung.* Im Allgemeinen gilt für $n\in\mathbb{N}$ nicht $A^n\in\mathcal{S}$, falls $A\in\mathcal{S}$ ist. Das heißt $\mathcal{S}$ ist im Allgemeinen *nicht* abgeschlossen gegenüber dem Potenzieren. Ein Gegenbeispiel liefert Beispiel 1.4. ♣

**Beispiel 1.4.** Betrachte die Menge aller positiven Zahlen deren erste signifikante Dezimalstelle gerade 1 ist, das heißt die Menge

$$A=\{x>0\,|\,d_1(x)=1\}=\{x>0\,|\,1\le S(x)<2\}=\bigcup_{k=-\infty}^{+\infty} 10^k\,[1,2)\in\mathcal{S}.$$

Nun ist gerade

$$3A=\bigcup_{k=-\infty}^{+\infty} 10^k\,[3,6)=\{x>0\,|\,d_1(x)\in\{3,4,5\}\}\in\mathcal{S},$$

und genauso erhält man

$$\begin{aligned}\sqrt{A}&=\left\{x>0\,\middle|\,x^2\in A\right\}=\left\{x>0\,\middle|\,1\le S(x)<\sqrt{2}\text{ oder }\sqrt{10}\le S(x)<\sqrt{20}\right\}\\&=\bigcup_{k=-\infty}^{+\infty} 10^k\left(\left[1,\sqrt{2}\right)\cup\left[\sqrt{10},\sqrt{20}\right)\right)\in\mathcal{S}.\end{aligned}$$

Nun ist jedoch

$$A^2=\left(\bigcup_{k=-\infty}^{+\infty} 10^k\,[1,2)\right)^2=\bigcup_{k=-\infty}^{+\infty} 10^{2k}\,[1,4)\,.$$

Damit gilt offenbar $S\left(A^2\right)=[1,4)$, aber

$$A^2\neq\bigcup_{k=-\infty}^{+\infty} 10^k S\left(A^2\right)=\bigcup_{k=-\infty}^{+\infty} 10^k\,[1,4)\,,$$

was wegen Satz 1.2 bedeutet $A^2\notin\mathcal{S}$.

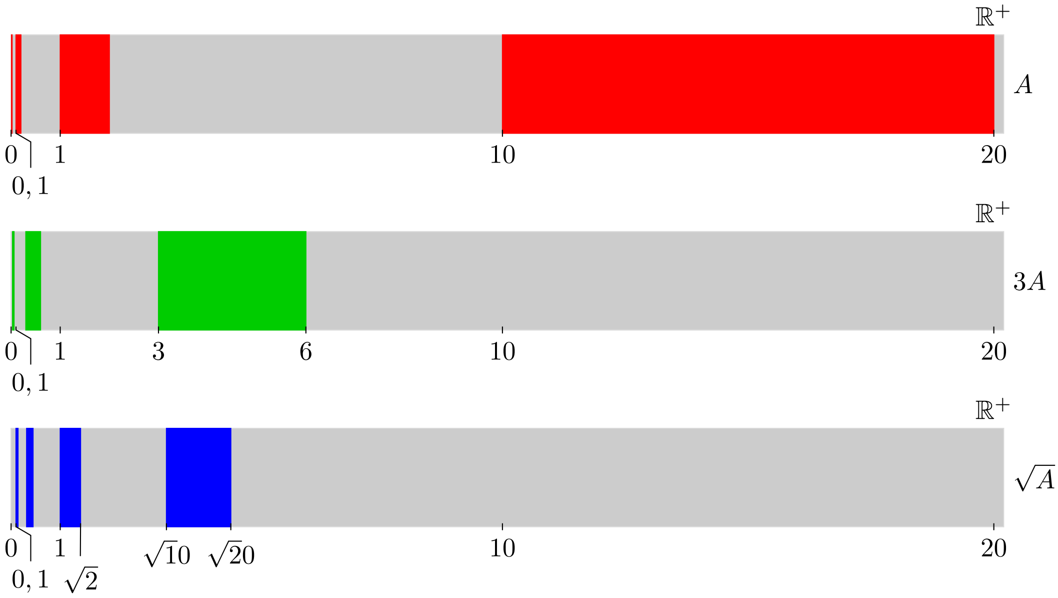


Abbildung 5: Veranschaulichung der Mengen $A$, $3A$ und $\sqrt{A}$ aus Beispiel 1.4.

Wir haben nun bereits viele Eigenschaften von $\mathcal{S}$ studiert und vor allem Unterschiede zur „klassischen" Borelschen $\sigma$-Algebra aufgezeigt: Satz 1.2 gab eine wichtige Aussage über die Struktur von $\mathcal{S}$ und liefert eine wichtige Eigenschaft von Mengen $A \in \mathcal{S}$. Schließlich liefert Lemma 1.2 interessante Abgeschlossenheitsaussagen über $\mathcal{S}$. Jedoch haben wir auch festgestellt, dass es grundlegende Unterschiede zur Borelschen $\sigma$-Algebra gibt – zum Beispiel besagt Beispiel 1.3, dass im Gegensatz zur Borelschen $\sigma$-Algebra keine endlichen Intervalle zu $\mathcal{S}$ gehören.

Das nachfolgende Resultat liefert nun eine fundamentale Aussage über die Beziehung zwischen $(\mathbb{R}^+, \mathcal{S})$ und $([0,1), \mathfrak{B}[0,1))$, in dem es eine Aussage über die Wahrscheinlichkeitsmaße auf den jeweiligen messbaren Räumen gibt. Für einen messbaren Raum $(\Omega, \Sigma)$ bezeichne im Weiteren $\mathcal{P}(\Omega, \Sigma)$ die Menge aller Wahrscheinlichkeitsmaße auf $(\Omega, \Sigma)$. Ferner sei an die Definition des Bildmaßes erinnert: Für einen Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, einen messbaren Raum $(\Omega', \Sigma')$ und einer messbaren Funktion $X\colon \Omega \longrightarrow \Omega'$ wird das *Bildmaß* $P_X$ auf $(\Omega', \Sigma')$ für $A \in \Sigma'$ definiert als

$$P_X(A) := P\left(X^{-1}(A)\right) = P\left(\{\omega \in \Omega \,|\, X(\omega) \in A\}\right).$$

Damit wird $(\Omega', \Sigma', P_X)$ ebenfalls zu einem Wahrscheinlichkeitsraum.

Es folgt nun das angesprochene Resultat über die Wahrscheinlichkeitsmaße auf $(\mathbb{R}^+, \mathcal{S})$ und $([0,1), \mathfrak{B}[0,1))$:

**Satz 1.3.** *Sei* $L\colon \mathbb{R}^+ \longrightarrow [0,1)$ *definiert gemäß* $L(x) := \log S(x)$. *Ferner sei dann* $\Phi\colon \mathcal{P}(\mathbb{R}^+, \mathcal{S}) \longrightarrow \mathcal{P}([0,1), \mathfrak{B}[0,1))$ *definiert gemäß* $\Phi(P) := P_L$. *Dann ist* $\Phi$ *eine bijektive Abbildung zwischen den Wahrscheinlichkeitsmaßen auf* $(\mathbb{R}^+, \mathcal{S})$ *und* $([0,1), \mathfrak{B}[0,1))$.

*Bemerkung.* Satz 1.3 besagt, dass jedes Wahrscheinlichkeitsmaß $Q$ auf $([0,1), \mathfrak{B}[0,1))$ in *eindeutiger* Art und Weise mit einem Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathcal{S})$ derart identifiziert werden kann, dass $Q = P_L$. ♣

Beweis. Zunächst zeigen wir die Wohldefiniertheit der Bildmaße $P_L$ auf $([0,1), \mathfrak{B}[0,1))$. Für $0 \leq a < b < 1$ gilt

$$L^{-1}([a,b]) = S^{-1}\left(\log^{-1}([a,b])\right) = S^{-1}\left(\left[10^a, 10^b\right]\right) = S^{-1}([s,t]),$$

mit $s := 10^a$, $t := 10^b$ und $1 \leq s < t < 10$. Da die Intervalle der Form $[a, b]$ bzw. $[s, t]$ gerade $\mathfrak{B}[0,1)$ bzw. $\mathfrak{B}[1,10)$ erzeugen, gilt somit $\sigma(L) = \mathbb{R}^+ \cap \sigma(S) = \mathcal{S}$. Damit ist $L$ insbesondere messbar. Für jedes Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathcal{S})$ ist also $P_L$ ein wohldefiniertes Wahrscheinlichkeitsmaß auf $([0,1), \mathfrak{B}[0,1))$. Wir zeigen nun noch: Für jedes $A \in \mathcal{S}$ gibt es ein eindeutiges $B \in \mathfrak{B}[0,1)$, so dass

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B, \tag{1.2}$$

wobei $10^B := \left\{10^b \,\middle|\, b \in B\right\}$. Wir wissen bereits nach Satz 1.2, dass es für jedes $A \in \mathcal{S}$ ein *eindeutiges* $B' \in \mathfrak{B}[1,10)$ gibt, so dass

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k B'.$$

Es gilt nun $B' = 10^B$ für

$$B := \log B' = \{\log b' \mid b' \in B'\} \in \mathfrak{B}[0,1).$$

Damit ist

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B.$$

**Zur Surjektivität:** Sei $Q$ ein Wahrscheinlichkeitsmaß auf $([0,1), \mathfrak{B}[0,1))$. Wir zeigen, dass es dann ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathcal{S})$ gibt mit $Q = \Phi(P) = P_L$. Ist die Darstellung (1.2) gegeben, so definieren wir ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathcal{S})$ durch

$$P(A) := Q(B).$$

Es gilt nun

$$\begin{aligned} L^{-1}(B) &= \{x > 0 \mid \log S(x) \in B\} = \left\{x > 0 \,\middle|\, S(x) \in 10^B\right\} \\ &= \left\{x > 0 \,\middle|\, \exists k \in \mathbb{Z}\; \exists s \in B\colon x = 10^k 10^s\right\} \\ &= \bigcup_{k=-\infty}^{+\infty} 10^k 10^B = A. \end{aligned}$$

Daraus ergibt sich

$$P_L(B) = P\left(L^{-1}(B)\right) = P(A) = Q(B),$$

das heißt $\Phi(P) = P_L = Q$.

**Zur Injektivität:** Seien $P$ und $P'$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^+, \mathcal{S})$ mit $P_L = P'_L$. Das heißt für alle $B \in \mathfrak{B}[0,1)$ gilt $P_L(B) = P'_L(B)$. Jedoch ist für $A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B \in \mathcal{S}$ wegen $A = L^{-1}(B)$

$$P(A) = P\left(L^{-1}(B)\right) = P_L(B) = P'_L(B) = P'\left(L^{-1}(B)\right) = P'(A).$$

Da $A \in \mathcal{S}$ beliebig gewählt war, gilt also $P = P'$. Damit ist die Abbildung $\Phi$ zudem injektiv. □

Die Abbildung $\Phi$ ist nicht die einzige bijektive Abbildung zwischen den Mengen $\mathcal{P}(\mathbb{R}^+, \mathcal{S})$ und $\mathcal{P}([0,1), \mathfrak{B}[0,1))$. Beispielsweise hätten wir genauso die Abbildung $L_0\colon \mathbb{R}^+ \longrightarrow [0,1)$ definiert gemäß

$$L_0(x) := \frac{1}{9}(S(x) - 1)$$

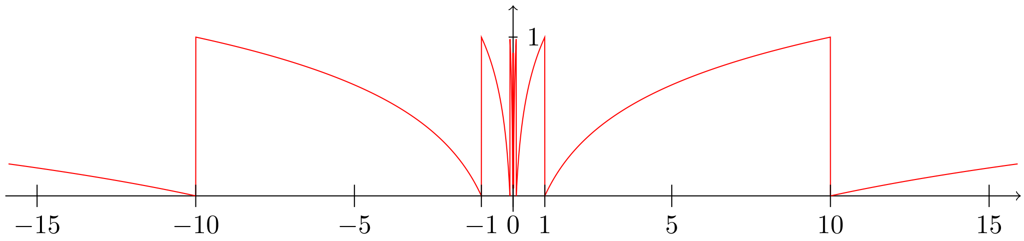


Abbildung 6: Der Graph von $L(x) = \log S(x) = \log|x| - \lfloor \log|x| \rfloor$.

und die zugehörige Abbildung $\Psi\colon \mathcal{P}(\mathbb{R}^+, \mathcal{S}) \longrightarrow \mathcal{P}([0,1), \mathfrak{B}[0,1))$ mit $\Psi(P) := P_{L_0}$ betrachten können. Die spezielle Rolle von

$$L(x) = \log S(x)$$

ist die Beziehung zu dem Benfordschen Gesetz: Um einen ersten Einblick darin zu gewinnen definiere durch

$$\mathbb{B}(\{x > 0 \,|\, S(x) \leq t\}) = \mathbb{B}\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1,t]\right) := \log t, \qquad \text{für alle } 1 \leq t < 10 \tag{1.3}$$

eine Mengenfunktion $\mathbb{B}$ auf $(\mathbb{R}^+, \mathfrak{E})$, wobei $\mathfrak{E} := \left\{\bigcup_{k=-\infty}^{+\infty} 10^k [1,t] \,\middle|\, 1 \leq t < 10\right\}$. Da nun $\{[1,t] \,|\, 1 \leq t < 10\}$ ein Erzeuger von $\mathfrak{B}[1,10)$ ist, ist dann

$$\mathbb{R}^+ \cap \left\{S^{-1}([1,t]) \,\middle|\, 1 \leq t < 10\right\} = \mathbb{R}^+ \cap \left\{\bigcup_{k=-\infty}^{+\infty} 10^k [1,t] \,\middle|\, 1 \leq t < 10\right\} = \mathfrak{E}$$

ein Erzeuger von $\mathcal{S}$, denn

$$\begin{aligned} \mathcal{S} &= \mathbb{R}^+ \cap \sigma(S) = \mathbb{R}^+ \cap S^{-1}(\mathfrak{B}[1,10)) = \mathbb{R}^+ \cap S^{-1}(\sigma(\{[1,t] \,|\, 1 \leq t < 10\})) \\ &= \mathbb{R}^+ \cap \sigma\left(S^{-1}(\{[1,t] \,|\, 1 \leq t < 10\})\right) = \sigma\left(\mathbb{R}^+ \cap S^{-1}(\{[1,t] \,|\, 1 \leq t < 10\})\right) \\ &= \sigma(\mathfrak{E}), \end{aligned}$$

wobei die Vertauschung von Urbild-Bildung und $\sigma$-Algebra-Erzeugung ausgenutzt wurde. Da $\mathfrak{E}$ zudem durchschnittsstabil ist, gibt es wegen dem Fortsetzungssatz von Carathéodory[1] dann *genau* ein Wahrscheinlichkeitsmaß $\mathbb{B}$ auf ganz $(\mathbb{R}^+, \mathcal{S})$ mit der Eigenschaft (1.3). Dieses werden wir später die *Benford-Verteilung*[2] auf $(\mathbb{R}^+, \mathcal{S})$ nennen. Sie spielt eine herausragende Rolle unter allen Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^+, \mathcal{S})$. Eines der „natürlichsten” Wahrscheinlichkeitsmaße auf $([0,1), \mathfrak{B}[0,1))$ ist die Gleichverteilung[3] $\lambda\!\!\lambda_{0,1}$. Die spezielle Wahl von $L$ in Satz 1.3 erklärt sich nun an der Tatsache

$$\mathbb{B}_L = \lambda\!\!\lambda_{0,1}. \tag{1.4}$$

Um dies zu erkennen, betrachten wir auf dem Erzeuger $\mathfrak{I} := \{[0,t] \,|\, 0 \leq t < 1\}$ von $\mathfrak{B}[0,1)$ gerade

$$\begin{aligned} \mathbb{B}_L([0,t]) &= \mathbb{B}\left(L^{-1}([0,t])\right) = \mathbb{B}\left(S^{-1}\left(\left[1, 10^t\right]\right)\right) = \mathbb{B}\left(\left\{x > 0 \,\middle|\, S(x) \leq 10^t\right\}\right) \\ &= \log 10^t = t = \lambda\!\!\lambda_{0,1}([0,t]), \end{aligned}$$

so dass die Aussage auf dem durchschnittsstabilen Erzeuger $\mathfrak{I}$ von $\mathfrak{B}[0,1)$ gültig ist und damit auch auf ganz $([0,1), \mathfrak{B}[0,1))$ richtig ist.

[1] siehe zum Beispiel [13] Kapitel 1.
[2] siehe Definition 2.5.
[3] Die Gleichverteilung auf $([0,1), \mathfrak{B}[0,1))$ ist das Lebesgue-Maß auf $([0,1), \mathfrak{B}[0,1))$.

# 2 Benford Eigenschaft und Charakterisierungen

Das von Newcomb und Benford beobachtete empirische Benfordsche Gesetz (1.1) wollen wir nun mathematisch präzisieren. Zunächst gilt es einmal zu präzisieren, welche Objekte betrachtet werden. In unserem Fall wird es sich um reelle Zahlenfolgen, reelle Funktionen mit Definitionsbereich $[0, +\infty)$, Zufallsgrößen und Wahrscheinlichkeitsmaße handeln. Wir werden dann formulieren, was es für die jeweiligen Objekte heißt, dass sie die sogenannte Benford-Eigenschaft besitzen. Danach werden wir verschiedene Kriterien kennenlernen, um die Benford-Eigenschaft für die verschiedenen Objekte zu charakterisieren.

## 2.1 Definitionen

Zunächst betrachten wir reelle Zahlenfolgen $(x_n)_{n\in\mathbb{N}}$. Es ist natürlich davon zu sprechen, dass eine Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ die *Benford-Eigenschaft* hat, falls im Grenzwert der Anteil der Folgenglieder mit signifikanter Anfangsziffer $d \in \mathbb{N}_{1,9}$ gerade gegen $\log\left(1 + \frac{1}{d}\right)$ konvergiert und analog für alle weiteren signifikanten Dezimalstellen. Wir werden jedoch stattdessen eine äquivalante Definition mithilfe der Signifikanten treffen. Dies führt auf folgende

**Definition 2.1.** Eine reelle Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ besitzt die *Benford-Eigenschaft*, falls

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in \mathbb{N}_{1,N} \mid S(x_n) \le t\} = \log t, \qquad \text{für alle } 1 \le t < 10.$$

Wir sagen auch kurz, dass die Folge $(x_n)_{n\in\mathbb{N}}$ Benford ist.

*Bemerkung.*

*(a)* Mittels charakteristischen Funktionen lässt sich obige Eigenschaft umschreiben in

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{0\}\cup[1,t]}(S(x_n)) = \log t, \qquad \text{für alle } 1 \le t < 10.$$

Hierbei handelt es sich um den Cesáro-Limes der Folge $\left(\mathbf{1}_{\{0\}\cup[1,t]}(S(x_n))\right)_{n\in\mathbb{N}}$.

*(b)* Äquivalent zu Definition 2.1 ist: Für alle $m \in \mathbb{N}$, alle $D_1 \in \mathbb{N}_{1,9}$ und alle $D_k \in \mathbb{Z}_{0,9}$ gilt

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in \mathbb{N}_{1,N} \mid \forall k \in \mathbb{N}_{1,m} \colon d_k(x_n) = D_k\} = \log\left(1 + \left(\sum_{k=1}^{m} 10^{m-k} D_k\right)^{-1}\right).$$

Dies ist der Tatsache geschuldet, dass nach Satz 1.1 $S$ eindeutig durch die signifikanten Dezimalziffern $d_1, d_2, \ldots$ bestimmt ist. ♣

Besitzt eine Folge $(x_n)_{n\in\mathbb{N}}$ die Benford-Eigenschaft, so gilt also insbesondere

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in \mathbb{N}_{1,N} \mid d_1(x_n) = 1\} = \log\left(1 + \frac{1}{1}\right) = \log 2,$$

so dass der Anteil der Folgenglieder mit signifikanter Anfangsziffer 1 gerade $\log 2$ ist und analog für alle anderen Anfangsziffern

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in \mathbb{N}_{1,N} \mid d_1(x_n) = d\} = \log\left(1 + \frac{1}{d}\right), \qquad \text{für alle } d \in \mathbb{N}_{1,9}.$$

Damit genügt eine Folge mit Benford-Eigenschaft dem empirischen Benfordschen Gesetz (1.1).

Zwei der prominentesten Folgen sind die Folgen der Primzahlen $(P_n)_{n\in\mathbb{N}} = (2, 3, 5, 7, 11, \dots)$ und die Folge der *Fibonacci-Zahlen* $(F_n)_{n\in\mathbb{N}} = (1, 1, 2, 3, 5, \dots)$, wobei diese durch die bekannte Rekursion

$$F_{n+1} = F_n + F_{n-1}, \qquad F_1 = F_2 = 1,$$

gebildet wird. Eine interessante Frage ist nun, ob diese beiden Folgen gerade die Benford-Eigenschaft haben. Es wird sich in Beispiel 2.4 zeigen, dass die Primzahlen *nicht* Benford sind, wohingegen die Fibonacci-Zahlen Benford sind. Abbildung 7 und Tabelle 1 geben einen ersten Hinweis hierfür: Für $N \in \mathbb{N}$ definieren wir für eine Folge $(x_n)_{n\in\mathbb{N}}$

$$\varrho_N(d) := \frac{1}{N} \operatorname{card}\{n \in \mathbb{N}_{1,N} \mid d_1(x_n) = d\}, \qquad \text{für } d \in \mathbb{N}_{1,9}.$$

Dann gibt $\varrho_N(d)$ den Anteil der ersten $N$ Folgenglieder an, welche die signifikante Anfangsziffer $d$ haben. Falls die Folge Benford ist, so gilt für hinreichend großes $N$ und beliebiges $\varepsilon > 0$ dann $\left|\varrho_N(d) - \log\left(1 + \frac{1}{d}\right)\right| < \varepsilon$, so dass $\varrho_N(d)$ sehr nahe bei $\log\left(1 + \frac{1}{d}\right)$ liegt.

Die Anzahl der jeweiligen Anfangsziffern der ersten 100 bzw. 10.000 Fibonacci-Zahlen und Primzahlen sind in Tabelle 1 gelistet. Bildet man die „Dichten" $\varrho_{10.000}(d)$ für beide Folgen, so erhält man Abbildung 7. Erstaunlich ist hierbei, wie genau die Fibonacci-Zahlen dem

| | $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $F_n$ | $N = 100$ | 30 | 18 | 13 | 9 | 8 | 6 | 5 | 7 | 4 |
| | $N = 10.000$ | 3011 | 1762 | 1250 | 968 | 792 | 668 | 580 | 513 | 456 |
| $P_n$ | $N = 100$ | 25 | 19 | 19 | 20 | 8 | 2 | 4 | 2 | 1 |
| | $N = 10.000$ | 1601 | 1129 | 1097 | 1069 | 1055 | 1013 | 1027 | 1003 | 1006 |

Tabelle 1: Vergleich der Anfangsziffern der ersten 100 und 10.000 Fibonacci- bzw. Primzahlen. Die Daten wurden aus [4] entnommen.

Benfordschen Gesetz genügen. Definiert man den Fehler $\varepsilon_N$ für $N \in \mathbb{N}$ zum Beispiel als maximale Abweichung vom empirischen Benfordschen Gesetz, das heißt

$$\varepsilon_N := \max_{d=1,\dots,9} \left|\varrho_N(d) - \log\left(1 + \frac{1}{d}\right)\right|,$$

so beträgt er bei den Fibonacci-Zahlen für $N = 10.000$ gerade einmal $0,0001574$.

Wir betrachten nun Borel-messbare Funktionen $f\colon [0, +\infty) \longrightarrow \mathbb{R}$. Wir gehen dabei stets implizit davon aus, dass eine $\sigma$-Algebra $\Sigma$ auf $[0, +\infty)$ gegeben ist und betrachten auf $\mathbb{R}$ stets die Borelsche $\sigma$-Algebra $\mathfrak{B}$. Das natürlichste Maß auf $(\mathbb{R}, \mathfrak{B})$ ist das Lebesgue-Maß $\lambda$.

In Anlehnung an Definition 2.1 für Folgen ist es sinnvoll zu sagen, dass eine Funktion $f\colon [0, +\infty) \longrightarrow \mathbb{R}$ die *Benford-Eigenschaft* besitzt, falls der Anteil jener Funktionswerte $f(\tau)$, wobei $0 \le \tau < T$, mit signifikanter Anfangsziffer $d$ gerade im Grenzwert $T \to +\infty$ $\log\left(1 + \frac{1}{d}\right)$ ist. Hierbei wird der „Anteil" gerade als Lebesgue-Maß interpretiert. Dies führt mittels der Signifikanten im Allgemeinen Fall auf folgende

**Definition 2.2.** Eine Borel-messbare Funktion $f\colon [0, +\infty) \longrightarrow \mathbb{R}$ besitzt die *Benford-Eigenschaft* oder heißt kurz *Benford*, falls

$$\lim_{T\to\infty} \frac{1}{T} \lambda\left(\{0 \le \tau < T \mid S(f(\tau)) \le t\}\right) = \log t, \qquad \text{für alle } 1 \le t < 10.$$

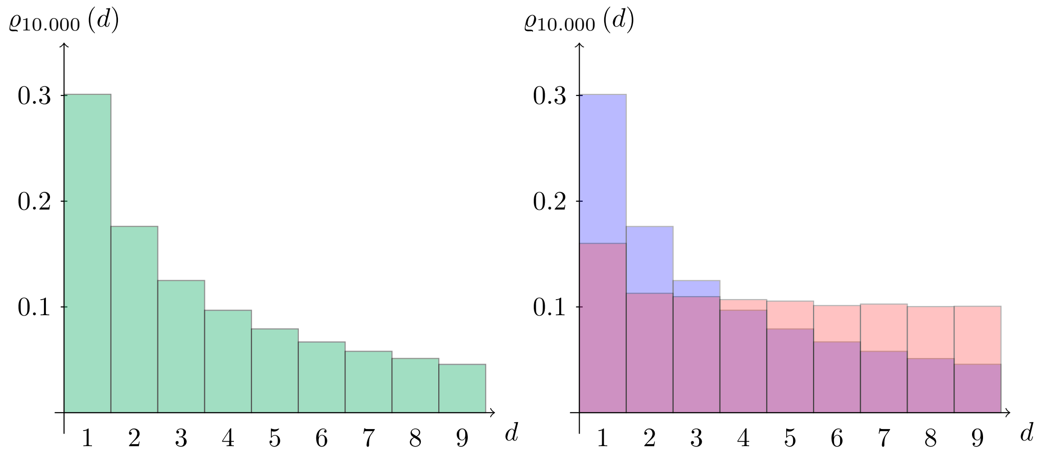


Abbildung 7: Während bei den Fibonacci-Zahlen (grün) mit bloßem Auge keine Abweichung vom Benfordschen Gesetz (blau) zu entdecken ist, sieht man bei den Primzahlen (rot) eine deutliche Abweichung.

*Bemerkung.* Wegen Satz 1.1 ist Folgendes äquivalent zur Definition 2.2: Für alle $m \in \mathbb{N}$, alle $D_1 \in \mathbb{N}_{1,9}$ und alle $D_k \in \mathbb{Z}_{0,9}$ gilt

$$\lim_{T\to\infty} \frac{1}{T} \lambda\!\!\lambda \left(\{0 \leq \tau < T \,|\, \forall k \in \mathbb{N}_{1,m} \colon d_k\left(f\left(\tau\right)\right) = D_k\}\right) = \log\left(1 + \left(\sum_{k=1}^{m} 10^{m-k} D_k\right)^{-1}\right).$$

♣

Der Ursprung des Benfordschen Gesetzes liegt in der Beobachtung, dass bestimmte Datensätze gerade der Gleichung (1.1) genügen – das heißt der Ursprung liegt gerade in der Statistik und Stochastik gewisser Daten. Um das Auftreten des Benfordschen Gesetzes in diesen Bereichen zu erklären, benötigen wir ähnliche Definitionen wie für Zahlenfolgen und Funktionen nun auch für Zufallsgrößen und Wahrscheinlichkeitsmaße.

Dabei sei daran erinnert, dass eine Zufallsgröße $X \colon \Omega \longrightarrow \mathbb{R}$ nichts anderes ist als eine messbare Abbildung zwischen den Räumen $(\Omega, \Sigma)$ und $(\mathbb{R}, \mathfrak{B})$. Ist die Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ gegeben, definiert man ihre *Verteilung* $P_X$ als

$$P_X\left((-\infty, t]\right) := P\left(X^{-1}\left((-\infty, t]\right)\right) = P\left(\{X \in (-\infty, t]\}\right) = P\left(\{X \leq t\}\right), \qquad \text{für alle } t \in \mathbb{R}.$$

Das so gegebene eindeutige Wahrscheinlichkeitsmaß $P_X$ macht $(\mathbb{R}, \mathfrak{B}, P_X)$ zu einem Wahrscheinlichkeitsraum. Ferner ist die *Verteilungsfunktion* $F_P$ eines Maßes $P$ auf $(\mathbb{R}, \mathfrak{B})$ definiert durch

$$F_P\left(t\right) := P\left((-\infty, t]\right), \qquad \text{für alle } t \in \mathbb{R}.$$

Wir schreiben kurz $F_X$ für die Verteilungsfunktion $F_{P_X}$ von $P_X$ und nennen $F_X$ die Verteilungsfunktion der Zufallsgröße $X$. Für die späteren Betrachtungen treffen wir noch folgende

**Definition 2.3.**

*(a)* Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ heißt *stetig*, falls $P(\{t\}) = 0$ für alle $t \in \mathbb{R}$. Genauso nennen wir eine Zufallsgröße $X\colon \Omega \longrightarrow \mathbb{R}$ *stetig*, falls deren Verteilung $P_X$ stetig ist.

*(b)* Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ heißt *absolutstetig* (bezüglich des Lebesgue-Maßes $\lambda\!\!\lambda$), falls für jedes $B \in \mathfrak{B}$ mit $\lambda(B) = 0$ auch $P(B) = 0$ ist. Genauso nennen wir eine Zufallsgröße $X\colon \Omega \longrightarrow \mathbb{R}$ *absolutstetig* (bezüglich des Lebesgue-Maßes $\lambda\!\!\lambda$), falls deren Verteilung $P_X$ absolutstetig ist.

Nach dem Satz von Radon-Nikodym ist die Absolutstetigkeit eines Wahrscheinlichkeitsmaßes äquivalent zur Existenz einer messbaren Abbildung $f_P\colon \mathbb{R} \longrightarrow [0, +\infty)$ mit

$$P([a,b]) = \int_a^b f_P(x)\, \lambda\!\!\lambda(\mathrm{d}x), \qquad \text{für alle } [a,b] \subseteq \mathbb{R} \text{ mit } a < b.$$

Dieses $f_P$ nennen wir *Dichte* von $P$. Für eine Zufallsgröße $X$ schreiben wir kurz $f_X$ für $f_{P_X}$ und nennen $f_X$ auch kurz Dichte von $X$. Da Dichten für festes $P$ nur $\lambda\!\!\lambda$-fast-überall eindeutig bestimmt sind, gibt es im Allgemeinen mehrere sogenannte Dichte-Versionen. Jedoch werden wir im Hinblick auf die $\lambda\!\!\lambda$-fast-überall Eindeutigkeit trotzdem von *der* Dichte $f_P$ sprechen.

Schließlich bezeichnen wir für ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ noch $|P|$ als jenes Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$, welches definiert ist durch

$$|P|(B) := P_{|\cdot|}(B) = P(\{x \in \mathbb{R} \mid |x| \in B\}), \qquad \text{für alle } B \in \mathfrak{B}.$$

Offenbar gilt $|P|([0, +\infty)) = 1$ und $|P|((-\infty, 0)) = 0$. Eine kurze Rechnung ergibt zudem

$$F_{|P|}(t) = (F_P(t) + F_P(-t) + P(\{-t\}))\, \mathbf{1}_{[0,+\infty)},$$

und

$$f_{|P|}(t) = (f_P(t) + f_P(-t))\, \mathbf{1}_{[0,+\infty)},$$

wobei der Transformationssatz für Dichten[4] angewandt wird, falls $P$ eine Dichte $f_P$ besitzt.

Wir kommen nun zur vorher angesprochenen Definition der Benford-Eigenschaft für Wahrscheinlichkeitsmaße und Zufallsgrößen:

**Definition 2.4.**

*(a)* Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ besitzt die *Benford-Eigenschaft* oder heißt kurz *Benford*, falls

$$P(\{x \in \mathbb{R} \mid S(x) \le t\}) = \log t, \qquad \text{für alle } 1 \le t < 10.$$

*(b)* Eine Zufallsgröße $X\colon \Omega \longrightarrow \mathbb{R}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ besitzt die *Benford-Eigenschaft* oder heißt kurz *Benford*, falls $P_X$ Benford ist, das heißt

$$P(\{S(X) \le t\}) = P_X(\{x \in \mathbb{R} \mid S(x) \le t\}) = \log t, \qquad \text{für alle } 1 \le t < 10.$$

*Bemerkung.*

*(a)* Wieder lassen sich mithilfe von Satz 1.1 äquivalente Definitionen mit den signifikanten Dezimalziffern treffen.

*(b)* Es ist

$$\begin{aligned} P(\{x \in \mathbb{R} \mid S(x) \le t\}) &= P(\{S \in \{0\} \cup [1,t]\}) = P\left(S^{-1}(\{0\} \cup [1,t])\right) \\ &= P_S(\{0\} \cup [1,t]), \end{aligned}$$

so dass die Benford-Eigenschaft für Wahrscheinlichkeitsmaße eine Forderung an das Bildmaß bezüglich der Signifikanten ist.

[4] siehe zum Beispiel [13] Kapitel 1.

*(c)* Für $t = 1$ erkennt man inbesondere

$$0 = \log 1 = P_S\left(\{0\} \cup [1,1]\right) = P_S\left(\{0\} \cup \{1\}\right) = P\left(\{S = 0\}\right) + P\left(\{S = 1\}\right),$$

so dass $P\left(\{S = 0\}\right) = P\left(\{S = 1\}\right) = 0$ eine notwendige Bedingung für die Benford-Eigenschaft von $P$ ist. ♣

**Beispiel 2.1.** Für $k \in \mathbb{Z}$ sei das Wahrscheinlichkeitsmaß $P^{(k)}$ definiert durch Angabe der Dichte (bezüglich des Lebesgue-Maßes $\lambda\!\!\!\lambda$)

$$f_k(x) := \frac{1}{x \cdot \ln 10} \mathbf{1}_{[10^k, 10^{k+1})}.$$

Dann ist $P^{(k)}$ stetig, auf dem Intervall $\left[10^k, 10^{k+1}\right)$ konzentriert und es ergibt sich

$$\begin{aligned} P^{(k)}\left(\{x \in \mathbb{R} \,|\, S(x) \le t\}\right) &= P^{(k)}\left(\bigcup_{\ell=-\infty}^{+\infty} 10^\ell [1,t]\right) = \sum_{\ell=-\infty}^{+\infty} P^{(k)}\left(\left[10^\ell, 10^\ell t\right]\right) \\ &= P^{(k)}\left(\left[10^k, 10^k t\right]\right) = \int_{10^k}^{10^k t} \frac{\lambda\!\!\!\lambda(\mathrm{d}x)}{x \cdot \ln 10} = \frac{1}{\ln 10}\left(\ln 10^k t - \ln 10^k\right) \\ &= \frac{1}{\ln 10} \ln t = \log t, \end{aligned}$$

so dass $P^{(k)}$ Benford ist. Ferner gilt für jede Folge $(\alpha_k)_{k\in\mathbb{Z}}$ nichtnegativer Zahlen mit der Eigenschaft $\sum_{k=-\infty}^{+\infty} \alpha_k = 1$ noch

$$\sum_{k=-\infty}^{+\infty} \alpha_k P^{(k)}\left(\{x \in \mathbb{R} \,|\, S(x) \le t\}\right) = \sum_{k=-\infty}^{+\infty} \alpha_k \log t = \log t,$$

so dass auch jede Konvexkombination $\sum_{k=-\infty}^{+\infty} \alpha_k P^{(k)}$ der Wahrscheinlichkeitsmaße $P^{(k)}$ wieder Benford ist. Dies verdeutlicht, dass es „sehr viele” Wahrscheinlichkeitsmaße mit Benford-Eigenschaft gibt.

**Beispiel 2.2.** Sei $U$ eine auf $[0,1)$ gleichverteilte Zufallsgröße, das heißt $P_U = \lambda\!\!\!\lambda_{0,1}$ bzw.

$$P_U([a,b]) = b - a, \qquad \text{für alle } [a,b] \subseteq [0,1) \text{ mit } a < b.$$

Dann ist wegen der Stetigkeit von $U$ insbesondere $P_U(\{0\}) = 0$ und es gilt

$$\begin{aligned} P\left(\{S(U) \le t\}\right) = P_U\left(\{S \le t\}\right) &= P_U\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1,t]\right) \\ &= \sum_{n=1}^{\infty} P_U\left(10^{-n}[1,t]\right) = (t-1) \cdot \sum_{n=1}^{\infty} 10^{-n} \\ &= \frac{t-1}{9} \ne \log t, \end{aligned}$$

so dass $U$ *nicht* Benford ist. Wir betrachten nun stattdessen die Zufallsgröße $X := 10^U$. Dann gilt für $1 \le t < 10$

$$\begin{aligned} P\left(\{S(X) \le t\}\right) &= P\left(\left\{S \circ 10^U \in [1,t]\right\}\right) = P\left(U^{-1}\left(\log\left(S^{-1}([1,t])\right)\right)\right) \\ &= P_U\left(\log\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1,t]\right)\right) = P_U\left(\bigcup_{k=-\infty}^{+\infty} [k, k + \log t]\right) \\ &= P_U([0, \log t]) = \lambda\!\!\!\lambda_{0,1}([0, \log t]) = \log t, \end{aligned}$$

so dass $X = 10^U$ die Benford-Eigenschaft besitzt.

Sei nun $X$ eine Zufallsgröße mit Benford-Eigenschaft, das heißt $P(\{S(X) \leq t\}) = \log t$ für alle $1 \leq t < 10$. Dann gilt für die Verteilungsfunktion der Zufallsgröße $S(X) : \Omega \longrightarrow [1, 10)$ dann

$$F_{S(X)}(t) = P(\{S(X) \leq t\}) = \begin{cases} 0, & \text{falls } t < 1, \\ \log t, & \text{falls } 1 \leq t < 10, \\ 1, & \text{falls } t > 10. \end{cases}$$

Da $F_{S(X)}$ $\lambda\!\!\!\lambda$-fast-überall differenzierbar ist, ergibt sich die Dichte von $S(X)$ dann als

$$f_{S(X)}(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_{S(X)}(t) = \frac{\mathrm{d}}{\mathrm{d}t} \log t = \frac{1}{\ln 10} \cdot \frac{1}{t}, \qquad \text{für } 1 \leq t < 10.$$

Daher ist

$$f_{S(X)}(t) = \begin{cases} 0, & \text{falls } t < 1, \\ \frac{1}{\ln 10} \cdot \frac{1}{t}, & \text{falls } 1 \leq t < 10, \\ 0, & \text{falls } t > 10, \end{cases}$$

die Dichte von $S(X)$. Damit ergibt sich aber auch

$$\begin{aligned} \mathrm{E}[S(X)] &= \int_\Omega S(X)\,\mathrm{d}P = \int_1^{10} x \cdot f_{S(X)}(x)\,\lambda\!\!\!\lambda(\mathrm{d}x) = \frac{1}{\ln 10} \int_1^{10} x \cdot \frac{1}{x} \lambda\!\!\!\lambda(\mathrm{d}x) \\ &= \frac{1}{\ln 10} \cdot (10 - 1) = \frac{9}{\ln 10} \approx 3{,}9087. \end{aligned}$$

Wie bereits am Ende des Kapitels 1 angedeutet, folgt nun die Definition der Benford-Verteilung auf $(\mathbb{R}^+, \mathcal{S})$:

**Definition 2.5.** Das eindeutig bestimmte Wahrscheinlichkeitsmaß $\mathbb{B}$ auf $(\mathbb{R}^+, \mathcal{S})$ mit der Eigenschaft

$$\mathbb{B}(\{S \leq t\}) = \mathbb{B}\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1, t]\right) = \log t, \qquad \text{für alle } 1 \leq t < 10,$$

heißt *Benford-Verteilung* auf $(\mathbb{R}^+, \mathcal{S})$.

Es folgt nun eine erste Charakterisierung von Benford-Wahrscheinlichkeitsmaßen mithilfe der Benford-Verteilung:

**Satz 2.1.** *Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ ist genau dann Benford, wenn*

$$|P|(A) = \mathbb{B}(A), \qquad \textit{für alle } A \in \mathcal{S}.$$

Insbesondere gilt falls $P(\mathbb{R}^+) = 1$, dass $P$ genau dann Benford ist, wenn $P(A) = \mathbb{B}(A)$ für alle $A \in \mathcal{S}$.

Beweis.

**„⇒":** Sei $P$ Benford, das heißt

$$P(\{x \in \mathbb{R} \mid S(x) \leq t\}) = \log t, \qquad \text{für alle } 1 \leq t < 10.$$

Speziell für $A = \bigcup_{k=-\infty}^{+\infty} 10^k [1, t]$ mit $1 \leq t < 10$ gilt dann

$$\begin{aligned} |P|(A) &= P_{|\cdot|}(A) = P_{|\cdot|}\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1, t]\right) = P\left(\left\{x \in \mathbb{R} \,\middle|\, |x| \in \bigcup_{k=-\infty}^{+\infty} 10^k [1, t]\right\}\right) \\ &= P(\{x \in \mathbb{R} \mid S(|x|) \leq t\}) = P(\{x \in \mathbb{R} \mid S(x) \leq t\}) = \log t \\ &= \mathbb{B}\left(\bigcup_{k=-\infty}^{+\infty} 10^k [1, t]\right) = \mathbb{B}(A), \end{aligned}$$

da $S(x) = S(|x|)$ für alle $x \in \mathbb{R}$. Somit gilt also $|P|(A) = \mathbb{B}(A)$ für alle $A \in \mathcal{S}$ der obigen Form. Da jedoch $\mathfrak{E} := \left\{ \bigcup_{k=-\infty}^{+\infty} 10^k [1,t] \,\middle|\, 1 \leq t < 10 \right\}$ ein durchschnittsstabiler Erzeuger von $\mathcal{S}$ ist und $|P| = \mathbb{B}$ auf $\mathfrak{E}$ gilt, muss auch $|P| = \mathbb{B}$ auf ganz $\mathcal{S}$ gelten.

**„$\Leftarrow$":** Gelte $|P|(A) = \mathbb{B}(A)$ für alle $A \in \mathcal{S}$. Dann gilt insbesondere für $A = \bigcup_{k=-\infty}^{+\infty} 10^k [1,t] \in \mathcal{S}$ mit $1 \leq t < 10$

$$\begin{aligned} P(\{x \in \mathbb{R} \mid S(x) \leq t\}) &= P(\{x \in \mathbb{R} \mid S(|x|) \leq t\}) = P(\{x \in \mathbb{R} \mid |x| \in A\}) \\ &= |P|(A) = \mathbb{B}(A) = \mathbb{B}\left( \bigcup_{k=-\infty}^{+\infty} 10^k [1,t] \right) = \log t, \end{aligned}$$

so dass $P$ Benford ist. $\square$

## 2.2 Die Gleichverteilungs-Charakterisierung

In diesem Abschnitt folgt ein erster zentraler Satz, der die Benford-Eigenschaft für Folgen, Funktionen, Wahrscheinlichkeitsmaße und Zufallsgrößen charakterisiert. Sei für eine reelle Zahl $x$ der *gebrochene Anteil* definiert als

$$\langle x \rangle := x - \lfloor x \rfloor .$$

In diesem Abschnitt werden wir wesentlich die Theorie der *Gleichverteilung modulo 1* benutzen. Dazu definieren wir:

**Definition 2.6.**

*(a)* Eine reelle Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ heißt *gleichverteilt modulo 1*, falls

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card} \{n \in \mathbb{N}_{1,N} \mid \langle x_n \rangle \leq s\} = s, \qquad \text{für alle } 0 \leq s < 1.$$

*(b)* Eine Borel-messbare Funktion $f\colon [0,+\infty) \longrightarrow \mathbb{R}$ heißt *gleichverteilt modulo 1*, falls

$$\lim_{T\to\infty} \frac{1}{T} \lambda\!\!\lambda (\{0 \leq \tau < T \mid \langle f(\tau) \rangle \leq s\}) = s, \qquad \text{für alle } 0 \leq s < 1.$$

*(c)* Eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ heißt *gleichverteilt modulo 1*, falls

$$P_{\langle X \rangle}([0,s]) = P(\{\langle X \rangle \leq s\}) = s, \qquad \text{für alle } 0 \leq s < 1.$$

*(d)* Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathfrak{B})$ heißt *gleichverteilt modulo 1*, falls

$$P(\{x \in \mathbb{R} \mid \langle x \rangle \leq s\}) = P\left( \bigcup_{k=-\infty}^{+\infty} [k, k+s] \right) = s, \qquad \text{für alle } 0 \leq s < 1.$$

*Bemerkung.* Teil *(c)* besagt gerade, dass eine Zufallsgröße $X$ genau dann gleichverteilt modulo 1 ist, wenn $P_{\langle X \rangle}$ der Gleichverteilung $\lambda\!\!\lambda_{0,1}$ auf $[0,1)$ entspricht. ♣

Es folgt nun der erste zentrale Satz in diesem Abschnitt, welcher in seiner ursprünglichen Form für Folgen in [8] zu finden ist. Hierbei wird wie im Weiteren auch, die Konvention $\log 0 := 0$ verwendet.

**Satz 2.2** (Gleichverteilungs-Charakterisierung)**.**

(a) *Eine Folge* $(x_n)_{n\in\mathbb{N}}$ *ist genau dann Benford, wenn die Folge* $(\log|x_n|)_{n\in\mathbb{N}}$ *gleichverteilt modulo 1 ist, das heißt wenn*

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid\langle\log|x_n|\rangle\le s\}=s,\qquad \textit{für alle } 0\le s<1.$$

(b) *Eine Borel-messbare Funktion* $f\colon[0,+\infty)\longrightarrow\mathbb{R}$ *ist genau dann Benford, wenn* $\log|f|$ *gleichverteilt modulo 1 ist, das heißt wenn*

$$\lim_{T\to\infty}\frac{1}{T}\lambda\!\!\lambda\left(\{0\le\tau<T\mid\langle\log|f(\tau)|\rangle\le s\}\right)=s,\qquad \textit{für alle } 0\le s<1.$$

(c) *Eine Zufallsgröße* $X$ *auf einem Wahrscheinlichkeitsraum* $(\Omega,\Sigma,P)$ *ist genau dann Benford, wenn* $\log|X|$ *gleichverteilt modulo* 1 *ist, das heißt wenn*

$$P\left(\langle\log|X|\rangle\le s\right)=s,\qquad \textit{für alle } 0\le s<1.$$

(d) *Ein Wahrscheinlichkeitsmaß* $P$ *auf* $(\mathbb{R},\mathfrak{B})$ *ist genau dann Benford, wenn*

$$P\left(\{x\in\mathbb{R}\mid\langle\log|x|\rangle\le s\}\right)=s,\qquad \textit{für alle } 0\le s<1.$$

Beweis. Der Beweis für die verschiedenen Objekte aus *(a)*-*(d)* verläuft vollkommen analog. Daher erfolgt der Beweis nur für Teil *(a)*. Es gilt zunächst für alle $0\le s<1$

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid\langle\log|x_n|\rangle\le s\}\\
&=\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\left\{n\in\mathbb{N}_{1,N}\;\middle|\;\log|x_n|\in\bigcup_{k=-\infty}^{+\infty}[k,k+s]\right\}\\
&=\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\left\{n\in\mathbb{N}_{1,N}\;\middle|\;|x_n|\in\bigcup_{k=-\infty}^{+\infty}10^k\,[1,10^s]\right\}\\
&=\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid S(|x_n|)\le 10^s\}\\
&=\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid S(x_n)\le 10^s\}.
\end{aligned}$$

**„⇒":** Sei $(x_n)_{n\in\mathbb{N}}$ Benford, das heißt $\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid S(x_n)\le 10^s\}=\log 10^s=s$, für alle $0\le s<1$. Dann gilt nach obiger Gleichheit, dass $(\log|x_n|)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist.

**„⇐":** Sei nun $(\log|x_n|)_{n\in\mathbb{N}}$ gleichverteilt modulo 1. Dann gilt also für alle $0\le s<1$, dass $\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid\langle\log|x_n|\rangle\le s\}=s$. Nach obiger Gleichheit gilt daher

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid S(x_n)\le 10^s\}=s,\qquad \text{für alle } 0\le s<1,$$

so dass mit $s:=\log t$ gilt

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{card}\{n\in\mathbb{N}_{1,N}\mid S(x_n)\le t\}=\log t,\qquad \text{für alle } 1\le t<10.$$

Also ist $(x_n)_{n\in\mathbb{N}}$ Benford. □

Die Theorie der Gleichverteilung modulo 1 ist weit entwickelt, so dass es viele verschiedene Kriterien gibt, um eine Folge, eine Funktion, eine Zufallsgröße oder ein Wahrscheinlichkeitsmaß auf Gleichverteilung modulo 1 zu testen. Es folgen nun zunächst einige Resultate, welche hier ohne Beweis erwähnt werden. Für Details siehe [14]:

**Lemma 2.1.**

(a) *Eine Folge* $(x_n)_{n\in\mathbb{N}}$ *ist genau dann gleichverteilt modulo* 1*, wenn für jedes* $k \in \mathbb{Z}\setminus\{0\}$ *und jedes* $b \in \mathbb{R}$ *die Folge* $(k \cdot x_n + b)_{n\in\mathbb{N}}$ *gleichverteilt modulo* 1 *ist.*

(b) *Eine Funktion* $f\colon [0, +\infty) \longrightarrow \mathbb{R}$ *ist genau dann gleichverteilt modulo* 1*, wenn für jedes* $k \in \mathbb{Z}\setminus\{0\}$ *und jedes* $b \in \mathbb{R}$ *die Funktion* $k \cdot f + b$ *gleichverteilt modulo* 1 *ist.*

(c) *Eine Zufallsgröße* $X$ *auf einem Wahrscheinlichkeitsraum* $(\Omega, \Sigma, P)$ *ist genau dann gleichverteilt modulo* 1*, wenn für jedes* $k \in \mathbb{Z}\setminus\{0\}$ *und jedes* $b \in \mathbb{R}$ *die Zufallsgröße* $k \cdot X + b$ *gleichverteilt modulo* 1 *ist.*

**Lemma 2.2.** *Sei* $(x_n)_{n\in\mathbb{N}}$ *eine reelle Zahlenfolge.*

(a) *Falls* $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$*, für ein irrationales* $\theta \in \mathbb{R}$*, so ist* $(x_n)_{n\in\mathbb{N}}$ *gleichverteilt modulo* 1*.*

(b) *Falls* $(x_n)_{n\in\mathbb{N}}$ *periodisch ist, das heißt falls* $x_{n+p} = x_n$ *für ein gewisses* $p \in \mathbb{N}$ *und für alle* $n \in \mathbb{N}$*, so ist die Folge* $(n\theta + x_n)_{n\in\mathbb{N}}$ *genau dann gleichverteilt modulo* 1*, wenn* $\theta$ *irrational ist.*

(c) *Falls die Folge* $(x_n)_{n\in\mathbb{N}}$ *gleichverteilt modulo* 1 *und monoton wachsend ist, so ist die Folge* $\left(\frac{x_n}{\log n}\right)_{n\in\mathbb{N}}$ *unbeschränkt.*

Wir erhalten eine direkte Folgerung aus Lemma 2.1:

**Folgerung 2.1.** *Eine Folge* $(x_n)_{n\in\mathbb{N}}$ *ist genau dann Benford, wenn für jedes* $k \in \mathbb{Z}$ *und jedes* $\alpha \in \mathbb{R}$ *mit* $\alpha \cdot k \neq 0$ *die Folge* $\left(\alpha \cdot x_n^k\right)_{n\in\mathbb{N}}$ *Benford ist.*

Beweis. Im Fall $\alpha = 0$ ist klar, dass die Folge $\left(\alpha \cdot x_n^k\right)_{n\in\mathbb{N}}$ nicht gleichverteilt modulo 1 ist. Für $\alpha \neq 0$ und $x_n \neq 0$ gilt

$$\log\left|\alpha x_n^k\right| = k \cdot \log|x_n| + \log|\alpha|\,.$$

Für jedes $b \in \mathbb{R}$ gibt es dann (mindestens) ein $\alpha \in \mathbb{R}\setminus\{0\}$ mit $b = \log|\alpha|$.

Wenden wir nun Teil *(a)* von Lemma 2.1 an, so ist die Folge $(\log|x_n|)_{n\in\mathbb{N}}$ genau dann gleichverteilt modulo 1, wenn für jedes $k \in \mathbb{Z}\setminus\{0\}$ und jedes $b \in \mathbb{R}$ dann die Folge $(k \cdot \log|x_n| + b)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist. Schreiben wir $b = \log|\alpha|$, so ist $(\log|x_n|)_{n\in\mathbb{N}}$ also genau dann gleichverteilt modulo 1, wenn für jedes $k \in \mathbb{Z}\setminus\{0\}$ und jedes $\alpha \in \mathbb{R}\setminus\{0\}$ die Folge $(k \cdot \log|x_n| + \log|\alpha|)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist. Die Bedingung $k \in \mathbb{Z}\setminus\{0\}$ und $\alpha \neq 0$ ist aber äquivalent zu $\alpha \cdot k \neq 0$. Mit obiger Gleichheit ist dann $(\log|x_n|)_{n\in\mathbb{N}}$ also genau dann gleichverteilt modulo 1, wenn für jedes $k \in \mathbb{Z}$ und jedes $\alpha \in \mathbb{R}$ mit $k \cdot \alpha \neq 0$ gerade $\left(\log\left|\alpha x_n^k\right|\right)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist. Mittels Satz 2.2 folgt dann die Behauptung. □

**Beispiel 2.3.** Wir betrachten die Folge $(2^n)_{n\in\mathbb{N}}$. Wegen

$$\log|2^n| = n \cdot \log 2$$

gilt mit $x_n \coloneqq n \cdot \log 2$

$$x_{n+1} - x_n = (n+1)\log 2 - n\log 2 = \log 2 \in \mathbb{R}\setminus\mathbb{Q}.$$

Daher ist die Folge $(x_n)_{n\in\mathbb{N}}$ nach Teil *(a)* von Lemma 2.2 gleichverteilt modulo 1. Nach Satz 2.2 ist die Folge $(2^n)_{n\in\mathbb{N}}$ also Benford. Etwas allgemeiner ist die Folge $(a^n)_{n\in\mathbb{N}}$ mit $a \neq 0$ Benford, falls $\log|a|$ irrational ist. So sind beispielsweise auch die Folgen $(3^n)_{n\in\mathbb{N}}$, $(\pi^n)_{n\in\mathbb{N}}$ oder $(1/4^n)_{n\in\mathbb{N}}$ Benford.

Zusammenfassend erhalten wir also

**Folgerung 2.2.** *Die Folge* $(a^n)_{n\in\mathbb{N}}$ *ist Benford genau dann, wenn* $\log|a|$ *irrational ist.*

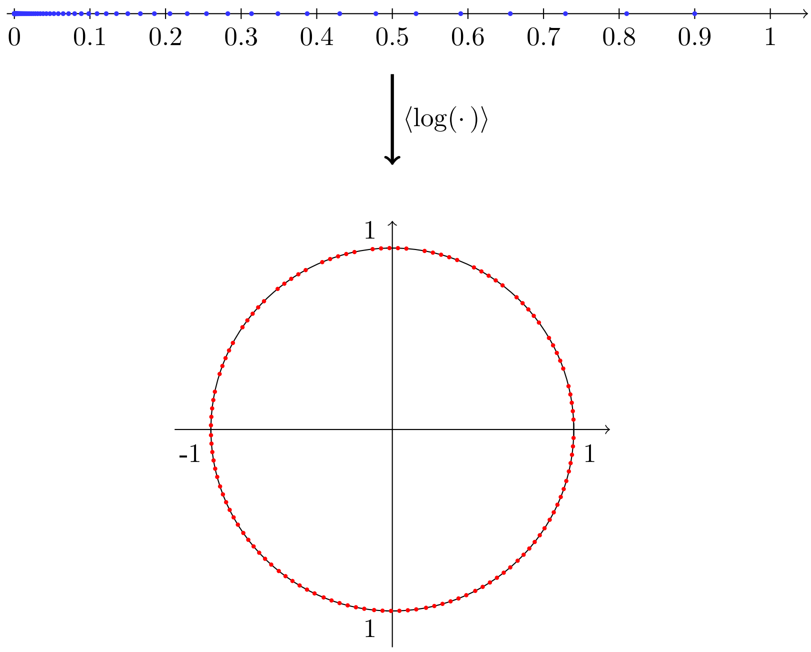


Abbildung 8: Verdeutlichung der Aussage von Satz 2.2 anhand der ersten 120 Folgenglieder der Folge $({}^{9^n}/{}_{10^n})_{n\in\mathbb{N}}$ .

Beweis. Nach Teil *(a)* von Satz 2.2 gilt, dass $(a^n)_{n\in\mathbb{N}}$ genau dann Benford ist, wenn $(n \cdot \log|a|)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist. Nach Teil *(b)* von Lemma 2.2 ist die Folge $(n \cdot \log|a| + 0)_{n\in\mathbb{N}}$ aber genau dann gleichverteilt modulo 1, wenn $\log|a|$ irrational ist. □

Es folgt nun noch ein wichtiges Kriterium für die Gleichverteilung modulo 1:

**Satz 2.3** (Kriterium von Weyl)**.** *Eine Folge $(x_n)_{n\in\mathbb{N}}$ ist gleichverteilt modulo* 1 *genau dann, wenn für jedes $k \in \mathbb{Z}\setminus\{0\}$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k\cdot x_n} = 0.$$

Für einen Beweis siehe [14]. Der nächste Satz basiert auf einem Resultat aus [5]:

**Satz 2.4.** *Seien $\lambda, \mu, x, y \in \mathbb{R}$ mit $\lambda \neq 0$ und $|x| > |y|$. Dann ist die Folge $(\lambda \cdot x^n + \mu \cdot y^n)_{n\in\mathbb{N}}$ genau dann Benford, wenn $\log|x|$ irrational ist.*

Beweis. Da $x \neq 0$ ist und $|x| > |y|$ gilt zunächst

$$\lim_{n\to\infty} \frac{\mu \cdot y^n}{\lambda \cdot x^n} = 0.$$

Wegen

$$\log|\lambda x^n + \mu y^n| - \log|\lambda x^n| = \log\left|1 + \frac{\mu y^n}{\lambda x^n}\right| \xrightarrow{n\to+\infty} 0,$$

ist die Folge $(\log|\lambda x^n + \mu y^n|)_{n\in\mathbb{N}}$ genau dann gleichverteilt modulo 1, wenn es $(\log|\lambda x^n|)_{n\in\mathbb{N}}$ ist. Es genügt also $(\log|\lambda x^n|)_{n\in\mathbb{N}}$ auf Gleichverteilung modulo 1 zu untersuchen. Wir erhalten nun

$$\lim_{n\to\infty} \left(\log\left|\lambda x^{n+1}\right| - \log|\lambda x^n|\right) = \lim_{n\to\infty} \log|x| = \log|x|.$$

Eine Anwendung von Teil *(a)* von Lemma 2.2 liefert nun, dass $(\log|\lambda x^n|)_{n\in\mathbb{N}}$ gleichverteilt modulo 1 ist, falls $\log|x| \in \mathbb{R}\setminus\mathbb{Q}$. Ist $\log|x| \in \mathbb{Q}$, so nimmt

$$\langle \log|\lambda x^n| \rangle = \langle \log|\lambda| + n \cdot \log|x| \rangle$$

nur endlich viele Werte an, womit $(\log|\lambda x^n|)_{n\in\mathbb{N}}$ nicht gleichverteilt modulo 1 sein kann. Die Behauptung folgt somit aus Satz 2.2. $\square$

**Beispiel 2.4** (Fibonacci-Zahlen und Primzahlen)**.**

*(a)* Wir betrachten zunächst die Folge der Fibonacci-Zahlen $(F_n)_{n\in\mathbb{N}}$. Nach der Formel von Binet gilt für jedes $n \in \mathbb{N}$

$$F_n = \frac{1}{\sqrt{5}}\left(\Phi^n - \left(-\frac{1}{\Phi}\right)^n\right), \qquad \text{mit } \Phi = \frac{1+\sqrt{5}}{2}.$$

Wir haben also eine Folge der Form $\left(\frac{1}{\sqrt{5}}\Phi^n - \frac{1}{\sqrt{5}}\left(-1/\Phi\right)^n\right)_{n\in\mathbb{N}}$ mit $|\Phi| > \left|\frac{1}{\Phi}\right|$. Da $\log\Phi$ irrational ist, ergibt eine Anwendung von Satz 2.4, dass $(F_n)_{n\in\mathbb{N}}$ Benford ist.

*(b)* Sei nun $(P_n)_{n\in\mathbb{N}}$ die Folge der Primzahlen. Nach dem Primzahlsatz gilt

$$\lim_{n\to\infty} \frac{P_n}{n\cdot\log n} = 1.$$

Insbesondere gibt es also ein $C > 0$ und ein $N \in \mathbb{N}$, so dass für alle $n \geq N$ gilt

$$P_n \leq C \cdot n\log n.$$

Damit ist

$$\frac{\log P_n}{\log n} \leq 1 + \frac{\log C}{\log n} + \frac{\log(\log n)}{\log n} \xrightarrow{n\to\infty} 1,$$

so dass wegen Teil *(c)* von Lemma 2.2 $(\log P_n)_{n\in\mathbb{N}}$ *nicht* gleichverteilt modulo 1 ist, was aber wegen Satz 2.2 bedeutet, dass $(P_n)_{n\in\mathbb{N}}$ *nicht* Benford ist.

Im Hinblick auf die folgenden Resultate sei an einen wichtigen Konvergenzbegriff für eine Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ erinnert: Sei für jedes $n \in \mathbb{N}$ dann $X_n$ eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega_n, \Sigma_n, P_n)$. Wir sagen die Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ konvergiert *in Verteilung* gegen eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, falls

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t), \qquad \text{für alle } t \in \mathbb{R} \text{ an denen } F_X \text{ stetig ist.}$$

Wir schreiben hierfür auch kurz $X_n \xrightarrow{D} X$. Da $F_{X_n}(t) = P_n(\{X_n \leq t\})$ und $F_X(t) = P(\{X \leq t\})$ lässt sich die Konvergenz in Verteilung auch mithilfe von Wahrscheinlichkeitsmaßen charakterisieren, denn $X_n \xrightarrow{D} X$ genau dann, wenn

$$\lim_{n\to\infty} P_n(\{X_n \leq t\}) = P(\{X \leq t\}), \qquad \text{für alle } t \in \mathbb{R} \text{ mit } P(\{X = t\}) = 0.$$

Für eine Zufallsgröße $Z$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, welche $P$-fast-sicher nur Werte in $[0,1)$ annimmt, so dass die Verteilung $P_Z$ ein Wahrscheinlichkeitsmaß auf $([0,1), \mathfrak{B}[0,1))$ ist, definiert man für $k \in \mathbb{Z}$ den *k-ten Fourier-Koeffizienten* von $P_Z$ als

$$\widehat{P_Z}(k) := \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\, P_Z(\mathrm{d}t) = \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\, F_Z(\mathrm{d}t).$$

Man nennt die komplexe Folge $\left(\widehat{P_Z}(k)\right)_{k\in\mathbb{Z}}$ dann die *Folge der Fourier-Koeffizienten* von $P_Z$. Nach dem Transformationssatz für die Integration bezüglich Bildmaßen gilt zudem für jedes $k \in \mathbb{Z}$

$$\widehat{P}_Z(k) = \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\, P_Z(\mathrm{d}t) = \int_\Omega \mathrm{e}^{\mathrm{i}2\pi kZ}\, \mathrm{d}P.$$

Wegen

$$\left|\widehat{P_Z}(k)\right| \leq \int_0^1 \left|e^{i2\pi kt}\right| P_Z(dt) = 1,$$

ist $\left(\widehat{P}_Z(k)\right)_{k\in\mathbb{Z}}$ beschränkt und zudem gilt stets

$$\widehat{P_Z}(0) = \int_0^1 e^{i2\pi 0t} P_Z(dt) = \int_0^1 P_Z(dt) = 1.$$

Das folgende Lemma beinhaltet nun einige wichtige Eigenschaften der Fourier-Koeffizienten, welche wir für die nachfolgenden Betrachtungen benötigen werden:

**Lemma 2.3.** *Seien $Z$ und $W$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, welche alle $P$-fast-sicher nur Werte in $[0,1)$ annehmen. Dann gilt:*

(a) *Es gilt $P_Z = P_W$ genau dann, wenn für alle $k \in \mathbb{Z}$*

$$\widehat{P_Z}(k) = \widehat{P_W}(k).$$

(b) *Sind $Z$ und $W$ unabhängig, so gilt für alle $k \in \mathbb{Z}$*

$$\widehat{P_{\langle Z+W\rangle}}(k) = \widehat{P_Z}(k) \cdot \widehat{P_W}(k).$$

(c) *Sei für jedes $n \in \mathbb{N}$ nun $Z_n$ eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $\left(\Omega^{(n)}, \Sigma^{(n)}, P^{(n)}\right)$, welche $P^{(n)}$-fast-sicher nur Werte in $[0,1)$ annimmt. Die Folge von Zufallsgrößen $(Z_n)_{n\in\mathbb{N}}$ konvergiert genau dann in Verteilung gegen $Z$, wenn*

$$\lim_{n\to\infty} \widehat{P^{(n)}_{Z_n}}(k) = \widehat{P_Z}(k), \qquad \textit{für alle } k \in \mathbb{Z}.$$

Beweis.

**Zu *(a)*:** Siehe zum Beispiel [2].

**Zu *(b)*:** Wegen der Unabhängigkeit von $Z$ und $W$ und der Messbarkeit von $\cos(2\pi\cdot)$ und $\sin(2\pi\cdot)$ sind für jedes $k \in \mathbb{Z}\setminus\{0\}$ auch $e^{i2\pi kZ} = \cos(2\pi k \cdot Z) + i\sin(2\pi k \cdot Z)$ und $e^{i2\pi kW} = \cos(2\pi k \cdot W) + i\sin(2\pi k \cdot Z)$ unabhängig. Daraus ergibt sich schließlich für $k \in \mathbb{Z}\setminus\{0\}$ mit der Unabhängigkeit

$$\begin{aligned}
\widehat{P_{\langle Z+W\rangle}}(k) &= \int_0^1 e^{i2\pi ks} P_{\langle Z+W\rangle}(ds) = \int_\Omega e^{i2\pi k\langle Z+W\rangle} dP = \int_\Omega e^{i2\pi k(Z+W)} dP \\
&= \int_\Omega e^{i2\pi kZ} e^{i2\pi kW} dP = \int_\Omega e^{i2\pi kZ} dP \cdot \int_\Omega e^{i2\pi kW} dP \\
&= \int_0^1 e^{i2\pi ks} P_Z(ds) \cdot \int_0^1 e^{i2\pi kt} P_W(dt) = \widehat{P}_Z(k) \cdot \widehat{P}_W(k).
\end{aligned}$$

Da $\widehat{P_{\langle Z+W\rangle}}(0) = 1 = 1 \cdot 1 = \widehat{P_Z}(0) \cdot \widehat{P_W}(0)$ gilt die Aussage also für alle $k \in \mathbb{Z}$.

**Zu *(c)*:** Siehe zum Beispiel [7]. □

Es folgt nun ein weiteres Lemma, welches Auskunft über die Unabhängigkeit einer Zufallsgröße von einer konstanten Zufallsgröße gibt:

**Lemma 2.4.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ und $Y : \Omega \longrightarrow \mathbb{R}$ definiert gemäß $Y(\omega) := c$ für ein $c \in \mathbb{R}$. Dann sind $X$ und $Y$ unabhängig.*

Beweis. Es reicht die Unabhängigkeit auf Mengen der Form $\{X \leq x\}$ und $\{Y \leq y\}$ zu überprüfen. Es ist zunächst für $y \in \mathbb{R}$

$$\{Y \leq y\} = \{c \leq y\} = \begin{cases} \varnothing, & \text{falls } c > y, \\ \Omega, & \text{falls } c \leq y. \end{cases}$$

Daher gilt nun für alle $x, y \in \mathbb{R}$

$$\begin{aligned} P(\{X \leq x\} \cap \{Y \leq y\}) &= P(\{X \leq x\} \cap \{c \leq y\}) \\ &= \begin{cases} 0, & \text{falls } c > y, \\ P(\{X \leq x\}), & \text{falls } c \leq y, \end{cases} \\ &= P(\{X \leq x\}) \cdot P(\{Y \leq y\}), \end{aligned}$$

so dass $X$ und $Y$ unabhängig sind. □

Sei nun $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktion $F_P$. Dann werden wir nun zeigen, dass es dann stets eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', \mu)$ gibt, so dass $F_P = F_X$ gilt.

**Lemma 2.5.** *Sei $P$ ein Maß auf $(\mathbb{R}, \mathfrak{B})$ und $F_P \colon \mathbb{R} \longrightarrow [0,1]$ die Verteilungsfunktion von $P$. Dann gibt es eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', \mu)$ mit $F_P = F_X$. Insbesondere gilt also $\mu_X = P$.*

Beweis. Sei $\Omega' \coloneqq (0,1)$, $\Sigma' \coloneqq \mathfrak{B}(0,1)$ und $\mu \coloneqq \lambda_{0,1}$. Definiere dann $F_P^- : \Omega' \longrightarrow \mathbb{R}$ gemäß

$$F_P^-(t) \coloneqq \inf\{x \in \mathbb{R} \mid F_P(x) \geq t\}.$$

Falls $F_P^-(t) \leq y$, so ist also $F_P(y) \geq t$. Ist andererseits nun $F_P(y) \geq t$, so ist $y \in \{x \in \mathbb{R} \mid F_P(x) \geq t\}$, also insbesondere $y \geq \inf\{x \in \mathbb{R} \mid F_P(x) \geq t\}$ und daher $y \geq F_P^-(t)$. Somit ist $F_P^-(t) \leq y$ genau dann, wenn $F_P(y) \geq t$. Damit ist

$$\left\{F_P^- \leq y\right\} = \left\{t \in (0,1) \,\middle|\, F_P^-(t) \leq y\right\} = \{t \in (0,1) \mid t \leq F_P(y)\} = (0, F_P(y)] \cap (0,1).$$

Also ist $F_P^-$ messbar und

$$\mu_{F_P^-}((-\infty, y]) = \mu\left(\left\{F_P^- \leq y\right\}\right) = \lambda_{0,1}\left(\left\{F_P^- \leq y\right\}\right) = \lambda_{0,1}((0, F_P(y)] \cap (0,1)) = F_P(y).$$

Damit ist $X \coloneqq F_P^-$ die behauptete Zufallsgröße auf $(\Omega', \Sigma', \mu)$. □

Ist eine Zufallsgröße $Z$ gleichverteilt auf $[0,1)$, das heißt $P_Z = \lambda_{0,1}$, so sind die Fourier-Koeffizienten besonders einfach, denn

$$\widehat{P_Z}(k) = \int_0^1 e^{i2\pi kt} \, \lambda_{0,1}(dt) = \int_0^1 e^{i2\pi kt} \, dt = \begin{cases} 1, & \text{falls } k = 0, \\ 0, & \text{falls } k \neq 0. \end{cases}$$

Nun folgt ein Resultat, welches Aussagen über die Gleichverteilung modulo 1 für Zufallsgrößen beeinhaltet:

**Satz 2.5.** *Seien $X$ und $Y$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$.*

(a) *Falls $X$ gleichverteilt modulo 1 ist und $Y$ unabhängig von $X$, so ist auch $X + Y$ gleichverteilt modulo 1.*

(b) *Falls $\langle X \rangle$ und $\langle X + \theta \rangle$ mit einem $\theta \in \mathbb{R} \backslash \mathbb{Q}$ die gleiche Verteilung besitzen – das heißt $P_{\langle X \rangle} = P_{\langle X+\theta \rangle}$ – so ist $X$ gleichverteilt modulo 1.*

(c) *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, so dass $X_1$ nicht nur abzählbar viele Werte annimmt*[5]*. Dann gilt*

$$\lim_{n\to\infty} P\left(\left\{\left\langle\sum_{k=1}^{n} X_k\right\rangle \leq s\right\}\right) = s, \qquad \text{für alle } 0 \leq s < 1.$$

*Das heißt es gibt eine Zufallsgröße $Z$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', \mu)$ mit*

$$\left\langle\sum_{k=1}^{n} X_k\right\rangle \xrightarrow[n\to+\infty]{D} Z, \qquad \text{mit } \mu_Z = \lambda\!\!\lambda_{0,1}.$$

Beweis.

**Zu *(a)*:** Für alle $k \in \mathbb{Z}$ gilt wegen Teil *(b)* von Lemma 2.3 wegen der Unabhängigkeit von $X$ und $Y$

$$\widehat{P_{\langle X+Y\rangle}}(k) = \widehat{P_{\langle X\rangle}}(k) \cdot \widehat{P_{\langle Y\rangle}}(k).$$

Da $X$ gleichverteilt modulo 1 ist, das heißt $P_{\langle X\rangle} = \lambda\!\!\lambda_{0,1}$ gilt $\widehat{P_{\langle X\rangle}}(k) = 0$ für alle $k \neq 0$ und damit auch

$$\widehat{P_{\langle X+Y\rangle}}(k) = 0,$$

unabhängig von $\widehat{P_{\langle Y\rangle}}(k)$. Da zudem stets $\widehat{P_{\langle Y\rangle}}(0) = \widehat{P_{\langle X\rangle}}(0) = 1$ gilt, erkennt man, dass insgesamt

$$\widehat{P_{\langle X+Y\rangle}}(k) = \widehat{\lambda\!\!\lambda_{0,1}}(k) = \begin{cases} 1, & \text{falls } k = 0, \\ 0, & \text{falls } k \neq 0, \end{cases}$$

so dass wegen Teil *(a)* von Lemma 2.3 ebenfalls $P_{\langle X+Y\rangle} = \lambda\!\!\lambda_{0,1}$ gilt, was aber gerade heißt, dass $X+Y$ gleichverteilt modulo 1 ist.

**Zu *(b)*:** Falls $\theta$ irrational ist, so ist auch $\langle\theta\rangle$ irrational, so dass wir ohne Beschränkung der Allgemeinheit annehmen können $\theta = \langle\theta\rangle$ – also $0 \leq \theta < 1$. Da die Zufallsgrößen $X$ und $Z\colon \Omega \longrightarrow \mathbb{R}$ mit $Z(\omega) := \theta$ nach Lemma 2.4 unabhängig sind, gilt wieder

$$\widehat{P_{\langle X+\theta\rangle}}(k) = \widehat{P_{\langle X+Z\rangle}}(k) = \widehat{P_{\langle X\rangle}}(k)\,\widehat{P_{\langle Z\rangle}}(k) = \widehat{P_{\langle X\rangle}}(k) \cdot \widehat{\delta_\theta}(k).$$

Wegen $\widehat{\delta_\theta}(k) = \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\,\delta_\theta(\mathrm{d}t) = \mathrm{e}^{\mathrm{i}2\pi k\theta}$ gilt also für alle $k \in \mathbb{Z}$

$$\widehat{P_{\langle X+\theta\rangle}}(k) = \mathrm{e}^{\mathrm{i}2\pi k\theta}\,\widehat{P_{\langle X\rangle}}(k).$$

Wenn $\langle X\rangle$ und $\langle X+\theta\rangle$ die gleiche Verteilung besitzen – also $P_{\langle X\rangle} = P_{\langle X+\theta\rangle}$ – so gilt für alle $k \in \mathbb{Z}$

$$\widehat{P_{\langle X\rangle}}(k) = \widehat{P_{\langle X+\theta\rangle}}(k) = \mathrm{e}^{\mathrm{i}2\pi k\theta}\,\widehat{P_{\langle X\rangle}}(k).$$

Da $\theta$ irrational, ist der Ausdruck $\mathrm{e}^{\mathrm{i}2\pi k\theta}$ für alle $k \neq 0$ ungleich 1, so dass die Gleichheit nur dann gelten kann, wenn für alle $k \neq 0$ gilt

$$\widehat{P_{\langle X\rangle}}(k) = 0.$$

Da zudem stets $\widehat{P_{\langle X\rangle}}(0) = 1$, gilt dann wieder $\widehat{P_{\langle X\rangle}}(k) = \widehat{\lambda\!\!\lambda_{0,1}}(k)$ für alle $k \in \mathbb{Z}$, so dass wegen Teil *(a)* von Lemma 2.3 gelten muss $P_{\langle X\rangle} = \lambda\!\!\lambda_{0,1}$. Also ist $X$ gleichverteilt modulo 1.

**Zu *(c)*:** In Weiterführung der Aussage von Teil *(a)* von Lemma 2.3 gilt für die Fourier-Koeffizienten von $\sum_{k=1}^{n} X_k$ einer unabhängigen Folge

$$\widehat{P_{\langle\sum_{k=1}^{n} X_k\rangle}}(\ell) = \prod_{k=1}^{n} \widehat{P_{\langle X_k\rangle}}(\ell), \qquad \text{für alle } \ell \in \mathbb{Z}.$$

[5]Das heißt $P(\{X_1 \in C\}) < 1$ für jede abzählbare Menge $C \in \mathfrak{B}$.

Da die Folge $(X_n)_{n\in\mathbb{N}}$ zudem identisch verteilt ist, gilt ferner

$$\widehat{P_{\langle \sum_{k=1}^n X_k \rangle}}(\ell) = \prod_{k=1}^n \widehat{P_{\langle X_1 \rangle}}(\ell) = \left[\widehat{P_{\langle X_1 \rangle}}(\ell)\right]^n, \qquad \text{für alle } \ell \in \mathbb{Z}.$$

Da für alle $\ell \in \mathbb{Z}$ dann $\left|\widehat{P_{\langle X_1 \rangle}}(\ell)\right| \leq 1$ ist, gilt

$$\lim_{n\to\infty} \widehat{P_{\langle \sum_{k=1}^n X_k \rangle}}(\ell) = \lim_{n\to\infty} \left[\widehat{P_{\langle X_1 \rangle}}(\ell)\right]^n = 0,$$

falls $\left|\widehat{P_{\langle X_1 \rangle}}(\ell)\right| < 1$ ist. Wir wissen zudem, dass stets $\widehat{P_{\langle X_1 \rangle}}(0) = 1$. Die Frage ist nun, ob es auch möglich ist, dass auch für andere $\ell_0 \neq 0$

$$\left|\widehat{P_{\langle X_1 \rangle}}(\ell_0)\right| = 1$$

gelten kann. Sei also $\ell_0 \in \mathbb{Z}$ mit $\ell_0 \neq 0$ und $\left|\widehat{P_{\langle X_1 \rangle}}(\ell_0)\right| = 1$. Dann gibt es ein $0 \leq \xi < 1$ mit $\widehat{P_{\langle X_1 \rangle}}(\ell_0) = \mathrm{e}^{\mathrm{i}2\pi\xi}$. Damit gilt mittels dem Transformationssatz für Lebesgue-Integrale dann

$$\begin{aligned}
0 &= 1 - \mathrm{e}^{-\mathrm{i}2\pi\xi}\,\widehat{P_{\langle X_1 \rangle}}(\ell_0) = 1 - \int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 t}\,\mathrm{e}^{-\mathrm{i}2\pi\xi}\,P_{\langle X_1 \rangle}(\mathrm{d}t)\\
&= 1 - \int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 (t - \xi/\ell_0)}\,P_{\langle X_1 \rangle}(\mathrm{d}t) = 1 - \int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 \langle t - \xi/\ell_0 \rangle}\,P_{\langle X_1 \rangle}(\mathrm{d}t)\\
&= 1 - \int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 s}\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s).
\end{aligned}$$

Damit muss $\int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 s}\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s) = 1 \in \mathbb{R}$ sein. Damit muss aber der Imaginärteil des Integrals

$$\begin{aligned}
&\int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 s}\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s)\\
&= \int_0^1 \cos(2\pi\ell_0 s)\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s) + \mathrm{i}\int_0^1 \sin(2\pi\ell_0 s)\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s)
\end{aligned}$$

verschwinden, das heißt also $\int_0^1 \sin(2\pi\ell_0 s)\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s) = 0$. Daraus ergibt sich

$$\begin{aligned}
0 = 1 - \int_0^1 \mathrm{e}^{\mathrm{i}2\pi\ell_0 s}\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s) &= 1 - \int_0^1 \cos(2\pi\ell_0 s)\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s)\\
&= \int_0^1 \underbrace{[1 - \cos(2\pi\ell_0 s)]}_{\geq 0}\,P_{\langle X_1 - \xi/\ell_0 \rangle}(\mathrm{d}s)\\
&= \int_\Omega \underbrace{[1 - \cos(2\pi\ell_0 \langle X_1 - \xi/\ell_0 \rangle)]}_{\geq 0}\,\mathrm{d}P,
\end{aligned}$$

wobei im letzten Satz der Transformationssatz für Lebesgue-Integrale angewandt wurde. Die Gleichheit kann aber nur gelten, falls für $P$-fast-alle $\omega \in \Omega$ gilt

$$\begin{aligned}
1 &= \cos(2\pi\ell_0 \langle X_1(\omega) - \xi/\ell_0 \rangle) = \cos(2\pi\ell_0 (X_1(\omega) - \xi/\ell_0 - \lfloor X_1(\omega) - \xi/\ell_0 \rfloor))\\
&= \cos(2\pi(\ell_0 X_1(\omega) - \xi)).
\end{aligned}$$

Das bedeutet aber $P$-fast-sicher $\ell_0 X_1 = \xi + k$ für ein $k \in \mathbb{Z}$. Damit ist

$$P\left(\{\ell_0 X_1 \in \xi + \mathbb{Z}\}\right) \geq P\left(\{\ell_0 X_1 = \xi + k\}\right) = 1.$$

Das heißt $P\left(\{\ell_0 X_1 \in \xi + \mathbb{Z}\}\right) = 1$ und $X_1$ nimmt insbesondere nur abzählbar viele Werte an, was einen Widerspruch zur Voraussetzung ist. Also kann es so ein $\ell_0 \neq 0$ nicht geben und es folgt

$$\lim_{n\to\infty} \widehat{P_{\langle \sum_{k=1}^n X_k \rangle}}(\ell) = \lim_{n\to\infty} \left[\widehat{P_{\langle X_1 \rangle}}(\ell)\right]^n = \widehat{\lambda\!\!\lambda_{0,1}}(\ell), \qquad \text{für alle } \ell \in \mathbb{Z}.$$

Nach Teil *(c)* von Lemma 2.3 gilt also

$$\lim_{n\to\infty} P\left(\left\{\left\langle \sum_{k=1}^n X_k \right\rangle \leq s\right\}\right) = s, \qquad \text{für alle } 0 \leq s < 1.$$

Nach Lemma 2.5 können wir nun eine Zufallsgröße $Z$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', \mu)$ wählen, so dass $\mu_Z = \lambda\!\!\lambda_{0,1}$, so dass also nach der Definition der Konvergenz in Verteilung

$$\left\langle \sum_{k=1}^n X_k \right\rangle \xrightarrow[n\to+\infty]{D} Z.$$

□

Mithilfe von Fourier-Koeffizienten kann man die Benford-Eigenschaft von Zufallsgrößen leicht charakterisieren: Eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ ist genau dann Benford, wenn

$$\widehat{P_{\langle \log|X| \rangle}}(k) = \begin{cases} 1, & \text{falls } k = 0, \\ 0, & \text{falls } k \neq 0. \end{cases}$$

Dann ist nämlich $\log|X|$ gleichverteilt modulo 1, was wegen Satz 2.2 äquivalent zur Benford-Eigenschaft von $X$ ist.

Im Hinblick auf das Grenzverhalten von Folgen von Zufallsgrößen, welches wir in Kapitel 3 näher betrachten, tätigen wir noch folgende

**Definition 2.7.** Eine Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung gegen das Benfordsche Gesetz*, falls

$$\lim_{n\to\infty} P\left(\{S(X_n) \leq t\}\right) = \log t, \qquad \text{für alle } 1 \leq t < 10.$$

*Bemerkung.* Die Definition besagt nun, dass keine der Zufallsgrößen $X_n$ Benford sein muss, jedoch die Zufallsgrößen $S(X_n)$ in Verteilung gegen $S(X)$ konvergieren, wobei $X$ eine Zufallsgröße ist, welche Benford ist. ♣

Mit dieser Begriffsbildung erhalten wir als direkte Folgerung aus Satz 2.5:

**Satz 2.6.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig identisch verteilter Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, so dass $X_1$ nicht nur abzählbar viele Werte annimmt. Dann konvergiert die Folge $\left(\prod_{k=1}^n X_k\right)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz, das heißt*

$$\lim_{n\to\infty} P\left(\left\{S\left(\prod_{k=1}^n X_k\right) \leq t\right\}\right) = \log t.$$

Beweis. Da $\log|\cdot| : \mathbb{R} \longrightarrow \mathbb{R}$ eine messbare Funktion ist, ist auch die Folge $(\log|X_n|)_{n\in\mathbb{N}}$ unabhängig und identisch verteilt. Natürlich nimmt dann auch $\log|X_1|$ nicht nur abzählbar viele Werte an. Nach Teil *(c)* von Satz 2.5 gilt dann aber

$$P\left(\left\{\left\langle \log \prod_{k=1}^n |X_k| \right\rangle \leq s\right\}\right) = P\left(\left\{\left\langle \sum_{k=1}^n \log|X_k| \right\rangle \leq s\right\}\right) \xrightarrow{n\to\infty} s, \quad \text{für alle } 0 \leq s < 1,$$

so dass $\left(\langle \log \prod_{k=1}^{n} |X_k| \rangle\right)_{n\in\mathbb{N}}$ in Verteilung gegen eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', \mu)$ konvergiert mit $\mu_X = \lambda_{0,1}$. Nach Definition 2.6 ist $X$ dann gleichverteilt modulo 1. Also konvergiert die Folge $\left(\langle \log \prod_{k=1}^{n} |X_k| \rangle\right)_{n\in\mathbb{N}}$ in Verteilung gegen eine Zufallsgröße, welche gleichverteilt modulo 1 ist, was wegen Satz 2.2 und Definition 2.7 heißt, dass $\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz konvergiert. □

*Bemerkung.* In Satz 3.9 werden wir noch sehen, dass mit denselben Voraussetzungen von Satz 2.6 sogar

$$P\left(\left\{\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1$$

gilt. Das heißt für $P$-fast-alle $\omega \in \Omega$ ist die Zahlenfolge $\left(\log \left|\prod_{k=1}^{n} X_k(\omega)\right|\right)_{n\in\mathbb{N}}$ gleichverteilt modulo 1, was in Hinblick auf Teil *(a)* von Satz 2.2 äquivalent dazu ist, dass die Folge $\left(\prod_{k=1}^{n} X_k(\omega)\right)_{n\in\mathbb{N}}$ für $P$-fast-alle $\omega \in \Omega$ Benford ist. ♣

**Beispiel 2.5.**

*(a)* Sei $X$ eine exponentialverteilte Zufallsgröße mit $\mathrm{E}[X] = 1$, das heißt

$$F_X(t) = \mathbf{1}_{[0,+\infty)}\left[1 - \mathrm{e}^{-t}\right], \qquad \text{für } t \in \mathbb{R}.$$

Dann ist für $1 \leq t < 10$

$$\begin{aligned} P(\{S(X) \leq t\}) &= P\left(\left\{X \in \bigcup_{k=-\infty}^{+\infty} 10^k [1,t]\right\}\right) = \sum_{k=-\infty}^{+\infty} P_X\left(10^k [1,t]\right) \\ &= \sum_{k=-\infty}^{+\infty} \left[F_X\left(10^k t\right) - F_X\left(10^k\right)\right] = \sum_{k=-\infty}^{+\infty} \left[\mathrm{e}^{-10^k} - \mathrm{e}^{-10^k t}\right] \\ &\not\equiv \log t, \end{aligned}$$

so dass $X$ *nicht* Benford ist. Betrachtet man jedoch eine Folge von unabhängigen exponentialverteilten Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ mit $\mathrm{E}[X_n] = \mathrm{E}[X_1] = 1$, so liefert Satz 2.6, dass jedoch

$$\lim_{n\to\infty} P\left(\left\{S\left(\prod_{k=1}^{n} X_k\right) \leq t\right\}\right) = \log t,$$

womit $\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz konvergiert.

*(b)* Sei $X$ eine standardnormalverteilte Zufallsgröße, das heißt

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \qquad \text{für } x \in \mathbb{R}.$$

Genau wie in *(b)* erhält man durch Nachrechnen

$$P(\{S(X) \leq t\}) = \sum_{k=-\infty}^{+\infty} \left[\Phi\left(10^k t\right) - \Phi\left(10^k\right)\right] \not\equiv \log t,$$

wobei $\Phi(s) := \int_{-\infty}^{s} f_X(x)\, \lambda(\mathrm{d}x)$ die Verteilungsfunktion der Standardnormalverteilung sei. Man schlussfolgert nun also wieder, dass $X$ *nicht* Benford ist. Betrachtet man aber wieder eine Folge von unabhängigen standardnormalverteilten Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ und wendet Satz 2.6 an, so konvergiert aber $\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz.

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsgrößen. Die Voraussetzung, dass $X_1$ nicht nur abzählbar viele Werte annimmt, ist zum Beispiel immer dann erfüllt, wenn $X_1$ eine Dichte $f_{X_1}$ besitzt, denn dann ist für jedes abzählbare $C = \bigcup_{n=1}^{\infty} \{c_n\} \in \mathfrak{B}$ nämlich

$$P(\{X_1 \in C\}) = \int\limits_C f_{X_1} \mathrm{d}\lambda\!\!\lambda = \sum_{n=1}^{\infty} \int\limits_{\{c_n\}} f_{X_1} \mathrm{d}\lambda\!\!\lambda = \sum_{n=1}^{\infty} 0 = 0 < 1.$$

Etwas allgemeiner ist die Voraussetzung aus Satz 2.6 auch dann erfüllt, wenn die Verteilungsfunktion $F_{X_1}$ linkseitig stetig auf $\mathbb{R}$ ist. Denn dann gilt

$$\begin{aligned} P(\{X_1 \in C\}) &= P\left(\left\{X_1 \in \bigcup_{n=1}^{\infty} \{c_n\}\right\}\right) = \sum_{n=1}^{\infty} P(\{X_1 = c_n\}) \\ &= \sum_{n=1}^{\infty} P(\{X_1 \leq c_n\} \setminus \{X_1 < c_n\}) = \sum_{n=1}^{\infty} \left(P(\{X_1 \leq c_n\}) - P(\{X_1 < c_n\})\right) \\ &= \sum_{n=1}^{\infty} \left(F_{X_1}(c_n) - \lim_{x \to c_n - 0} F_{X_1}(c_n)\right) = \sum_{n=1}^{\infty} \left(F_{X_1}(c_n) - F_{X_1}(c_n)\right) = 0 < 1. \end{aligned}$$

## 2.3 Skaleninvarianz-Charakterisierung

Wie bereits von Benford selbst beobachtet, tritt das Benfordsche Gesetz in den unterschiedlichsten Bereichen – das heißt auch auf unterschiedlichen Skalen auf. Sollte es sich beim Benfordschen Gesetz wirklich um ein „Naturgesetz" – im Sinne von einer Eigenschaft des Dezimalsystems – handeln, so sollte es unabhängig von der Skala, das heißt unabhängig von Maßeinheiten gelten. Genügt also zum Beispiel ein Datensatz in Metern dem Benfordschen Gesetz, so sollte dieser dann auch umgerechnet in Fuß oder Meilen dem Benfordschen Gesetz genügen.

Wir werden nun zeigen, dass dies in geeigneter Formulierung eine Charakterisierung der Benford-Eigenschaft ist. Jedoch kommt es hierbei wesentlich auf die richtige Formulierung an: Betrachten wir zum Beispiel den messbaren Raum $(\mathbb{R}^+, \mathfrak{B}^+)$ und ein Wahrscheinlichkeitsmaß $P$ auf diesem. Dann könnte man $P$ skaleninvariant nennen, wenn

$$P(\alpha B) = P(B), \qquad \text{für jedes } B \in \mathfrak{B}^+ \text{ und jedes } \alpha > 0 \tag{2.1}$$

gilt. Angenommen es gäbe so ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathfrak{B}^+)$. Dann bezeichne $F_P \colon \mathbb{R} \longrightarrow [0,1]$ die Verteilungsfunktion

$$F_P(x) = P((0,x]).$$

Aus der Skaleninvarianz (2.1) folgt dann

$$F_P(1) = P((0,1]) = P(\alpha(0,1]) = P((0,\alpha]) = F_P(\alpha), \qquad \text{für alle } \alpha > 0.$$

Das heißt $F_P$ ist auf ganz $\mathbb{R}^+$ konstant. Da aber $F_P(0) = 0$ und $\lim_{x \to +\infty} F_P(x) = P(\mathbb{R}^+) = 1$ kann es ein solches $P$ nicht geben. Es wird sich zeigen, dass diese Bedingung nicht für *alle* $B \in \mathfrak{B}^+$ gefordert werden darf.

Genauso könnte man auf die Idee kommen eine Zufallsgröße $X$ auf $(\Omega, \Sigma, P)$, welche $P$-fast-sicher nur Werte in $\mathbb{R}^+$ annimmt, skaleninvariant zu nennen, wenn $X$ und $\alpha X$ für alle $\alpha > 0$ die selbe Verteilung besitzen, das heißt wenn

$$P_{\alpha X} = P_X, \qquad \text{für alle } \alpha > 0. \tag{2.2}$$

Wir erkennen sofort, dass dann die Zufallsgröße $\overline{X}$, welche $P$-fast-sicher konstant 0 ist, skaleninvariant ist, denn $\alpha\overline{X}$ ist dann für jedes $\alpha > 0$ ebenfalls $P$-fast-sicher konstant 0. Jedoch ist dann $\overline{X}$ auch die einzige Zufallsgröße, welche gemäß (2.2) skaleninvariant ist. Denn angenommen es gibt eine Zufallsgröße $X$, welche (2.2) erfüllt und nicht $P$-fast-überall 0 ist. Dann gibt es ein $M > 0$ und ein $p > 0$, so dass

$$P\left(\{|X| > M\}\right) = p > 0.$$

Damit gilt dann für $\alpha > 0$ auch $P\left(\{|\alpha X| > M\}\right) = P\left(\{|X| > M/\alpha\}\right)$ und wegen der Oberhalbstetigkeit[6] von $P$ gilt dann

$$\lim_{\alpha \to 0+0} P\left(\left\{|X| > \frac{M}{\alpha}\right\}\right) = P\left(\{|X| = +\infty\}\right) = 0.$$

Somit gilt für hinreichend kleines $\alpha > 0$ dann

$$P\left(\{|\alpha X| > M\}\right) < p = P\left(\{|X| > M\}\right),$$

so dass $X$ nicht (2.2) erfüllen kann. Somit gibt es keine Zufallsgröße außer jene die $P$-fast-überall 0 ist, welche gemäß (2.2) skaleninvariant ist.

Definieren wir nun aber $\mu$ auf $(\mathbb{R}^+, \mathfrak{B}^+)$ durch

$$\mu\left([a,b]\right) := \int_a^b \frac{\lambda\!\!\lambda\,(\mathrm{d}t)}{t} = \ln \frac{b}{a}, \qquad \text{für alle } [a,b] \subseteq \mathbb{R}^+ \text{ mit } a < b.$$

Dann gilt für jedes $\alpha > 0$

$$\mu\left(\alpha\,[a,b]\right) = \mu\left([\alpha a, \alpha b]\right) = \int_{\alpha a}^{\alpha b} \frac{\mathrm{d}t}{t} = \ln b + \ln \alpha - (\ln a + \ln \alpha) = \ln \frac{b}{a}, \qquad \text{für alle } [a,b] \subseteq \mathbb{R}^+,$$

so dass $\mu$ gemäß (2.1) – in voller Übereinstimmung mit dem bereits Gezeigten – skaleninvariant ist. $\mu$ ist nämlich kein endliches Maß, denn $\mu\left(\mathbb{R}^+\right) = +\infty$.

Wir erkennen, dass Wahrscheinlichkeitsmaße und Zufallsgrößen *nicht* skaleninvariant gemäß (2.1) und (2.2) sein können. Trotzdem können Wahrscheinlichkeitsmaße und Zufallsgrößen skaleninvariante *signifikante Dezimalziffern* besitzen:

**Definition 2.8.** Sei $\mathfrak{A}$ eine $\sigma$-Algebra auf $\mathbb{R}^+$ mit $\mathfrak{A} \supseteq \mathcal{S}$. Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \mathfrak{A})$ hat *skaleninvariante signifikante Dezimalziffern*, wenn

$$P\left(\alpha A\right) = P\left(A\right), \qquad \text{für alle } A \in \mathcal{S} \text{ und alle } \alpha > 0.$$

Zunächst ist für alle $\alpha > 0$ dann $\alpha A$ nach Lemma 1.2 wieder in $\mathcal{S}$, so dass die obige Definition sinnvoll ist. Bis jetzt ist aber nicht klar, ob es überhaupt ein Wahrscheinlichkeitsmaß gibt, welches diese Eigenschaft besitzen kann. Sei nun $\mathbb{B}$ die Benford-Verteilung auf $(\mathbb{R}^+, \mathcal{S})$. Dann gilt für $A = \bigcup_{k=-\infty}^{+\infty} 10^k\,[a,b] \in \mathcal{S}$ mit $1 \le a < b < 10$

$$\alpha A = \bigcup_{k=-\infty}^{+\infty} 10^{k+\log \alpha}\,[a,b] = \bigcup_{k=-\infty}^{+\infty} 10^{k+\langle \log \alpha \rangle}\,[a,b] = \bigcup_{k=-\infty}^{+\infty} 10^k C,$$

wobei sich durch Nachrechnen zeigt, dass

$$C = \begin{cases} \left[10^{\langle \log \alpha \rangle} a, 10^{\langle \log \alpha \rangle} b\right], & \text{falls } 0 \le \langle \log \alpha \rangle < 1 - \log b, \\ \left[1, 10^{\langle \log \alpha \rangle - 1} b\right] \cup \left[10^{\langle \log \alpha \rangle} a, 10\right), & \text{falls } 1 - \log b \le \langle \log \alpha \rangle < 1 - \log a, \\ \left[10^{\langle \log \alpha \rangle - 1} a, 10^{\langle \log \alpha \rangle - 1} b\right], & \text{falls } 1 - \log a \le \langle \log \alpha \rangle < 1. \end{cases}$$

[6] Ein Maß $\mu$ auf einem messbaren Raum $(\Omega, \Sigma)$ heißt oberhalbstetig, wenn für jede antitone Mengenfolge $(A_n)_{n \in \mathbb{N}}$ aus $\Sigma$ dann $\lim_{n\to\infty} \mu\left(A_n\right) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right)$.

Damit ist zunächst $C \in \mathfrak{B}[1,10)$. Nach Definition 2.5 gilt zudem für alle $1 \leq s < t < 10$

$$\mathbb{B}\left(\bigcup_{k=-\infty}^{+\infty} 10^k [s,t]\right) = \log t - \log s.$$

Hieraus ergibt sich

$$\begin{aligned}
&\mathbb{B}(\alpha A) \\
=&\mathbb{B}\left(\bigcup_{k=-\infty}^{+\infty} 10^k C\right) \\
=&\begin{cases} \log\left(10^{\langle\log\alpha\rangle}b\right) - \log\left(10^{\langle\log\alpha\rangle}a\right), & \text{falls } 0 \leq \langle\log\alpha\rangle < 1-\log b, \\ \log 10 - \log\left(10^{\langle\log\alpha\rangle}a\right) + \log\left(10^{\langle\log\alpha\rangle-1}b\right), & \text{falls } 1-\log b \leq \langle\log\alpha\rangle < 1-\log a, \\ \log\left(10^{\langle\log\alpha\rangle-1}b\right) - \log\left(10^{\langle\log\alpha\rangle-1}a\right), & \text{falls } 1-\log a \leq \langle\log\alpha\rangle < 1 \end{cases} \\
=&\log b - \log a,
\end{aligned}$$

so dass $\mathbb{B}$ skaleninvariante signifikante Dezimalziffern hat.

Der folgende Satz zeigt nun, dass die Benford-Verteilung $\mathbb{B}$ auch das *einzige* Wahrscheinlichkeitsmaß auf $(\mathbb{R}^+, \mathcal{S})$ ist, welches diese Eigenschaft besitzt:

**Satz 2.7** (Skaleninvarianz-Charakterisierung)**.** *Sei $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^+, \mathfrak{A})$ mit $\mathfrak{A} \supseteq \mathcal{S}$. Dann sind folgende Aussagen äquivalent:*

(i) *$P$ hat skaleninvariante signifikante Dezimalziffern.*

(ii) *Für alle $A \in \mathcal{S}$ gilt $P(A) = \mathbb{B}(A)$, das heißt $P$ ist Benford.*

Beweis. Sei $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^+, \mathfrak{A})$ und bezeichne $P^*$ die Einschränkung von $P$ auf $(\mathbb{R}^+, \mathcal{S})$. Es sei an die Funktion $L\colon \mathbb{R}^+ \longrightarrow [0,1)$ definiert gemäß $L(x) \coloneqq \log S(x)$ aus Satz 1.3 erinnert. Es bezeichne nun $Q$ das Bildmaß von $P^*$ unter $L$, das heißt

$$Q \coloneqq P^*_L.$$

Sei nun $A = \bigcup_{k=-\infty}^{+\infty} 10^k B' \in \mathcal{S}$ mit $B' \in \mathfrak{B}[1,10)$ und $\alpha > 0$. Wie im Beweis von Satz 1.3 gezeigt, gibt es dann genau ein $B \in \mathfrak{B}[0,1)$ mit $A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B$. Es gilt zunächst

$$\begin{aligned}
\alpha A &= \bigcup_{k=-\infty}^{+\infty} 10^{k+\log\alpha} 10^B = \bigcup_{k=-\infty}^{+\infty} 10^k 10^{\langle\log\alpha\rangle} 10^B \\
&= \bigcup_{k=-\infty}^{+\infty} 10^k 10^{\langle\log\alpha\rangle + B},
\end{aligned}$$

wobei $\langle\log\alpha\rangle + B \coloneqq \{\langle\log\alpha\rangle + b \,|\, b \in B\}$[7]. Sei nun $t \coloneqq \langle\log\alpha\rangle \in [0,1)$. Damit gilt also

$$\begin{aligned}
\alpha A &= \bigcup_{k=-\infty}^{+\infty} 10^k 10^{t+B} = \bigcup_{k=-\infty}^{+\infty} 10^k 10^{\lfloor t+B\rfloor + \langle t+B\rangle} = \bigcup_{k=-\infty}^{+\infty} 10^{k+\lfloor t+B\rfloor} 10^{\langle t+B\rangle} \\
&= \bigcup_{\ell=-\infty}^{+\infty} 10^{\ell} 10^{\langle t+B\rangle}.
\end{aligned}$$

Die Aussage *(i)* bedeutet gerade

$$P^*(A) = P^*(\alpha A), \qquad \text{für alle } \alpha > 0 \text{ und alle } A \in \mathcal{S}.$$

[7] Im Allgemeinen sei für eine reelle Zahl $x$ und eine Menge $A \subseteq \mathbb{R}$ dann $\langle x + A\rangle \coloneqq \{\langle x+a\rangle \,|\, a \in A\}$.

Dies ist nun unter Ausnutzung von $A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B$ und $\alpha A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^{\langle t+B \rangle}$ unter der in Satz 1.3 eingeführten Korrespondenz[8] äquivalent zu

$$Q(B) = Q(\langle t+B \rangle), \qquad \text{für alle } t \in [0,1) \text{ und alle } B \in \mathfrak{B}[0,1). \tag{2.3}$$

Nach Lemma 2.5 können wir nun eine Zufallsgröße $X$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mu)$ wählen, deren Verteilung gerade $Q$ ist, das heißt $\mu_X = Q$. Da dann $X$ $\mu$-fast-sicher nur Werte in $[0,1)$ annimmt, ist dann insbesondere $Q = \mu_X = \mu_{\langle X \rangle}$. Dann ist (2.3) wiederum äquivalent dazu, dass $\langle X \rangle$ und $\langle X + t \rangle$ für alle $t \in [0,1)$ dieselbe Verteilung besitzen. Nach Teil *(b)* von Satz 2.5 folgt daraus, dass $X$ gleichverteilt modulo 1 ist. Ist umgekehrt $X$ gleichverteilt modulo 1, so ist nach Teil *(a)* von Satz 2.5 dann wegen der Unabhängigkeit von $X$ und $t$ dann $\langle X + t \rangle$ gleichverteilt modulo 1, da $X$ und $t$ wegen Lemma 2.4 unabhängig sind. Das heißt insgesamt ist (2.3) äquivalent dazu, dass die Zufallsgröße $X$ gleichverteilt modulo 1 ist. Dies bedeutet aber

$$Q = \mu_{\langle X \rangle} = \lambda\!\!\lambda_{0,1}.$$

Es gilt nun aber auch $Q = P_L^*$. Wegen (1.4) gilt auch für die Benford-Verteilung $\mathbb{B}_L = \lambda\!\!\lambda_{0,1}$. Das heißt wir haben $P_L^* = \mathbb{B}_L$ – es gilt also nach Satz 1.3

$$P^* = \mathbb{B},$$

so dass für alle $A \in \mathcal{S}$ gilt

$$P(A) = P^*(A) = \mathbb{B}(A). \qquad \square$$

## 2.4 Spieltheoretische Charakterisierung

Wir werden nun eine Anwendung der Theorie über das Benfordsche Gesetz und zugleich eine weitere interessante Charakterisierung der Benford-Verteilung kennenlernen. Dabei handelt es sich um das sogenannte *Multiplikationsspiel* oder auch *multiplication game*. Bei der Untersuchung dieses einfachen Spiels werden wir dann in natürlicher Art und Weise auf die Benford-Verteilung $\mathbb{B}$ stoßen:

**Regeln des Multiplikationsspiels.** Wir haben zwei Spieler – im Weiteren stets Spieler $A$ und Spieler $B$ genannt. Die beiden Spieler wählen unabhängig voneinander eine natürliche Zahl. Spieler $A$ gewinnt, wenn das Produkt der beiden Zahlen mit den Ziffern 1, 2 oder 3 anfängt – andernfalls gewinnt Spieler $B$.

Wir formalisieren dieses Spiel zunächst, so dass es sich einfacher mathematisch behandeln lässt. Da jeder Spieler eine natürliche Zahl wählt, können wir die Ergebnisse dieser Auswahl als Tupel $(i,j) \in \mathbb{N} \times \mathbb{N}$ schreiben. Die Ergebnisse des Spiels – das heißt gewinnt $A$ oder gewinnt $B$ – können wir demzufolge in eine unendliche Matrix $\mathbf{M} = (a_{ij})_{i,j \in \mathbb{N}}$ mit

$$a_{ij} := \begin{cases} 1, & \text{falls } i \cdot j \text{ mit } 1,2 \text{ oder } 3 \text{ beginnt,} \\ 0, & \text{sonst,} \end{cases}$$

schreiben. Das heißt, wählt $A$ die Zahl $i$ und $B$ die Zahl $j$, so steht eine 1 in der $i$-ten Zeile und $j$-ten Spalte, falls $A$ gewinnt und ansonsten eine 0. $A$ wird nun mit einer gewissen Wahrscheinlichkeit $p_i$ die Zahl $i$ auswählen, während $B$ mit der Wahrscheinlichkeit $q_j$ die Zahl $j$ auswählt. Damit erhalten wir die Folgen

$$P = (p_1, p_2, \ldots) \in [0,1]^{\mathbb{N}} \qquad \text{und} \qquad Q = (q_1, q_2, \ldots) \in [0,1]^{\mathbb{N}},$$

[8] Dabei wurden Wahrscheinlichkeitsmaße $\mu$ auf $(\mathbb{R}^+, \mathcal{S})$ und $\nu$ auf $([0,1), \mathfrak{B}[0,1))$ eindeutig miteinander identifiziert, falls für alle $A = \bigcup_{k=-\infty}^{+\infty} 10^k 10^B$ dann $\mu(A) = \nu(B)$ gilt.

mit

$$\sum_{i=1}^{\infty} p_i = 1 \qquad \text{und} \qquad \sum_{j=1}^{\infty} q_j = 1,$$

welche dann als Wahrscheinlichkeitsmaße auf $(\mathbb{N}, \mathfrak{P}(\mathbb{N}))$ aufgefasst werden können. Eine Strategie für Spieler $A$ in diesem Spiel ist also nichts weiter als ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{N}, \mathfrak{P}(\mathbb{N}))$, wobei $P(\{i\}) = p_i$ die Wahrscheinlichkeit angibt, dass $A$ die natürliche Zahl $i$ wählt und analog für Spieler $B$ mit einem Wahrscheinlichkeitsmaß $Q$. Insofern bezeichnen wir in diesem Abschnitt Wahrscheinlichkeitsmaße auch als Strategien. Bezeichnen wir nun mit $w(P, Q)$ die Wahrscheinlichkeit, dass $A$ gewinnt, wenn $A$ die Strategie $P$ wählt und $B$ die Strategie $Q$, so gilt offenbar[9]

$$w(P, Q) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_i q_j a_{ij}.$$

Wir bezeichnen im Weiteren mit $\mathfrak{W}$ die Menge aller Strategien. Um möglichst oft zu gewinnen wird $A$ nun versuchen $P$ so zu wählen, dass egal wie $B$ seine Strategie $Q$ wählt, der Wert $w(P, Q)$ möglichst groß wird. Da $A$ davon ausgeht, dass $B$ rational handelt, wird er versuchen den „worst-case-Fall" zu maximieren. Spieler $A$ sucht also den Wert

$$V_A \coloneqq \sup_{P \in \mathfrak{W}} \inf_{Q \in \mathfrak{W}} w(P, Q).$$

Umgekehrt wird Spieler $B$ versuchen den seine Strategie so zu wählen, dass egal wie $A$ die Strategie $P$ wählt der Wert $w(P, Q)$ möglichst klein wird. Da auch $B$ davon ausgeht, dass $A$ rational handelt, wird er ebenfalls versuchen den für ihn schlechtesten Fall zu minimieren, also sucht $B$ den Wert

$$V_B \coloneqq \inf_{Q \in \mathfrak{W}} \sup_{P \in \mathfrak{W}} w(P, Q).$$

Es stellt sich nun heraus, dass zwischen $V_A$ und $V_B$ folgender Zusammenhang besteht:

**Lemma 2.6.** *Es gilt stets $V_A \leq V_B$.*

Beweis. Für alle $P_0 \in \mathfrak{W}$ gilt offenbar

$$\inf_{Q \in \mathfrak{W}} w(P_0, Q) \leq \inf_{Q \in \mathfrak{W}} \sup_{P \in \mathfrak{W}} w(P, Q).$$

Bilden wir das Supremum über $P \in \mathfrak{W}$ auf der linken Seite, so erhalten wir also

$$V_A = \sup_{P \in \mathfrak{W}} \inf_{Q \in \mathfrak{W}} w(P, Q) \leq \inf_{Q \in \mathfrak{W}} \sup_{P \in \mathfrak{W}} w(P, Q) = V_B.$$

□

Unser Ziel ist nun zu zeigen, dass $V_A = V_B$ ist und dann optimale Strategien $P^*$ und $Q^*$ für $A$ und $B$ zu finden. Dabei nennen wir Strategien $P^*$ und $Q^*$ *optimal*, falls $V_A = V_B = w(P^*, Q^*)$. Dazu modifizieren wir das Spiel noch etwas. Dazu stellen wir zunächst fest: Wählt $A$ die Zahl $i$ und $B$ die Zahl $j$ und beginnt das Produkt $i \cdot j$ mit der Ziffer $d$, so beginnt auch $(i10^n) \cdot (j10^m)$ für alle $n, m \in \mathbb{N}$ mit der Ziffer $d$, das heißt $A$ hätte auch die Zahl $i \cdot 10^n$ und $B$ die Zahl $j \cdot 10^m$ wählen können ohne das Ergebnis zu verändern. Somit sind die zugehörigen Einträge in der Matrix $\mathbf{M}$ redundant. Ferner erkennt man leicht, dass es *nicht* notwendig ist natürliche Zahlen zu wählen, denn wählt $A$ die Zahl $i = i_1 i_2 i_3 \ldots$ und $B$ die Zahl $j = j_1 j_2 j_3 \ldots$, so hat das Produkt

$$\underbrace{(i_1, i_2 i_3 \ldots)}_{\in [1,10)} \cdot \underbrace{(j_1, j_2 j_3 \ldots)}_{\in [1,10)}$$

dieselbe Anfangsziffer wie das ursprüngliche Produkt $i \cdot j$. Dies entspricht gerade den Übergang zur Signifikanten $S(i \cdot j)$.

[9] Die Wahrscheinlichkeit, dass $B$ gewinnt, falls $A$ Strategie $P$ wählt und $B$ Strategie $Q$ ist dann $1 - w(P, Q)$.

Wir können das Multiplikationsspiel also derart modifizieren, dass wir statt natürlichen Zahlen nun sogar *reelle* Zahlen aus dem Intervall $[1, 10)$ wählen. Statt abzählbar vielen Möglichkeiten die Zahlen zu wählen, haben wir nun *überabzählbar* viele Möglichkeiten. Wählt Spieler $A$ nun die Zahl $x$ und Spieler $B$ die Zahl $y$, so gewinnt also Spieler $A$ wenn

$$1 \leq x \cdot y < 4 \qquad \text{oder} \qquad 10 \leq x \cdot y < 40.$$

Letztlich ist es noch Vorteilhaft nicht Zahlen $x, y \in [1, 10)$ zu wählen, sondern Zahlen $s, t \in [0, 1)$ zu wählen und diese dann mittels $x = 10^s$ und $y = 10^t$ bzw.

$$s = \log x \qquad \text{und} \qquad t = \log y$$

ineinander umzurechnen. Dann gewinnt $A$, falls

$$0 \leq s + t < \log 4 \qquad \text{oder} \qquad 1 \leq s + t < \log 40 = 1 + \log 4.$$

Dies können wir noch zusammenfassen zu

$$0 \leq \langle s + t \rangle < \log 4.$$

Die Gewinnregion von Spieler $A$ ist in Abbildung 9 grafisch verdeutlicht. Wählt Spieler $B$ die Zahl $t \in [0, 1)$, so gewinnt Spieler $A$, wenn er eine Zahl $s$ wählt, so dass das Tupel $(s, t)$ in dem roten Bereich liegt. Zusammenfassend erhalten wir folgende Modifikation des Multiplikationsspiels:

**Regeln des modifizierten Multiplikationsspiels.** Die Spieler $A$ und $B$ wählen unabhängig voneinander eine reelle Zahl aus dem Intervall $[0, 1)$. Spieler $A$ gewinnt, wenn die Summe der beiden Zahlen modulo 1 im Intervall $[0, \log 4)$ liegt – andernfalls gewinnt Spieler $B$.

Wir wollen nun für diese Modifikation des Spiels zeigen, dass $V_A = V_B$ ist. Die zugehörigen Strategien sind nun Wahrscheinlichkeitsmaße auf $([0, 1), \mathfrak{B}[0, 1))$. Wählt Spieler $A$ die Strategie $P$ und Spieler $B$ die Strategie $Q$, so wird in [6] gezeigt, dass dann

$$w(P, Q) = \int_0^1 \int_0^1 \mathbf{1}_{[0,\log 4)}(\langle s + t \rangle) \, P(\mathrm{d}s) \, Q(\mathrm{d}t) = \int_0^1 \int_0^1 \mathbf{1}_{[0,\log 4)}(\langle s + t \rangle) \, Q(\mathrm{d}t) \, P(\mathrm{d}s)$$

gilt. Eine einfache Möglichkeit für Spieler $A$ besteht darin, die Zahlen *gleichverteilt* zu wählen, das heißt gemäß der Gleichverteilung $\lambda_{0,1}$ auf $[0, 1)$. Wählt nun Spieler $B$ die Strategie $\delta_t$ – er wählt also stets die Zahl $t \in [0, 1)$ – so gewinnt $A$ er wenn eine Zahl $s \in [0, 1)$ so wählt, dass das Tupel $(s, t)$ im rot schraffierten Bereich von Abbildung 9 liegt. Genauer bedeutet das

$$\begin{aligned} w(\lambda_{0,1}, \delta_t) &= \int_0^1 \int_0^1 \mathbf{1}_{[0,\log 4)}(\langle s + t' \rangle) \, \delta_t(\mathrm{d}t') \, \lambda_{0,1}(\mathrm{d}s) = \int_0^1 \mathbf{1}_{[0,\log 4)}(\langle s + t \rangle) \, \lambda_{0,1}(\mathrm{d}s) \\ &= \int_0^1 \mathbf{1}_{\langle [0,\log 4) - t \rangle}(s) \, \lambda_{0,1}(\mathrm{d}s) = \lambda_{0,1}(\langle [0, \log 4) - t \rangle) = \lambda_{0,1}([0, \log 4)) \\ &= \log 4, \end{aligned}$$

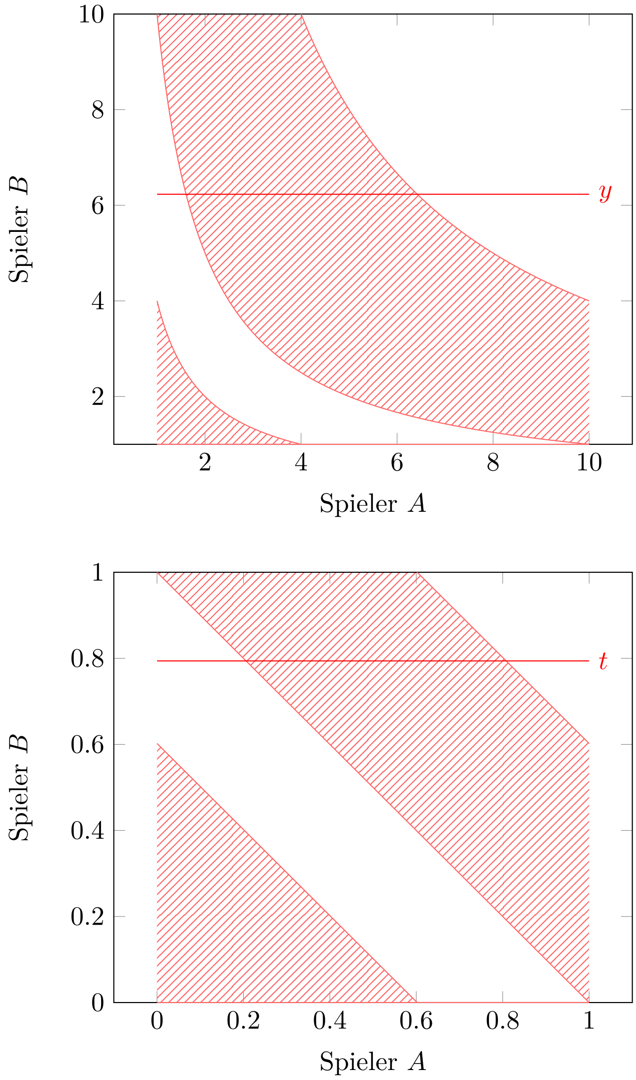


Abbildung 9: Die rote Fläche ist die Gewinnregion von $A$. Oben im Bereich $[1, 10)$ und unten nach dem Übergang zum Logarithmus im Bereich $[0, 1)$.

wobei die Translationsinvarianz des Lebesgue-Maßes ausgenutzt wurde. Man beachte, dass $w(\lambda\!\!\lambda_{0,1}, \delta_t) = \log 4$ unabhängig davon welches $t$ Spieler $B$ wählt. Betrachten wir nun den Fall, dass $B$ eine beliebige Strategie $Q$ wählt, so gilt

$$\begin{aligned} w(\lambda\!\!\lambda_{0,1}, Q) &= \int_0^1 \int_0^1 \mathbf{1}_{[0,\log 4)}(\langle s+t \rangle)\, \lambda\!\!\lambda_{0,1}(\mathrm{d}s)\, Q(\mathrm{d}t) = \int_0^1 \int_0^1 \mathbf{1}_{\langle [0,\log 4) - t \rangle}(s)\, \lambda\!\!\lambda_{0,1}(\mathrm{d}s)\, Q(\mathrm{d}t) \\ &= \int_0^1 \lambda\!\!\lambda_{0,1}(\langle [0,\log 4) - t \rangle)\, Q(\mathrm{d}t) = \int_0^1 \lambda\!\!\lambda_{0,1}([0,\log 4))\, Q(\mathrm{d}t) = \lambda\!\!\lambda_{0,1}([0,\log 4)) \\ &= \log 4. \end{aligned}$$

Damit gilt in jedem Falle

$$V_A = \sup_{P \in \mathfrak{W}} \inf_{Q \in \mathfrak{W}} w(P, Q) \geq \inf_{Q \in \mathfrak{W}} w(\lambda\!\!\lambda_{0,1}, Q) = \inf_{Q \in \mathfrak{W}} \log 4 = \log 4.$$

Wählt umgekehrt Spieler $B$ seine Zahl $t$ gemäß $\lambda\!\!\lambda_{0,1}$ und Spieler $A$ seine Zahl gemäß $\delta_s$, so erhalten wir analog

$$w\left(\delta_s, \lambda\!\!\lambda_{0,1}\right) = \log 4$$

und dies wieder unabhängig davon welche Zahl $s$ Spieler $A$ wählt und im Falle einer beliebigen Strategie $P$ ebenfalls

$$w\left(P, \lambda\!\!\lambda_{0,1}\right) = \log 4, \qquad \text{für alle } P \in \mathfrak{W}.$$

Wir erhalten also

$$V_B = \inf_{Q \in \mathfrak{W}} \sup_{P \in \mathfrak{W}} w\left(P, Q\right) \leq \sup_{P \in \mathfrak{W}} w\left(P, \lambda\!\!\lambda_{0,1}\right) = w\left(P, \lambda\!\!\lambda_{0,1}\right) = \log 4.$$

Daher gilt $V_B \leq \log 4 \leq V_A$. Da wegen Lemma 2.6 stets $V_A \leq V_B$ gilt, haben wir also bewiesen

$$V_A = V_B = \log 4.$$

Man nennt $V \coloneqq V_A = V_B$ dann auch den *Wert des Spiels* für Spieler $A$. Wenn also sowohl Spieler $A$ als auch Spieler $B$ ihre Zahlen $s, t \in [0, 1)$ gemäß der Gleichverteilung $\lambda\!\!\lambda_{0,1}$ wählen, so gilt nach obigen Gleichungen $w\left(\lambda\!\!\lambda_{0,1}, \lambda\!\!\lambda_{0,1}\right) = \log 4$. Also ist $\lambda\!\!\lambda_{0,1}$ eine optimale Strategie für beide Spieler. Spielt $A$ immer gemäß der Gleichverteilung auf $[0, 1)$, so gewinnt er (auf lange Sicht) *unabhängig* davon wie $B$ spielt, mindestens mit einer Wahrscheinlichkeit von $\log 4 \approx 0,6021$ – also in etwa 60% der Fälle. Dies ist insofern überraschend, da der Ausgangspunkt den Anschein eines unfairen Spiels für Spieler $A$ machte[10]: Schließlich gewinnt $A$ „nur" falls drei von neun möglichen Ziffern als Anfangsziffer des Produkts auftauchen. Die Ausführungen zeigen aber das Gegenteil – es handelt sich um ein unfaires Spiel für Spieler $B$.

Jedoch interessieren wir uns nicht für die Verteilung der $s, t \in [0, 1)$, sondern eher für die Verteilung der Zahlen $x = 10^s$ und $y = 10^t$ im Intervall $[1, 10)$. Wir erinnern uns an das Vorgehen, um das Spiel zu modifizieren und betrachten nun die finale Version des Multiplikationsspiels:

**Finale Version des Multiplikationsspiels.** Die Spieler $A$ und $B$ wählen unabhängig voneinander je eine reelle Zahl größer Null. Spieler $A$ gewinnt, wenn das Produkt der beiden Zahlen die erste signifikante Dezimalziffer 1, 2 oder 3 besitzt – andernfalls gewinnt Spieler $B$.

Die von $A$ bzw. $B$ ausgewählten reellen Zahlen $a > 0$ bzw. $b > 0$ werden so normiert, dass sie im Intervall $[1, 10)$ liegen – wir bilden also die Signifikante. Dies ändert nichts an der signifikanten Anfangsziffer von $a \cdot b$. Wir bezeichnen wie vorher

$$x = S\left(a\right) \qquad \text{und} \qquad y = S\left(b\right).$$

Wir schreiben nun $x = 10^s$ und $y = 10^t$, so dass

$$s = \log S\left(a\right) \qquad \text{und} \qquad t = \log S\left(b\right).$$

Mit der in Satz 1.3 eingeführten Abbildung $L \colon \mathbb{R}^+ \longrightarrow [0, 1)$ gemäß $L = \log \circ S$ gilt also $s = L\left(a\right)$ und $t = L\left(b\right)$. Da die Zahlen $s, t \in [0, 1)$ für die optimale Strategie gleichverteilt gewählt werden sollen, suchen wir also ein Wahrscheinlichkeitsmaß $\mu$ mit der Eigenschaft

$$\mu_L = \lambda\!\!\lambda_{0,1}. \tag{2.4}$$

Nach Satz 1.3 gibt es nun *genau* ein Wahrscheinlichkeitsmaß $\mu$ auf $\left(\mathbb{R}^+, \mathcal{S}\right)$ mit dieser Eigenschaft, welches in (1.4) als die Benford-Verteilung $\mathbb{B}$ charakterisiert wurde. Die von Spieler $A$ bzw. Spieler $B$ ausgewählten Zahlen $a$ und $b$ müssen bei der optimalen Strategie also gemäß der Benford-Verteilung gewählt werden. Zusammenfassend erhalten wir also

[10] Man würde das Spiel unfair für Spieler $A$ nennen, wenn $V_A < 1/2$.

**Satz 2.8** (Spieltheoretische Charakterisierung)**.** *Die Benford-Verteilung* $\mathbb{B}$ *ist sowohl für Spieler* $A$ *als auch Spieler* $B$ *eine optimale Strategie für die Wahl von* $a > 0$ *und* $b > 0$ *für das Multiplikationsspiel.*

Wir verallgemeinern das bisherige Spiel so, dass Spieler $A$ das Multiplikationsspiel dann gewinnt, wenn für ein gewisses $B \in \mathfrak{B}\,[1,10)$ gilt

$$x \cdot y \in B \qquad \text{oder} \qquad \frac{x \cdot y}{10} \in B.$$

Im bisherigen Spiel war $B = [1,4)$. Gehen wir dann ebenfalls zum Logarithmus über und bilden $\log B \in \mathfrak{B}\,[0,1)$, so gewinnt $A$ wenn mit $s := \log x$ und $t := \log y$ gilt

$$\langle s+t \rangle \in \log B.$$

Wählt Spieler $A$ als Strategie wieder die Gleichverteilung $\lambda_{0,1}$ und Spieler $B$ die Strategie $Q$ so erhalten wir wieder wegen der Translationsinvarianz des Lebesgue-Maßes

$$\begin{aligned} w\left(\lambda_{0,1}, Q\right) &= \int_0^1 \int_0^1 \mathbf{1}_{\log B}\left(\langle s+t \rangle\right) \lambda_{0,1}\left(\mathrm{d}s\right) Q\left(\mathrm{d}t\right) = \int_0^1 \int_0^1 \mathbf{1}_{\langle \log B - t \rangle}\left(s\right) \lambda_{0,1}\left(\mathrm{d}s\right) Q\left(\mathrm{d}t\right) \\ &= \int_0^1 \lambda_{0,1}\left(\langle \log B - t \rangle\right) Q\left(\mathrm{d}t\right) = \int_0^1 \lambda_{0,1}\left(\log B\right) Q\left(\mathrm{d}t\right) = \lambda_{0,1}\left(\log B\right). \end{aligned}$$

Damit gilt wieder $V_A \geq \lambda_{0,1}(\log B)$. Analog zeigt man $V_B \leq \lambda_{0,1}(\log B)$, so dass auch hier

$$V_A = V_B = \lambda_{0,1}\left(\log B\right)$$

gilt. Auch hier ist es für beide Spieler die optimale Strategie die Zahlen $s, t \in [0,1)$ gleichverteilt auf $[0,1)$ zu wählen. Transformiert man dies wieder auf das ursprüngliche Intervall $[1,10)$ zurück, so ergibt sich wie in (2.4), dass es für beide optimal ist die Zahlen $x$ und $y$ gemäß der Benford-Verteilung $\mathbb{B}$ zu wählen. Der Wert $V$ des Spiels für Spieler $A$ ist dann $V = \lambda_{0,1}(\log B)$.

Man kann das Multiplikationsspiel auch als Spiel auf bestimmten kompakten topologischen Gruppen betrachten. Das soll hier nun aber nur skizziert werden – für die Details siehe [15]. Die zugrundeliegende Menge $X := [1,10)$ ist nämlich mit der Multiplikation

$$x * y := \begin{cases} x \cdot y, & \text{falls } 1 \leq x \cdot y < 10, \\ \frac{x \cdot y}{10}, & \text{falls } 10 \leq x \cdot y < 100 \end{cases}$$

eine kommutative Gruppe, welche mittels

$$X \ni x \longmapsto \mathrm{e}^{\mathrm{i}2\pi \log x} \in \mathbb{T}$$

isomorph zum Einheitskreis $(\mathbb{T}, \cdot)$ ist. Demnach handelt es sich um eine kompakte topologische Gruppe $(X, *)$. Genauso handelt es sich bei der Menge $Y := [0,1)$ versehen mit der Addition Modulo 1 – das heißt $s+t := \langle s+t \rangle$ – um eine kommutative Gruppe, welche ebenfalls mittels

$$Y \ni y \longmapsto \mathrm{e}^{\mathrm{i}2\pi y} \in \mathbb{T}$$

isomorph zum Einheitskreis $(\mathbb{T}, \cdot)$ ist. Daher ist $(Y, +)$ ebenfalls eine kompakte topologische Gruppe. Die Abbildung $\Phi\colon Y \longrightarrow X$ definiert gemäß $\Phi(y) := 10^y$ ist dann ein Gruppenisomorphismus zwischen $X$ und $Y$. Betrachten wir auf $Y = [0,1)$ die Borelsche $\sigma$-Algebra $\mathfrak{B}\,[0,1)$ und darauf die Gleichverteilung $\lambda_{0,1}$ – also das Lebesgue-Maß auf $[0,1)$ – so gilt für das Lebesgue-Maß bekanntermaßen die *Translationsinvarianz*

$$\lambda_{0,1}\left(y+B\right) = \lambda_{0,1}\left(B\right), \qquad \text{für alle } y \in Y \text{ und alle } B \in \mathfrak{B}\,[0,1), \tag{2.5}$$

wobei „+" im Sinne der Gruppenoperation Modulo 1 zu verstehen ist. Im Sinne der Theorie der kompakten topologischen Gruppen ist das Lebesgue-Maß $\lambda\!\!\lambda_{0,1}$ nämlich das *Haar-Maß* der Gruppe $(Y,+)$. Mittels des Isomorphismus' $\Phi$ überträgt sich (2.5) auf $(X,*)$ ähnlich wie in (2.4) zu

$$\mathbb{B}\left(x * A\right) = \mathbb{B}\left(A\right), \qquad \text{für alle } x \in X \text{ und alle } A = 10^{B} \text{ mit } B \in [0,1)\,.$$

Dies bedeutet im Sinne der Theorie kompakter topologischer Gruppen, dass die Benford-Verteilung auf $[1,10)$ gerade das Haar-Maß der Gruppe $(X,*)$ ist und spiegelt die Eigenschaft der skaleninvarianten signifikanten Dezimalziffern gemäß Satz 2.7 wider.

## 2.5 Basis-Invarianz-Charakterisierung

Genügt Datensatz dem empirischen Benfordschen Gesetz (1.1), so sollte die Gültigkeit *nicht* davon abhängen zu welcher Basis $b$ man die Daten betrachtet. Sei nun $A \in \mathcal{S}$. Wie in Lemma 1.2 gezeigt, gilt dann für $n \in \mathbb{N}$ auch $\sqrt[n]{A} \in \mathcal{S}$. Nach [10] ist es nun so, dass ein Basiswechsel von der Basis $b$ zur Basis $b^n$ dann gerade dem Übergang von $\sqrt[n]{A}$ zu $A$ bedeutet. Deshalb ist es sinnvoll, folgende Definition zu tätigen:

**Definition 2.9.** Sei $\Sigma$ eine $\sigma$-Algebra auf $\mathbb{R}^+$ mit $\Sigma \supseteq \mathcal{S}$. Ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \Sigma)$ hat *Basis-invariante signifikante Ziffern*, falls

$$P\left(A\right) = P\left(\sqrt[n]{A}\right), \qquad \text{für alle } A \in \mathcal{S} \text{ und alle } n \in \mathbb{N}.$$

**Beispiel 2.6.**

*(a)* Wir betrachten das Dirac-Maß $\delta_1$, welches im Punkt 1 konzentriert ist. Für alle $A \in \mathcal{S}$ gilt dann

$$\begin{aligned}\delta_1\left(A\right) &= \begin{cases} 1, & \text{falls } 1 \in A, \\ 0, & \text{falls } 1 \notin A, \end{cases} \\ &= \begin{cases} 1, & \text{falls } 1 \in \sqrt[n]{A}, \\ 0, & \text{falls } 1 \notin \sqrt[n]{A}, \end{cases} \\ &= \delta_1\left(\sqrt[n]{A}\right),\end{aligned}$$

da für alle $n \in \mathbb{N}$ dann $1 \in A$ genau dann, wenn $1 \in \sqrt[n]{A} = \{x > 0 \,|\, x^n \in A\}$. Nach Definition 2.9 hat dann $\delta_1$ Basis-invariante signifikante Ziffern. Andererseits hat das Dirac-Maß $\delta_3$, welches im Punkt 3 konzentriert ist, keine Basis-invariante signifikanten Ziffern, denn mit $A = \{x > 0 \,|\, S\left(x\right) \in [3,4)\}$ gilt $\delta_3\left(A\right) = 1$ wohingegen für

$$\sqrt{A} = \left\{x > 0 \,\middle|\, S\left(x\right) \in \left[\sqrt{3}, 2\right) \cup \left[\sqrt{30}, \sqrt{40}\right)\right\} \subseteq [1,2] \cup [5,7]$$

gilt

$$\delta_3\left(\sqrt{A}\right) = 0.$$

Damit haben wir also $\delta_3\left(A\right) \neq \delta_3\left(\sqrt{A}\right)$ und $\delta_3$ hat *keine* Basis-invariante signifikanten Ziffern.

*(b)* Betrachten wir die Gleichverteilung $\lambda\!\!\lambda_{0,1}$ auf $[0,1)$. Sei nun

$$A = \{x > 0 \,|\, d_1\left(x\right) = 1\} = \{x > 0 \,|\, S\left(x\right) \in [1,2)\} = \bigcup_{k=-\infty}^{\infty} 10^k\,[1,2)\,.$$

Da $\lambda\!\!\!\lambda_{0,1}\left(\mathbb{R}^+ \setminus [0,1)\right) = 0$ reicht es

$$A' = \bigcup_{n=1}^{\infty} 10^{-n}\,[1,2)$$

zu betrachten. Es ist nun

$$\lambda\!\!\!\lambda_{0,1}\left(A'\right) = \sum_{n=1}^{\infty} \lambda\!\!\!\lambda_{0,1}\left(10^{-n}\,[1,2)\right) = \sum_{n=1}^{\infty} 10^{-n} = \frac{1}{9}.$$

Andererseits ist aber

$$\sqrt{A'} = \{x > 0 \,|\, x \in A'\} = \bigcup_{n=1}^{\infty} 10^{-n}\left(\left[1, \sqrt{2}\right) \cup \left[\sqrt{10}, \sqrt{20}\right)\right),$$

woraus sich dann sofort ergibt

$$\begin{aligned}
\lambda\!\!\!\lambda_{0,1}\left(\sqrt{A'}\right) &= \sum_{n=1}^{\infty} \lambda\!\!\!\lambda_{0,1}\left(10^{-n}\left(\left[1, \sqrt{2}\right) \cup \left[\sqrt{10}, \sqrt{20}\right)\right)\right) \\
&= \sum_{n=1}^{\infty} 10^{-n} \underbrace{\left(\sqrt{2} - 1 + \sqrt{20} - \sqrt{10}\right)}_{=\left(\sqrt{2}-1\right)\left(1+\sqrt{10}\right)} \\
&= \frac{1}{9} \underbrace{\left(\sqrt{2} - 1\right)\left(1 + \sqrt{10}\right)}_{>1} > \frac{1}{9}.
\end{aligned}$$

Das heißt es gilt $\lambda\!\!\!\lambda_{0,1}\left(A'\right) < \lambda\!\!\!\lambda_{0,1}\left(\sqrt{A'}\right)$, woraus sich sofort ergibt, dass $\lambda\!\!\!\lambda_{0,1}$ *keine* Basis-invarianten signifikanten Ziffern besitzt.

*(c)* Wir betrachten nun die Benford-Verteilung $\mathbb{B}$ auf $(\mathbb{R}^+, \mathcal{S})$. Wir betrachten nun zunächst

$$A = \bigcup_{k=-\infty}^{+\infty} 10^k\,[1, 10^s)\,, \qquad \text{mit einem } 0 \le s < 1.$$

Sei nun $n \in \mathbb{N}$. Im Beweis von Teil *(c)* von Lemma 1.2 wurde gezeigt

$$\sqrt[n]{A} = \bigcup_{k=-\infty}^{+\infty} 10^k \bigcup_{j=0}^{n-1} \left[10^{j/n}, 10^{(j+s)/n}\right).$$

Damit haben wir

$$\begin{aligned}
\mathbb{B}\left(\sqrt[n]{A}\right) &= \sum_{j=0}^{n-1} \mathbb{B}\left(\left[10^{j/n}, 10^{(j+s)/n}\right)\right) = \sum_{j=0}^{n-1}\left(\log 10^{(j+s)/n} - \log 10^{j/n}\right) \\
&= \sum_{j=0}^{n-1}\left(\frac{j+s}{n} - \frac{j}{n}\right) = s = \log 10^s - \log 1 = \mathbb{B}\,(A)\,.
\end{aligned}$$

Damit gilt die Gleichheit für alle $A \in \mathcal{S}$ der obigen Form. Analog zeigt man die Eigenschaft für beliebige $A = \bigcup_{k=-\infty}^{+\infty} 10^k B \in \mathcal{S}$.

Beispiel 2.6 zeigt insbesondere, dass das Dirac-Maß $\delta_1$ und die Benford-Verteilung $\mathbb{B}$ Basis-invariante signifikante Ziffern besitzen. Jede Konvexkombination

$$P_\alpha \coloneqq \alpha\delta_1 + (1-\alpha)\,\mathbb{B}, \qquad \text{mit einem } 0 \le \alpha \le 1,$$

hat dann wegen

$$\begin{aligned} P_\alpha(A) &= \alpha\delta_1(A) + (1-\alpha)\mathbb{B}(A) \\ &= \alpha\delta_1\left(\sqrt[n]{A}\right) + (1-\alpha)\mathbb{B}\left(\sqrt[n]{A}\right) \\ &= P_\alpha\left(\sqrt[n]{A}\right) \end{aligned}$$

auch Basis-invariante signifikante Ziffern. Damit gibt es also *überabzählbar* viele Wahrscheinlichkeitsmaße $P_\alpha$ mit Basis-invarianten signifikanten Ziffern. Das Hauptresultat dieses Abschnitts wird nun aber zeigen, dass es eben *genau* nur diese Wahrscheinlichkeitsmaße $P_\alpha$ sind, welche diese Eigenschaft erfüllen.

Für das Hauptresultat dieses Abschnitts benötigen wir noch etwas Vorbereitung: Sei $(\Omega, \Sigma, P)$ ein Wahrscheinlichkeitsraum und $T\colon \Omega \longrightarrow \Omega$ eine messbare Abbildung. Wir nennen $T$ dann *$P$-maßerhaltend* oder sagen $P$ ist *$T$-invariant*, falls $P_T = P$, das heißt

$$P_T(A) = P\left(T^{-1}(A)\right) = P(A), \qquad \text{für alle } A \in \Sigma.$$

Für $n \in \mathbb{N}$ betrachten wir nun die Abbildung $T_n\colon [0,1) \longrightarrow [0,1)$ definiert gemäß

$$T_n(x) := \langle nx \rangle .$$

Wir suchen nun alle Wahrscheinlichkeitsmaße $P$ auf $([0,1), \mathfrak{B}[0,1))$, welche $T_n$-invariant für alle $n \in \mathbb{N}$ sind.

**Lemma 2.7.** *Ein Wahrscheinlichkeitsmaß $P$ auf $([0,1), \mathfrak{B}[0,1))$ ist genau dann für alle $n \in \mathbb{N}$ $T_n$-invariant, wenn für ein gewisses $0 \leq \alpha \leq 1$ gilt*

$$P = \alpha\delta_0 + (1-\alpha)\lambda_{0,1}.$$

Beweis. Den Beweis führen wir mithilfe der Fourier-Koeffizienten

$$\widehat{P}(k) := \int_0^1 e^{i2\pi ks}\, P(ds), \qquad \text{für } k \in \mathbb{Z},$$

welche wir in Abschnitt 2.1 eingeführt hatten. Wir wollen nun zunächst die Fourier-Koeffizienten des Bildmaßes $P_{T_n}$ bestimmen. Es gilt

$$\begin{aligned} \widehat{P_{T_n}}(k) &= \int_0^1 e^{i2\pi ks}\, P_{T_n}(ds) = \int_0^1 e^{i2\pi kT_n(s)}\, P(ds) \\ &= \int_0^1 e^{i2\pi k\langle ns\rangle}\, P(ds) = \int_0^1 e^{i2\pi k(ns-\lfloor ns\rfloor)}\, P(ds) \\ &= \int_0^1 e^{i2\pi kns} \underbrace{e^{-i2\pi k\lfloor ns\rfloor}}_{=1} P(ds) = \int_0^1 e^{i2\pi (kn)s}\, P(ds) \\ &= \widehat{P}(kn). \end{aligned} \tag{2.6}$$

Nun werden wir die Äquivalenz zeigen:

**$\Leftarrow$:** Sei $P = \alpha\delta_0 + (1-\alpha)\lambda_{0,1}$. Zunächst gilt

$$\widehat{\delta_0}(k) = \int_0^1 e^{i2\pi ks}\, \delta_0(ds) = e^{i2\pi k\cdot 0} = 1, \qquad \text{für alle } k \in \mathbb{Z}.$$

Ferner ist bereits bekannt

$$\widehat{\lambda\!\!\lambda_{0,1}}(k) = \begin{cases} 1, & \text{für } k = 0, \\ 0, & \text{falls } k \neq 0. \end{cases}$$

Fasst man dies nun zusammen haben wir also

$$\widehat{P}(k) = \begin{cases} 1, & \text{falls } k = 0, \\ \alpha, & \text{falls } k \neq 0. \end{cases}$$

Damit gilt wegen (2.6) für $n \in \mathbb{N}$ und $k \in \mathbb{Z}$ mit $k \neq 0$

$$\widehat{P_{T_n}}(k) = \widehat{P}(nk) = \widehat{P}(k).$$

und natürlich auch $\widehat{P_{T_n}}(0) = 1$, so dass

$$\widehat{P_{T_n}}(k) = \widehat{P}(k), \qquad \text{für alle } k \in \mathbb{Z} \text{ und alle } n \in \mathbb{N}.$$

Da die Fourier-Koeffizienten das Maß eindeutig bestimmen, gilt also $P_{T_n} = P$ – also ist $P$ $T_n$-invariant für alle $n \in \mathbb{N}$.

**⇒:** Sei nun $P$ $T_n$-invariant für alle $n \in \mathbb{N}$. Dann gilt wegen (2.6) für alle $n \in \mathbb{N}$ insbesondere

$$\widehat{P}(n) = \widehat{P}(n \cdot 1) = \widehat{P_{T_n}}(1) = \widehat{P}(1) =: \alpha$$

und genauso

$$\widehat{P}(-n) = \widehat{P}(-n \cdot 1) = \widehat{P_{T_n}}(-1) = \widehat{P}(-1).$$

Da für Fourier-Koeffizienten stets gilt

$$\overline{\widehat{P}(k)} = \overline{\int_0^1 e^{i2\pi ks}\, P(ds)} = \int_0^1 e^{-i2\pi ks}\, P(ds) = \widehat{P}(-k), \qquad \text{für alle } k \in \mathbb{Z},$$

erhalten wir also

$$\widehat{P}(k) = \begin{cases} \alpha, & \text{falls } k \geq 1, \\ 1, & \text{falls } k = 0, \\ \overline{\alpha}, & \text{falls } k \leq -1. \end{cases}$$

Nun ist für $x \notin \mathbb{Z}$

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{i2\pi nx} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(e^{i2\pi x}\right)^n = \lim_{N\to\infty} \frac{1}{N} \left( \sum_{n=0}^{N} \left(e^{i2\pi x}\right)^n - 1 \right) \\ &= \lim_{N\to\infty} \frac{1}{N} \left( \frac{\left(e^{i2\pi x}\right)^{N+1} - 1}{e^{i2\pi x} - 1} - 1 \right) = 0 \end{aligned}$$

und für $x \in \mathbb{Z}$ andererseits

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{i2\pi nx} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1 = 1.$$

Also haben wir

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{i2\pi nx} = \begin{cases} 0, & \text{falls } x \in \mathbb{R}\backslash\mathbb{Z}, \\ 1, & \text{falls } x \in \mathbb{Z}. \end{cases}$$

Daraus ergibt sich mittels des Satzes von der majorisierten Konvergenz

$$\begin{aligned} P(\{0\}) &= \int_0^1 1_{\{0\}}(s)\, P(\mathrm{d}s) = \int_0^1 \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi ns}\, P(\mathrm{d}s) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_0^1 \mathrm{e}^{\mathrm{i}2\pi ns}\, P(\mathrm{d}s) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \widehat{P}(n) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \widehat{P}(1) = \alpha. \end{aligned}$$

Damit ist $\alpha$ reell und $0 \le \alpha \le 1$. Somit hat $P$ diegleichen Fourier-Koeffizienten wie $\alpha\delta_0 + (1-\alpha)\,\lambda\!\!\!\lambda_{0,1}$ und da die Fourier-Koeffizienten das Maß eindeutig bestimmen, gilt demnach $P = \alpha\delta_0 + (1-\alpha)\,\lambda\!\!\!\lambda_{0,1}$. □

Es folgt nun das bereits erwähnte Hauptresultat dieses Abschnitts, welches eine weitere Charakterisierung der Benford-Verteilung mittels Basis-invarianten signifikanten Ziffern liefert:

**Satz 2.9** (Basis-Invarianz-Charakterisierung)**.** *Sei $P$ ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\mathbb{R}^+, \Sigma)$, wobei $\Sigma \supseteq \mathcal{S}$ gelte. Dann sind folgende Aussagen äquivalent:*

(i) *$P$ hat Basis-invariante signifikante Ziffern.*

(ii) *Für alle $A \in \mathcal{S}$ gilt mit einem gewissen $0 \le \alpha \le 1$*

$$P(A) = \alpha\delta_1(A) + (1-\alpha)\,\mathbb{B}(A).$$

Bevor wir Satz 2.9 beweisen, erhalten wir zunächst direkt eine wichtige

**Folgerung 2.3.** *Sei $P$ ein stetiges Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\mathbb{R}^+, \Sigma)$, wobei $\Sigma \supseteq \mathcal{S}$ gelte. Dann sind folgende Aussagen äquivalent:*

(i) *$P$ hat Basis-invariante signifikante Ziffern.*

(ii) *Für alle $A \in \mathcal{S}$ gilt $P(A) = \mathbb{B}(A)$, das heißt $P$ ist Benford.*

Beweis von Satz 2.9. Wie im Beweis von Satz 2.7 bezeichne $P^*$ die Einschränkung von $P$ auf $(\mathbb{R}^+, \mathcal{S})$ und

$$Q := P^*_L$$

mit der in Satz 1.3 definierten Abbildung $L(x) := \log S(x)$. Zunächst zeigen wir, dass $P^*$ genau dann Basis-invariante signifikante Ziffern besitzt, wenn $Q$ für alle $n \in \mathbb{N}$ $T_n$-invariant ist. Für $A = \bigcup_{k=-\infty}^{+\infty} 10^k [1, 10^s)$ mit einem $0 \le s < 1$ haben wir nämlich einerseits

$$\begin{aligned} Q_{T_n}([0,s)) &= Q\left(T_n^{-1}([0,s))\right) = Q\left(\bigcup_{j=0}^{n-1} \left[\frac{j}{n}, \frac{j+s}{n}\right)\right) \\ &= Q\left(L\left(\bigcup_{k=-\infty}^{+\infty} 10^k \bigcup_{j=0}^{n-1} \left[10^{j/n}, 10^{(j+s)/n}\right)\right)\right) \\ &= P^*\left(\bigcup_{k=-\infty}^{+\infty} 10^k \bigcup_{j=0}^{n-1} \left[10^{j/n}, 10^{(j+s)/n}\right)\right) \\ &= P^*\left(\sqrt[n]{A}\right) \end{aligned}$$

und andererseits

$$Q\left([0,s)\right) = P^*\left(\bigcup_{k=-\infty}^{+\infty} 10^k 10^{[0,s)}\right) = P^*\left(\bigcup_{k=-\infty}^{+\infty} 10^k\left[1,10^s\right)\right) = P^*\left(A\right).$$

Falls also $P^*\left(A\right) = P^*\left(\sqrt[n]{A}\right)$ für alle $n \in \mathbb{N}$ gilt, so haben wir

$$Q\left([0,s)\right) = P^*\left(A\right) = P^*\left(\sqrt[n]{A}\right) = Q_{T_n}\left([0,s)\right), \tag{2.7}$$

also ist $Q$ für alle $n \in \mathbb{N}$ $T_n$-invariant. Falls andererseits $Q$ für alle $n \in \mathbb{N}$ $T_n$-invariant ist folgt aus der Gleichheit in (2.7) das $P^*$ Basis-invariante signifikante Ziffern besitzt.

Aussage *(i)* ist also äquivalent dazu, dass $Q$ für alle $n \in \mathbb{N}$ $T_n$-invariant ist. Nach Lemma 2.7 ist dies äquivalent zu

$$Q = \alpha\delta_0 + (1-\alpha)\,\lambda_{0,1}, \qquad \text{mit einem gewissen } 0 \leq \alpha \leq 1,$$

was wegen $Q = P_L^*$, $\left(\delta_1\right)_L = \delta_0$ und der Gleichung (1.4) äquivalent ist mit

$$P^* = \alpha\delta_1 + (1-\alpha)\,\mathbb{B},$$

was gerade *(ii)* bedeutet. □

Das Kapitel wird nun mit zwei Folgerungen abgeschlossen, welche Aufschluss über den Zusammenhang zwischen skaleninvarianten signifikanten Dezimalziffern und Basis-invarianten signifikanten Ziffern geben:

**Folgerung 2.4.** *Hat ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^+, \Sigma)$ mit $\Sigma \supseteq \mathcal{S}$ skaleninvariante signifikante Dezimalziffern, so besitzt es auch Basis-invariante signifikante Ziffern.*

Beweis. Nach Satz 2.7 ist die Skaleninvarianz äquivalent dazu, dass $P$ Benford ist. Bezeichnet $P^*$ wieder die Einschränkung von $P$ auf $\mathcal{S}$, so impliziert die Darstellung $P^* = 0 \cdot \delta_1 + (1-0)\,\mathbb{B}$ mit Satz 2.9, dass $P$ Basis-invariante signifikante Ziffern besitzt. □

Die Umkehrung von Folgerung 2.4 gilt im Allgemeinen nicht. So besitzt nach Satz 2.9 dann $P \coloneqq 1 \cdot \delta_1 + (1-1)\,\mathbb{B} = \delta_1$ Basis-invariante signifikante Ziffern, aber $\delta_1$ hat *keine* skaleninvarianten signifikanten Dezimalziffern: Für $A = \bigcup_{k=-\infty}^{+\infty} 10^k\,[1,2)$ gilt nämlich $\delta_1\left(A\right) = 1$, aber für $2 \cdot A = \bigcup_{k=-\infty}^{+\infty} 10^k\,[2,4)$ gilt $\delta_1\left(2 \cdot A\right) = 0$. Dieser Fall tritt immer dann ein, wenn $P = \alpha\delta_1 + (1-\alpha)\,\mathbb{B}$ mit einem $\alpha \neq 0$. Daraus erhalten wir abschließend noch

**Folgerung 2.5.** *Sei $P$ ein stetiges Wahrscheinlichkeitsmaß auf $(\mathbb{R}^+, \Sigma)$ mit $\Sigma \supseteq \mathcal{S}$. Dann sind folgende Aussagen äquivalent:*

(i) *$P$ ist Benford.*

(ii) *$P$ hat Basis-invariante signifikante Ziffern.*

(iii) *$P$ hat skaleninvariante signifikante Dezimalziffern.*

# 3 Asymptotisches Benfordsches Gesetz für Produkte von Zufallsgrößen

In diesem Kapitel steht die Untersuchung von Folgen von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ mit Blick auf die Benford-Eigenschaft im Grenzverhalten $n \longrightarrow +\infty$ im Vordergrund. Insbesondere interessieren wir uns für das Grenzverhalten des Produkts

$$\prod_{k=1}^{n} X_k = X_1 \cdot X_2 \cdot \ldots \cdot X_n, \qquad \text{für } n \in \mathbb{N}.$$

Ein erstes wichtiges Resultat in dieser Thematik ist bereits Satz 2.6. Dieser besagt, dass für eine unabhängig und identisch verteilte Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$, so dass $X_1$ nicht nur abzählbar viele Werte annimmt, die Folge der Produkte $\left(\prod_{k=1}^{n} X_n\right)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz konvergieren, also

$$P\left(\left\{S\left(\prod_{k=1}^{n} X_k\right) \leq t\right\}\right) \xrightarrow{n\to+\infty} \log t, \qquad \text{für alle } 1 \leq t < 10.$$

Wir wollen nun weitere Resultate dieser Art herleiten.

## 3.1 Abhängige Zufallsgrößen

In Erinnerung an das im Abschnitt 2.2 eingeführte Konzept der Konvergenz in Verteilung sei an die Definition 2.7 erinnert: Eine Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ konvergiert in Verteilung gegen das Benfordsche Gesetz, falls

$$\lim_{n\to\infty} P\left(\left\{S\left(X_n\right) \leq t\right\}\right) = \log t, \qquad \text{für alle } 1 \leq t < 10.$$

Sei nun $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$. Wir untersuchen nun das Verhalten der geometrischen Folge

$$X_n := X^n = \prod_{k=1}^{n} X.$$

Damit ist die Folge $(X_n)_{n\in\mathbb{N}}$ insbesondere *nicht* stochastisch unabhängig. Unser Ziel ist nun eine Verallgemeinerung der Resultate für Zahlenfolgen der Form$(a^n)_{n\in\mathbb{N}}$ auf Folgen von Zufallsgrößen der Form $(X^n)_{n\in\mathbb{N}}$. Das Hauptresultat für Zahlenfolgen war hierbei Satz 2.2, welcher besagt, dass die Benford-Eigenschaft einer Folge $(a^n)_{n\in\mathbb{N}}$ äquivalent zu $\log|a| \in \mathbb{R}\backslash\mathbb{Q}$ ist. Das Analogon für Zufallsgrößen ist folgender

**Satz 3.1.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$. Dann sind folgende Aussagen äquivalent:*

(i) $P\left(\left\{(X^n)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1$.

(ii) $P\left(\left\{\log|X| \in \mathbb{Q}\right\}\right) = 0$.

Beweis. Für $\omega \in \Omega$ ist die Zahlenfolge $(X^n(\omega))_{n\in\mathbb{N}}$ wegen Satz 2.2 genau dann Benford, wenn $\log|X(\omega)|$ irrational ist. Daraus ergibt sich direkt die behauptete Äquivalenz. □

Zur Vorbereitung auf das nächste Resultat benötigen wir noch folgendes

**Lemma 3.1.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, welche eine Riemann-integrierbare Dichte $f_X$ besitzt. Dann gilt*

$$\lim_{n\to\infty} P\left(\{\langle nX\rangle \le s\}\right) = s, \qquad \textit{für alle } 0 \le s < 1.$$

Beweis. Da $\langle nX\rangle = \langle n\,\langle X\rangle\rangle$ gilt, können wir ohne Einschränkung annehmen, dass $X$ nur Werte in $[0,1)$ annimmt. Die Dichte $f_X\colon [0,1] \longrightarrow \mathbb{R}$ von $X$ erfüllt dann für alle $0 \le s < 1$ die Bedingung

$$P\left(\{X \le s\}\right) = \int_0^s f_X(t)\,\mathrm{d}t, \qquad \text{für alle } 0 \le s < 1.$$

Da $\{\langle nX\rangle \le s\} = \left\{X \in \bigcup_{k=0}^{n-1} \left[\frac{k}{n}, \frac{k+s}{n}\right)\right\}$ erhalten wir mittels der Substitution $t = \frac{k+z}{n}$ dann

$$\begin{aligned}
P\left(\{\langle nX\rangle \le s\}\right) &= P\left(\left\{X \in \bigcup_{k=0}^{n-1}\left[\frac{k}{n}, \frac{k+s}{n}\right)\right\}\right) = \sum_{k=0}^{n-1} P\left(\left\{X \in \left[\frac{k}{n}, \frac{k+s}{n}\right)\right\}\right) \\
&= \sum_{k=0}^{n-1} \int_{k/n}^{(k+s)/n} f_X(t)\,\mathrm{d}t = \sum_{k=0}^{n-1} \int_0^s \frac{1}{n} f_X\left(\frac{k+z}{n}\right)\mathrm{d}z \\
&= \int_0^s \frac{1}{n}\sum_{k=0}^{n-1} f_X\left(\frac{k+z}{n}\right)\mathrm{d}z,
\end{aligned}$$

so dass die Dichte von $\langle nX\rangle$ gegeben ist durch

$$f_{\langle nX\rangle}(z) = \frac{1}{n}\sum_{k=0}^{n-1} f_X\left(\frac{k+z}{n}\right), \qquad \text{für alle } 0 \le z \le 1.$$

Die rechte Seite sind gerade Riemann-Zwischensummen mit äquidistanter Zerlegung der Breite $1/n$. Da $f_X$ Riemann-integrierbar ist, gilt daher für alle $0 \le z \le 1$

$$\lim_{n\to\infty} f_{\langle nX\rangle}(z) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_X\left(\frac{k+z}{n}\right) = \int_0^1 f_X(t)\,\mathrm{d}t = 1. \tag{3.1}$$

Sei $\varepsilon > 0$. Wegen (3.1) gibt es dann ein $N_0 \in \mathbb{N}$, so dass für alle $n \ge N_0$ gilt

$$\left|f_{\langle nX\rangle}(z) - 1\right| < \varepsilon, \qquad \text{für alle } 0 \le z \le 1.$$

Nach der Definition der Riemann-Integrierbarkeit hängt dieses $N_0$ insbesondere nicht von $z$ ab. Damit erhalten wir für alle $n \ge N_0$

$$\int_0^1 \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z \le \int_0^1 \varepsilon\mathrm{d}z = \varepsilon,$$

so dass in jedem Falle $\limsup_{n\to\infty} \int_0^1 \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z \le \varepsilon$. Da $\varepsilon > 0$ beliebig war, ist damit gezeigt das

$$\lim_{n\to\infty} \int_0^1 \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z = 0.$$

Damit erhalten wir schließlich

$$\begin{aligned} \left|P\left(\{\langle nX\rangle \le s\} - s\right)\right| &= \left|\int_0^s f_{\langle nX\rangle}(z)\,\mathrm{d}z - s\right| = \left|\int_0^s \left(f_{\langle nX\rangle}(z) - 1\right)\mathrm{d}z\right| \\ &\le \int_0^s \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z \le \int_0^1 \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z. \end{aligned}$$

Da $\lim_{n\to\infty}\int_0^1 \left|f_{\langle nX\rangle}(z) - 1\right|\mathrm{d}z = 0$, folgt also

$$\left|P\left(\{\langle nX\rangle \le s\}\right) - s\right| \xrightarrow{n\to\infty} 0, \qquad \text{für alle } 0 \le s < 1. \qquad \square$$

Mit diesem Lemma können wir nun folgendes Resultat beweisen:

**Satz 3.2.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, welche eine Riemann-integrierbare Dichte $f_X$ besitzt. Dann gilt:*

(a) *Die Folge $(X^n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen das Benfordsche Gesetz.*

(b) *$(X^n)_{n\in\mathbb{N}}$ ist mit Wahrscheinlichkeit 1 Benford, das heißt*

$$P\left(\left\{(X^n)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1.$$

BEWEIS.

**Zu *(a)*:** Es gilt

$$\begin{aligned} P\left(\{S(X^n) \le t\}\right) &= P\left(\left\{|X|^n \in \bigcup_{k=-\infty}^{+\infty} 10^k\,[1,t]\right\}\right) \\ &= P\left(\left\{\log|X|^n \in \bigcup_{k=-\infty}^{+\infty} [k, k+\log t)\right\}\right) \\ &= P\left(\{\langle \log|X|^n\rangle \in [0,\log t)\}\right) = P\left(\{\langle n\log|X|\rangle \le \log t\}\right). \end{aligned}$$

Da $X$ eine Riemann-integrierbare Dichte besitzt, besitzt auch $\log|X|$ eine Riemann-integrierbare Dichte, denn für $s \in \mathbb{R}$ haben wir mit der Substitution $t = \log z$ und $\mathrm{d}z = \ln 10 \cdot 10^t \mathrm{d}t$ dann

$$\begin{aligned} P\left(\{\log|X| \le s\}\right) &= P\left(\{|X| \le 10^s\}\right) = P\left(\{0 \le X \le 10^s\} \cup \{-10^s \le X < 0\}\right) \\ &= \int_0^{10^s} f_X(z)\,\mathrm{d}z + \int_{-10^s}^0 f_X(u)\,\mathrm{d}u = \int_0^{10^s} \left(f_X(z) + f_X(-z)\right)\mathrm{d}z \\ &= \int_{-\infty}^s \ln 10 \cdot 10^t \left(f_X\left(10^t\right) + f_X\left(-10^t\right)\right)\mathrm{d}t, \end{aligned}$$

so dass

$$f_{\log|X|}(t) = \ln 10 \cdot 10^t \left(f_X\left(10^t\right) + f_X\left(-10^t\right)\right)$$

die Dichte von $\log|X|$ ist. Lemma 3.1 liefert also für alle $1 \le t < 10$ dann

$$P\left(\{S(X^n) \le t\}\right) = P\left(\{\langle n\log|X|\rangle \le \log t\}\right) \xrightarrow{n\to\infty} \log t.$$

Damit konvergiert $(X^n)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz.

**Zu *(b)*:** Wie in *(a)* gezeigt, besitzt die Zufallsgröße $\log|X|$ eine Dichte und ist damit insbesondere stetig, das heißt $P(\{\log|X| = t\}) = 0$ für alle $t \in \mathbb{R}$. Deshalb gilt

$$P(\{\log|X| \in \mathbb{Q}\}) = P\left(\bigcup_{q\in\mathbb{Q}} \{\log|X| = q\}\right) = \sum_{q\in\mathbb{Q}} \underbrace{P(\{\log|X| = q\})}_{=0} = 0,$$

so dass $\log|X|$ mit Wahrscheinlichkeit 1 irrationale Werte annimmt. Nach Satz 3.1 ist dies äquivalent zu $P\left(\left\{(X^n)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1$. □

**Beispiel 3.1.** Sei $X$ gleichverteilt auf $[0,1)$. Nach Satz 3.2 konvergiert die Folge $(X^n)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz. Wir wollen nun die Konvergenzgeschwindigkeit feststellen. Da $X$ $P$-fast-sicher nur Werte auf $[0,1)$ annimmt, gilt

$$\{S(X^n) \le t\} = \left\{X^n \in \bigcup_{k=1}^{\infty} 10^{-k}[1,t]\right\}, \qquad \text{für alle } 1 \le t < 10.$$

Damit haben wir für alle $1 \le t < 10$ und alle $n \in \mathbb{N}$

$$\begin{aligned}
F_{S(X^n)}(t) &= P(\{S(X^n) \le t\}) = \sum_{k=1}^{\infty} P\left(\left\{10^{-k} \le X^n \le 10^{-k}t\right\}\right) \\
&= \sum_{k=1}^{\infty} P\left(\left\{10^{-k/n} \le X \le 10^{-k/n}t^{1/n}\right\}\right) = \sum_{k=1}^{\infty}\left(F_X\left(10^{-k/n}t^{1/n}\right) - F_X\left(10^{-k/n}\right)\right) \\
&= \sum_{k=1}^{\infty}\left(t^{1/n}10^{-k/n} - 10^{-k/n}\right) = \sum_{k=1}^{\infty} 10^{-k/n}\left(t^{1/n} - 1\right) \\
&= \left(t^{1/n} - 1\right)\sum_{k=1}^{\infty} 10^{-k/n} = \frac{t^{1/n} - 1}{10^{1/n} - 1}.
\end{aligned}$$

Mittels der Regel von L'Hospital erhalten wir schließlich

$$\begin{aligned}
\lim_{n\to\infty} \frac{t^{1/n} - 1}{10^{1/n} - 1} &= \lim_{n\to\infty} \frac{\mathrm{e}^{\frac{1}{n}\cdot\ln t} - 1}{\mathrm{e}^{\frac{1}{n}\ln 10} - 1} = \lim_{n\to\infty} \frac{-\frac{1}{n^2}\cdot\ln t\cdot \mathrm{e}^{\frac{1}{n}\cdot\ln t}}{-\frac{1}{n^2}\cdot\ln 10\cdot\mathrm{e}^{\frac{1}{n}\ln 10}} \\
&= \lim_{n\to\infty} \frac{\ln t\cdot\mathrm{e}^{\frac{1}{n}\cdot\ln t}}{\ln 10\cdot\mathrm{e}^{\frac{1}{n}\cdot\ln 10}} = \frac{\ln t}{\ln 10} = \log t,
\end{aligned}$$

so dass

$$\left|F_{S(X^n)}(t) - \log t\right| \xrightarrow{n\to+\infty} 0, \qquad \text{für alle } 1 \le t < 10.$$

Dies bestätigt die Aussage von Teil *(a)* des Satzes 3.2 für eine gleichverteilte Zufallsgröße $X$, welche als Dichte gerade $f_X(t) = \mathbf{1}_{[0,1)}(t)$ besitzt. Wir wollen nun zeigen, dass

$$\lim_{n\to\infty} n\cdot\left(F_{S(X^n)}(t) - \log t\right)$$

existiert und diesen Grenzwert bestimmen. Nach zweimaligen Anwenden der Regel von L'Hospital erhalten wir (ohne jeden Rechenschritt detailliert anzugeben) für alle $1 < t < 10$

$$\begin{aligned}
\lim_{n\to\infty} n\left(F_{S(X^n)}(t) - \log t\right) &= \lim_{n\to\infty} \frac{nt^{1/n} - n - n\log t\left(10^{1/n} - 1\right)}{10^{1/n} - 1} \\
&= \lim_{n\to\infty} \frac{t^{1/n} - 1 + \frac{1}{n}\ln t\left(10^{1/n} - t^{1/n}\right) - \log t\left(10^{1/n} - 1\right)}{-\frac{1}{n^2}\ln 10\cdot 10^{1/n}} \\
&= \lim_{n\to\infty} \frac{-\ln 10\cdot\ln t\cdot 10^{1/n} + (\ln t)^2 t^{1/n}}{2\ln 10\cdot 10^{1/n} + \frac{1}{n}(\ln 10)^2 t^{1/n}} = \frac{-\ln 10\cdot\ln t + (\ln t)^2}{2\ln 10} \\
&= \frac{\frac{\log t}{\log \mathrm{e}}\left(\frac{\log t}{\log \mathrm{e}} - \frac{\log 10}{\log \mathrm{e}}\right)}{2\frac{\log 10}{\log \mathrm{e}}} = \frac{\log t(\log t - 1)}{2\log \mathrm{e}}.
\end{aligned}$$

Damit gilt

$$n\left(F_{S(X^n)}(t) - \log t\right) \xrightarrow{n \to +\infty} \frac{\log t\,(\log t - 1)}{2 \log \mathrm{e}}, \qquad \text{für alle } 1 < t < 10.$$

Für $t = 1$ ist $F_{S(X^n)}(1) - \log 1 = 0 - 0 = 0$, so dass in diesem Fall der Ausdruck $n\left(F_{S(X^n)}(1) - \log 1\right)$ für alle $n \in \mathbb{N}$ gleich 0 ist. Damit gilt $\left|F_{S(X^n)}(t) - \log t\right| \in o\left({}^1\!/\!_n\right)$.

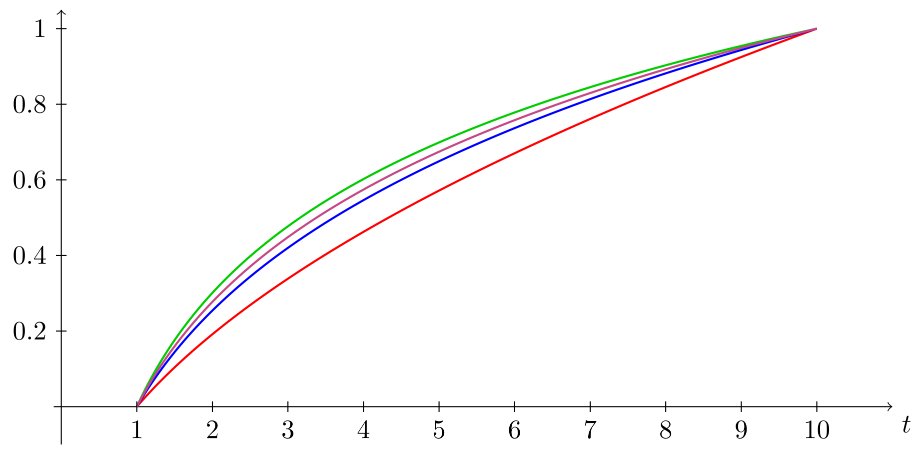


Abbildung 10: Veranschaulichung der Konvergenz von $F_{S(X^n)}(t)$ gegen $\log t$ (grün) für $n = 2$ (rot), $n = 5$ (blau) und $n = 10$ (violett).

**Beispiel 3.2** (Zufallszahlen am Computer)**.** Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, welche gleichverteilt auf $\left[1, 10^6\right)$ ist – das heißt $P_X$ ist gerade das normierte Lebesgue-Maß auf $\left[1, 10^6\right)$. Nach Satz 3.2 konvergiert dann die Folge der Produkte $(X^n)_{n \in \mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz – das heißt

$$P\left(\{S(X^n) \le t\}\right) \xrightarrow{n \to +\infty} \log t, \qquad \text{für alle } 1 \le t < 10.$$

Für große $N \in \mathbb{N}$ gilt also etwa $P\left(\left\{S\left(X^N\right) \le t\right\}\right) \approx \log t$. Für die erste signifikante Ziffer ist also für alle $d \in \mathbb{N}_{1,9}$ dann $P\left(\left\{d_1\left(X^N\right) = d\right\}\right) \approx \log\left(1 + \frac{1}{d}\right)$. Simuliert man die Realisierungen der gleichverteilten Zufallsgröße $X$ mittels des „*Mersenne-Twister-Algorithmus*" am Computer, so erhält man sogenannte Pseudo-Zufallszahlen, welche als Realisierungen einer solchen Zufallsgröße interpretiert werden können.

Mittels des Mersenne-Twister-Algorithmus wurden 100 Realisierungen der Zufallsgröße $X$ simuliert. Diese Realisierungen seien mit $x_1^{(1)}, x_2^{(1)}, \ldots, x_{100}^{(1)}$ bezeichnet. Führt man diese Prozedur nun $N$-mal durch, und betrachtet die Produkte

$$\prod_{k=1}^{N} x_1^{(k)}, \prod_{k=1}^{N} x_2^{(k)}, \ldots, \prod_{k=1}^{N} x_{100}^{(k)},$$

so besagt Satz 3.2, dass die so gewonnenen 100 Zahlen in etwa dem Benfordschen Gesetz genügen sollten. Tabelle 2 zeigt die Ergebnisse einer solchen Simulation für $N = 10$ und Abbildung 11 visualisiert die Resultate.

| $d$ | $d_1\left(\prod_{k=1}^{10} X_k\right) = d$ | $\log\left(1+\frac{1}{d}\right)\cdot 100$ | $\left\lvert d_1\left(\prod_{k=1}^{10} X_k\right) - \log\left(1+\frac{1}{d}\right)\cdot 100\right\rvert$ |
|---|---|---|---|
| 1 | 30 | 30,10 | 0,10 |
| 2 | 17 | 17,60 | 0,60 |
| 3 | 14 | 12,50 | 1,50 |
| 4 | 10 | 9,69 | 0,31 |
| 5 | 9 | 7,92 | 1,08 |
| 6 | 6 | 6,69 | 0,69 |
| 7 | 8 | 5,80 | 2,20 |
| 8 | 6 | 5,12 | 0,88 |
| 9 | 0 | 4,58 | 4,58 |
| $\sum$ | 100 | 100 | – |

Tabelle 2: signifikante Anfangsziffern des Produktes der Mersenne-Twister-Pseudo-Zufallszahlen für $N = 10$.

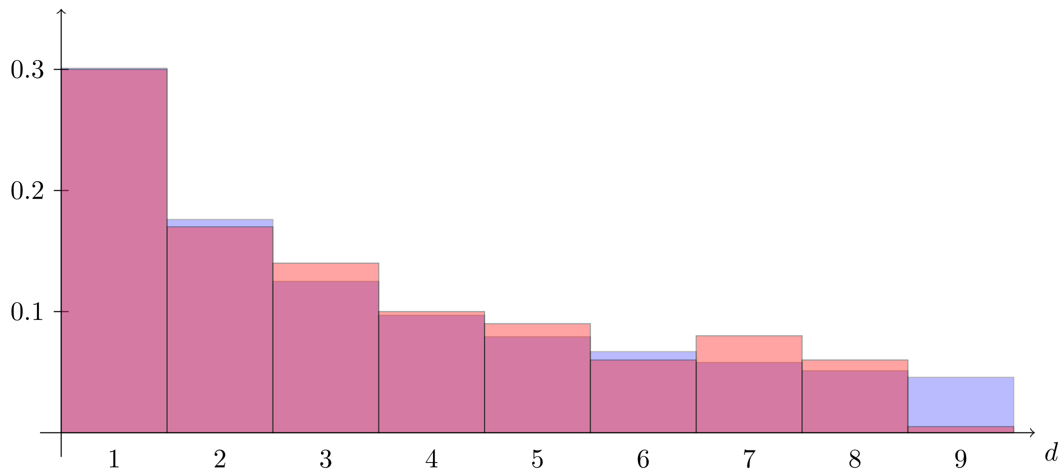


Abbildung 11: Gegenüberstellung des Benfordschen Gesetzes (blau) mit den Ergebnissen der Simulation (rot).

Trotz des relativ kleinen $N$ erkennt man schon eine deutliche Ähnlichkeit mit dem empirischen Benfordschen Gesetz (1.1). Für deutlich größere Werte von $N$ sind im Hinblick auf Satz 3.2 noch bessere Resultate zu erwarten. Beispiel 3.2 zeigt, dass die Konvergenzgeschwindigkeit für große $N$ in etwa $1/N$ ist.

**Satz 3.3.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$. Dann sind folgende Aussagen äquivalent:*

(i) $(X^n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung gegen das Benfordsche Gesetz.*

(ii) *Es gilt* $\lim_{n\to\infty} \mathrm{E}\left[\mathrm{e}^{\mathrm{i}2\pi n \log|X|}\right] = 0$.

Beweis.

***(ii)⇒(i):*** Sei für $n \in \mathbb{N}$

$$Y_n := \log S\left(X^n\right).$$

Es gibt nun ein eindeutiges $\ell \in \mathbb{Z}$, so dass $S\left(X^n\right) = 10^{-\ell}\left|X\right|^n$. Dann gilt

$$Y_n = \log\left(10^{-\ell}\left|X\right|^n\right) = n \cdot \log|X| - \ell$$

und da $0 \le Y_n < 1$ und $\ell \in \mathbb{Z}$, gilt per Definition von $\langle \cdot \rangle$

$$Y_n = \langle n \cdot \log |X| \rangle .$$

Die Folge $(X^n)_{n\in\mathbb{N}}$ konvergiert nun in Verteilung gegen das Benfordsche Gesetz, falls für alle $0 \le s < 1$

$$\begin{aligned} P\left(\{\langle n \cdot \log |X|\rangle \le s\}\right) &= P\left(\{\log S\left(X^n\right) \le s\}\right) \\ &= P\left(\{S\left(X^n\right) \le 10^s\}\right) \xrightarrow{n\to+\infty} \log 10^s = s, \end{aligned}$$

das heißt falls die Folge $\left(\langle n \cdot \log |X|\rangle\right)_{n\in\mathbb{N}}$ in Verteilung gegen eine auf $[0,1)$ gleichverteilte Zufallsgröße konvergiert. Nach Teil *(c)* von Lemma 2.3 ist dies genau dann der Fall, wenn

$$\widehat{P_{\langle n\cdot\log|X|\rangle}}(k) \xrightarrow{n\to+\infty} \widehat{\lambda\!\!\lambda_{0,1}}(k), \qquad \text{für alle } k \in \mathbb{Z}.$$

Wegen *(ii)* gilt nun für jedes $k \in \mathbb{N}$

$$\begin{aligned} 0 = \lim_{n\to\infty} \mathrm{E}\left[\mathrm{e}^{\mathrm{i}2\pi k n \log|X|}\right] &= \lim_{n\to\infty} \int\limits_\Omega \mathrm{e}^{\mathrm{i}2\pi k n \log|X|}\,\mathrm{d}P \\ &= \lim_{n\to\infty} \int\limits_0^1 \mathrm{e}^{\mathrm{i}2\pi k s}\, P_{\langle n\log|X|\rangle}(\mathrm{d}s) = \lim_{n\to\infty} \widehat{P_{\langle n\log|X|\rangle}}(k). \end{aligned}$$

Wegen $\overline{\widehat{P_{\langle n\log|X|\rangle}}(k)} = \widehat{P_{\langle n\log|X|\rangle}}(-k)$ gilt zudem auch für jedes $k \in \mathbb{N}$

$$\lim_{n\to\infty} \widehat{P_{\langle n\log|X|\rangle}}(-k) = 0.$$

Da stets $\widehat{P_{\langle n\log|X|\rangle}}(0) = \widehat{\lambda\!\!\lambda_{0,1}}(0) = 1$, haben wir somit

$$\widehat{P_{\langle n\log|X|\rangle}}(k) \xrightarrow{n\to+\infty} \widehat{\lambda\!\!\lambda_{0,1}}(k), \qquad \text{für alle } k \in \mathbb{Z}.$$

***(i)⇒(ii)*:** Angenommen *(ii)* gilt nicht, das heißt

$$\limsup_{n\to\infty} \left|\mathrm{E}\left[\mathrm{e}^{\mathrm{i}2\pi n \log|X|}\right]\right| > 0.$$

Dann gibt es eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ aus $\mathbb{N}$ und ein $\delta > 0$, so dass $\left|\mathrm{E}\left[\mathrm{e}^{\mathrm{i}2\pi n_k \log|X|}\right]\right| \ge \delta$ für alle $k \in \mathbb{N}$. Dann haben wir jedoch mit den Bezeichnungen wie vorher

$$\left|\widehat{P_{Y_{n_k}}}(1)\right| = \left|\widehat{P_{\langle n_k\log|X|\rangle}}(1)\right| = \left|\mathrm{E}\left[\mathrm{e}^{\mathrm{i}2\pi n_k \log|X|}\right]\right| \ge \delta,$$

so dass $\langle n_k \log |X|\rangle$ nicht in Verteilung gegen eine auf $[0,1)$ gleichverteilte Zufallsgröße konvergiert, so dass *(i)* nicht gelten kann. □

Mittels Satz 3.3 erhalten wir nun als direkte Folgerung eine vollständige Charakterisierung der Konvergenz in Verteilung von $(X^n)_{n\in\mathbb{N}}$ gegen das Benfordsche Gesetz mittels Fourier-Koeffizienten:

**Folgerung 3.1.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum* $(\Omega, \Sigma, P)$. *Dann sind folgende Aussagen äquivalent:*

(i) $(X^n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung gegen das Benfordsche Gesetz.*

(ii) *Es gilt* $\widehat{P_{\langle\log|X|\rangle}}(n) \xrightarrow{n\to+\infty} 0.$[11]

[11] Wahrscheinlichkeitsmaße $\mu$ mit der Eigenschaft $\widehat{\mu}(n) \xrightarrow{n\to+\infty} 0$ heißen *Rajchman-Maße*.

Beweis. In (2.6) haben wir gesehen, dass für die Fourier-Koeffizienten mit einer Zufallsgröße $Y$ stets gilt

$$\widehat{P_{\langle nY\rangle}}(k) = \widehat{P_{\langle Y\rangle}}(n\cdot k), \qquad \text{für alle } k\in\mathbb{Z} \text{ und alle } n\in\mathbb{N}.$$

Im Beweis von Satz 3.3 haben wir zudem gesehen, dass *(i)* genau dann gilt, wenn für alle $k\in\mathbb{N}$ gilt $\widehat{P_{\langle n\log|X|\rangle}}(k) \xrightarrow{n\to+\infty} 0$. Kombiniert man diese beiden Tatsachen, so gilt *(i)* genau dann, wenn

$$\widehat{P_{\langle\log|X|\rangle}}(k\cdot n) \xrightarrow{n\to+\infty} 0, \qquad \text{für alle } k\in\mathbb{N},$$

was aber äquivalent ist zu

$$\widehat{P_{\langle\log|X|\rangle}}(n) \xrightarrow{n\to+\infty} 0.$$

□

## 3.2 Unabhängige Zufallsgrößen

Auch in diesem Abschnitt interessieren wir uns für das asymptotische Verhalten der Folge der Produkte $\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}}$ von Zufallsgrößen, welche unabhängig sind.

Zunächst erhalten wir ein Resultat für zwei Zufallsgrößen:

**Satz 3.4.** *Seien $X$ und $Y$ unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega,\Sigma,P)$ mit $P(\{X\cdot Y=0\})=0$. Dann gilt:*

(a) *Wenn $X$ Benford ist, so ist auch die Zufallsgröße $X\cdot Y$ Benford.*

(b) *Wenn $S(X)$ und $S(X\cdot Y)$ dieselbe Verteilung besitzen, so ist entweder $X$ Benford oder es gilt $P\left(\left\{\log S(Y)\in\frac{1}{m}\mathbb{Z}\right\}\right)=1$ für ein $m\in\mathbb{Z}\setminus\{0\}$.*

Beweis.

**Zu *(a)*:** Wir beweisen die Aussage mithilfe von Fourier-Koeffizienten. Wenn also $X$ Benford ist, so ist gilt nach Satz 2.2

$$P_{\langle\log|X|\rangle} = \lambda\!\!\lambda_{0,1}.$$

Es gilt nun mit $\ell := \lfloor\log|X|\rfloor\in\mathbb{Z}$

$$\begin{aligned}\langle\log|X|\rangle &= \log|X| - \lfloor\log|X|\rfloor = \log|X|-\ell = \log|X|-\ell\cdot\log 10\\ &= \log|X| - \log 10^{\ell} = \log\left(10^{-\ell}|X|\right).\end{aligned}$$

Da $0\le\langle\log|X|\rangle<1$, gilt dann $1\le 10^{-\ell}|X|<10$, das heißt nach Definition der Signifikanten $S(X)=10^{-\ell}|X|$, also

$$\langle\log|X|\rangle = \log S(X).$$

Somit ist die Benford-Eigenschaft von $X$ äquivalent zu $P_{\log S(X)} = \lambda\!\!\lambda_{0,1}$ oder mittels Fourier-Koeffizienten ausgedrückt

$$\widehat{P_{\log S(X)}}(k) = \widehat{\lambda\!\!\lambda_{0,1}}(k), \qquad \text{für alle } k\in\mathbb{Z}.$$

Analog zeigt man $\langle\log|X\cdot Y|\rangle = \log S(X\cdot Y)$. Da $P(\{X\cdot Y=0\})=0$, gilt

$$0 = P(\{X\cdot Y=0\}) = P(\{X=0\}\cup\{Y=0\}) \ge P(\{X=0\}) \ge 0,$$

also $P(\{X=0\})=0$ und analog $P(\{Y=0\})=0$. Daher gilt wegen $\log 0 := 0$ auch $P(\{\log S(X)=0\}) = P(\{\log S(Y)=0\}) = 0$. Zusammenfassend gilt dann also $P$-fast-sicher die folgende Gleichungskette

$$\log S(X\cdot Y) = \langle\log|X\cdot Y|\rangle = \langle\log|X|+\log|Y|\rangle = \langle\log S(X)+\log S(Y)\rangle. \tag{3.2}$$

Da $X$ und $Y$ unabhängig sind, gilt dies auch für $\log S(X)$ und $\log S(Y)$ und Teil *(b)* von Lemma 2.3 liefert

$$\begin{aligned}\widehat{P_{\log S(X\cdot Y)}}(k) &= \widehat{P_{\langle \log S(X)+\log S(Y)\rangle}}(k) \\ &= \widehat{P_{\log S(X)}}(k)\cdot \widehat{P_{\log S(Y)}}(k), \qquad \text{für alle } k\in\mathbb{Z}. \qquad (3.3)\\ &= \widehat{\lambda\!\!\lambda_{0,1}}(k)\cdot \widehat{P_{\log S(Y)}}(k),\end{aligned}$$

Da nun $\widehat{\lambda\!\!\lambda_{0,1}}(0)=1$ und sonst $\widehat{\lambda\!\!\lambda_{0,1}}(k)=0$ erhalten wir aus (3.3) dann $\widehat{P_{\log S(X\cdot Y)}}(0)=1$ und $\widehat{P_{\log S(X\cdot Y)}}(k)=0$ für alle $k\in\mathbb{Z}$ mit $k\neq 0$. Also gilt wieder $\widehat{P_{\log S(X\cdot Y)}}(k)=\widehat{\lambda\!\!\lambda_{0,1}}(k)$ für alle $k\in\mathbb{Z}$ und damit

$$P_{\langle \log S(X\cdot Y)\rangle} = \lambda\!\!\lambda_{0,1}.$$

Dies ist dann wieder nach Satz 2.2 äquivalent zur Benford-Eigenschaft von $X\cdot Y$.

**Zu *(b)*:** Wenn $S(X)$ und $S(X\cdot Y)$ dieselbe Verteilung besitzen – das heißt also $P_{S(X)}=P_{S(X\cdot Y)}$ – so gilt natürlich auch $P_{\log S(X)}=P_{\log S(X\cdot Y)}$ und (3.3) liefert

$$\widehat{P_{\log S(X)}}(k)=\widehat{P_{\log S(X)}}(k)\cdot \widehat{P_{\log S(Y)}}(k), \qquad \text{für alle } k\in\mathbb{Z}.$$

Also ist

$$\widehat{P_{\log S(X)}}(k)\cdot\Big(1-\widehat{P_{\log S(Y)}}(k)\Big)=0, \qquad \text{für alle } k\in\mathbb{Z}.$$

**Fall 1:** Falls $\widehat{P_{\log S(Y)}}(k)\neq 1$ für alle $k\neq 0$, so muss $\widehat{P_{\log S(X)}}(k)=0$ für alle $k\neq 0$ gelten. Es ist dann wieder wegen $\widehat{P_{\log S(X)}}(0)=\widehat{\lambda\!\!\lambda_{0,1}}(0)=1$

$$\widehat{P_{\log S(X)}}(k)=\widehat{\lambda\!\!\lambda_{0,1}}(k), \qquad \text{für alle } k\in\mathbb{Z}.$$

Wir haben also $P_{\log S(X)}=\lambda\!\!\lambda_{0,1}$ und wegen Satz 2.2 ist daher $X$ Benford.

**Fall 2:** Ist andererseits $\widehat{P_{\log S(Y)}}(k_0)=1$ für ein $k_0\neq 0$, so gilt

$$\widehat{P_{\log S(Y)}}(k_0)=\mathrm{e}^{\mathrm{i}2\pi\cdot 0}.$$

Im Beweis von Teil *(c)* von Satz 2.5 wurde gezeigt[12], dass dann gilt

$$P\left(\{k_0\log S(Y)=0+\mathbb{Z}\}\right)=1.$$

Daraus ergibt sich schließlich $P\left(\left\{\log S(Y)=\frac{1}{k_0}\mathbb{Z}\right\}\right)=1.$ □

*Bemerkung.* In Satz 3.4 gibt es die Voraussetzung $P(\{X\cdot Y=0\})=0$. Das man diese Voraussetzung wirklich benötigt erkennt man an folgendem Beispiel: Wäre zum Beispiel $X$ Benford und $Y$ $P$-fast-sicher konstant 0, so wären $X$ und $Y$ unabhängig, aber $X\cdot Y$ wäre ebenfalls $P$-fast-sicher konstant 0, so dass $X\cdot Y$ nicht Benford ist. ♣

Die Aussage von Teil *(a)* von Satz 3.4 ist dahingehend überraschend, da über die Verteilung von $Y$ *keinerlei* Voraussetzungen außer $P(\{X\cdot Y=0\})=0$ gemacht werden. Lediglich die stochastische Unabhängigkeit ist eine Voraussetzung die *nicht* ohne Weiteres wegfallen darf. Denn ist $U$ eine auf $[0,1)$ gleichverteilte Zufallsgröße, so ist die Zufallsgröße $X:=10^U$ nach Beispiel 2.2 Benford. Da die Zufallsgröße $1-U$ ebenfalls gleichverteilt auf $[0,1)$ ist, ist die $Y:=10^{1-U}$ ebenfalls Benford. Jedoch ist $X\cdot Y=10^U\cdot 10^{1-U}=10$ *nicht* Benford, obwohl *beide* Zufallsgrößen Benford waren. Dieses Beispiel zeigt, dass die Voraussetzung der stochastischen Unabhängigkeit in Satz 3.4 nicht wegfallen darf.

[12] in diesem Fall ist mit den Bezeichnungen aus dem Beweis von Satz 2.5 $\xi=0$.

Im Hinblick auf die Thematik der skaleninvarianten signifikanten Dezimalziffern aus Abschnitt 2.3 erhalten wir noch folgende interessante Eigenschaft von Benford-Zufallsgrößen: Sei $X$ eine Benford-Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ und sei für ein $\alpha > 0$ dann $Y := \alpha$ die konstante Zufallsgröße mit Wert $\alpha$. Nach Lemma 2.4 sind dann $X$ und $Y$ unabhängig und es gilt $P(\{X \cdot Y\} = 0) = P(\{\alpha X = 0\}) = P(\{X = 0\}) = 0$, also gilt nach Teil *(a)* von Satz 3.4, dass auch $Y \cdot X = \alpha X$ Benford ist, das heißt es gilt

$$P(\{S(X) \leq t\}) = P(\{S(\alpha X) \leq t\}) = \log t, \qquad \text{für alle } 1 \leq t < 10.$$

Eine weitere interessante Folgerung aus Satz 3.4 ist:

**Folgerung 3.2.** *Sei $X$ eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$ mit $P(\{X = 0\}) = 0$. Gilt für ein $\alpha \in \mathbb{R}$ mit der Eigenschaft $\log|\alpha| \in \mathbb{R}\backslash\mathbb{Q}$, dass $S(X)$ und $S(\alpha X)$ dieselbe Verteilung besitzen, so ist $X$ Benford.*

Beweis. Die Zufallsgröße $Y = \alpha$ ist nach Lemma 2.4 unabhängig von der Zufallsgröße $X$. Nach Voraussetzung ist dann auch $P(\{\alpha X = 0\}) = P(\{X \cdot Y = 0\}) = 0$. Nach Teil *(b)* von Satz 3.4 ist dann entweder $P\left(\left\{\log S(\alpha) \in \frac{1}{m}\mathbb{Z}\right\}\right) = 1$ für ein $m \in \mathbb{Z}$ oder $X$ ist Benford.

Falls für ein $m \in \mathbb{Z}$ $P\left(\left\{\log S(\alpha) \in \frac{1}{m}\mathbb{Z}\right\}\right) = 1$, ist wegen $P\left(\left\{\log S(\alpha) = \frac{1}{m}\mathbb{Z}\right\}\right) \leq P(\{\log S(\alpha) \in \mathbb{Q}\})$ dann $P(\{\log S(\alpha) \in \mathbb{Q}\}) = 1$. Daher gilt entweder $P(\{\log S(\alpha) \in \mathbb{Q}\}) = 1$ oder $X$ ist Benford. Angenommen $P(\{\log S(\alpha) \in \mathbb{Q}\}) = 1$. Im Beweis von Satz 3.4 wurde gezeigt, dass die Beziehung $\log S(Y) = \langle \log|Y| \rangle$ besteht. Angewandt auf $Y = \alpha$ haben wir also

$$\log S(\alpha) = \langle \log|\alpha| \rangle.$$

Da $\langle \log|\alpha| \rangle = \log|\alpha| - \lfloor \log|\alpha| \rfloor$ ist $\langle \log|\alpha| \rangle$ rational genau dann, wenn $\log|\alpha|$ rational ist. Nach Voraussetzung ist $\log|\alpha|$ irrational, also haben wir $\{\log S(\alpha) \in \mathbb{Q}\} = \{\log|\alpha| \in \mathbb{Q}\} = \varnothing$, so dass $P(\{\log S(\alpha) \in \mathbb{Q}\}) = 0$ gilt. Somit muss $X$ Benford sein. □

Wir betrachten nun eine Folge stochastisch unabhängiger Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$. In Fortführung der Aussage von Teil *(a)* von Satz 3.4 ist das endliche Produkt $\prod_{k=1}^{n} X_k$ Benford, wenn mindestens einer der Faktoren $X_1, \ldots, X_n$ selbst Benford ist. Sind die Zufallsgrößen $X_1, \ldots, X_n$ auch noch identisch verteilt, so müsste also *jede* Zufallsgröße Benford sein damit $\prod_{k=1}^{n} X_k$ selbst Benford ist. Viel sinnvoller erscheint es in diesem Falle asymptotische Betrachtungen anzustellen – das heißt wir möchten $\prod_{k=1}^{n} X_k$ für $n \to \infty$ auf die Konvergenz in Verteilung gegen das Benfordsche Gesetz untersuchen.

Für das Hauptresultat dieser Thematik, benötigen wir noch einige Begriffsbildungen aus der Ergodentheorie:

**Definition 3.1.** Sei $(\Omega, \Sigma, P)$ ein Wahrscheinlichkeitsraum und $T\colon \Omega \longrightarrow \Omega$ eine messbare Abbildung.

*(a)* Das Tripel $(\Omega, \Sigma, T, P)$ heißt ein *maßtheoretisches dynamisches System*.

*(b)* Die Abbildung $T$ heißt *maßerhaltend* oder $P$ heißt $T$*-invariant*, falls $P = P_T$.

*(c)* Eine Menge $A \in \Sigma$ heißt $T$*-invariant*, falls $T^{-1}(A) = A$ (bis auf eine $P$ Nullmenge). Die Menge aller $T$-invarianten $A \in \Sigma$ bezeichnen wir mit $\mathcal{T}(\Sigma) \subseteq \Sigma$.

*(d)* Das maßtheoretische dynamische System $(\Omega, \Sigma, T, P)$ heißt *ergodisch*, falls die Abbildung $T$ maßerhaltend ist und für alle $T$-invarianten Mengen $A \in \Sigma$ gilt $P(A) = 0$ oder $P(\Omega\backslash A) = 0$.

Wir geben nun zwei Resultate ohne Beweis an, welche wir für den Beweis des Hauptresultates benötigen:

**Satz 3.5** (Bedingte Erwartung). *Sei $(\Omega, \Sigma, P)$ ein Wahrscheinlichkeitsraum und $X$ eine Zufallsgröße aus $\mathscr{L}^1(\Omega, \Sigma, P)$. Sei $\Sigma' \subseteq \Sigma$ eine Unter-$\sigma$-Algebra. Dann gibt es eine Zufallsgröße $Z$ auf $(\Omega, \Sigma, P)$, welche die folgenden Eigenschaften besitzt:*

(I) *$Z$ ist $\Sigma'$-messbar, das heißt für alle $B \in \mathfrak{B}$ ist $Z^{-1}(B) \in \Sigma'$.*

(II) *Für jedes $A \in \Sigma'$ gilt*
$$\int_\Omega 1_A Z \mathrm{d}P = \int_\Omega 1_A X \mathrm{d}P.$$

*Die Zufallsgröße $Z$ ist $P$-fast-eindeutig bestimmt.*

Beweis. Siehe zum Beispiel [13] Kapitel 8. ◊

Insbesondere aufgrund der $P$-fast-sicheren Eindeutigkeit ist die folgende Definition gerechtfertigt:

**Definition 3.2** (Bedingte Erwartung)**.** Die Zufallsgröße $Z$ aus Satz 3.5 nennen wir *bedingte Erwartung* von $X$ bezüglich $\Sigma'$ und schreiben hierfür

$$Z = \mathrm{E}\left[X \mid \Sigma'\right].$$

Der folgende Satz ist einer der Hauptsätze der Ergodentheorie:

**Satz 3.6** (Birkhoffscher Ergodensatz)**.** *Sei $(\Omega, \Sigma, T, P)$ ein maßtheoretisches dynamisches System, $T$ maßerhaltend und $Y$ eine Zufallsgröße aus $\mathscr{L}^1(\Omega, \Sigma, P)$. Dann gilt*

$$\frac{1}{N}\sum_{n=0}^{N-1} Y \circ T^n \xrightarrow{N \to +\infty} \mathrm{E}\left[Y \mid \mathcal{T}(\Sigma)\right], \qquad P\text{-fast-sicher.}$$

*Ist das maßtheoretische dynamische System sogar ergodisch, so gilt*

$$\frac{1}{N}\sum_{n=0}^{N-1} Y \circ T^n \xrightarrow{N \to +\infty} \mathrm{E}\left[Y\right], \qquad P\text{-fast-sicher.}$$

Beweis. Siehe zum Beispiel [13] Kapitel 20. ◊

*Bemerkung.* Ausführlich geschrieben bedeutet die Aussage aus dem Birkhoffschen Ergodensatz gerade

$$\frac{1}{N}\sum_{n=0}^{N-1} Y\left(T^n(\omega)\right) \xrightarrow{N \to +\infty} \mathrm{E}\left[Y \mid \mathcal{T}(\Sigma)\right](\omega), \qquad \text{für } P\text{-fast-alle } \omega \in \Omega,$$

und für ergodische Systeme

$$\frac{1}{N}\sum_{n=0}^{N-1} Y\left(T^n(\omega)\right) \xrightarrow{N \to +\infty} \int_\Omega Y \mathrm{d}P, \qquad \text{für } P\text{-fast-alle } \omega \in \Omega.$$

Das heißt falls Ergodizität vorliegt ist der Grenzwert von $\frac{1}{N}\sum_{n=0}^{N-1} Y\left(T^n(\omega)\right)$ $P$-fast-überall konstant. ♣

**Folgerung 3.3** (Birkhoffscher Ergodensatz in $\mathbb{C}$)**.** *Sei $(\Omega, \Sigma, T, P)$ ein maßtheoretisches dynamisches System, $T$ maßerhaltend und $f \colon \Omega \longrightarrow \mathbb{C}$ eine Abbildung aus $\mathscr{L}^1(\Omega, \Sigma, P)$. Dann gilt*

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \xrightarrow{N \to +\infty} \mathrm{E}\left[f \mid \mathcal{T}(\Sigma)\right], \qquad P\text{-fast-sicher.}$$

*Ist das maßtheoretische dynamische System sogar ergodisch, so gilt*

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \xrightarrow{N \to +\infty} \mathrm{E}\left[f\right], \qquad P\text{-fast-sicher.}$$

Beweis. Wegen $f = \Re f + \mathrm{i}\Im f$ erfolgt der Beweis mittels Satz 3.6 durch Zerlegung von $f$ in Real- und Imaginärteil. □

Als nächstes erhalten wir nun eine nützliche Charakterisierung der Ergodizität von maßtheoretischen dynamischen Systemen:

**Satz 3.7.** *Sei $(\Omega, \Sigma, T, P)$ ein maßtheoretisches dynamisches System. Für $1 \leq p < +\infty$ sei der lineare Operator $\mathtt{T} \colon \mathscr{L}^p(\Omega, \Sigma, P) \longrightarrow \mathscr{L}^p(\Omega, \Sigma, P)$ definiert gemäß*

$$(\mathtt{T}f)(\omega) := f(T(\omega)).$$

*Ferner sei* $\operatorname{Fix}\mathtt{T} := \{f \in \mathscr{L}^p(\Omega, \Sigma, P) \mid \mathtt{T}f = f\}$. *Dann sind folgende Aussagen äquivalent:*

(i) $(\Omega, \Sigma, T, P)$ *ist ergodisch.*

(ii) $\operatorname{Fix}\mathtt{T}$ *besteht nur aus $P$-fast-überall konstanten Funktionen.*

Beweis.

**„*(i)⇒(ii)*":** Sei $\mathtt{T}f = f$. Dann definiere für $\alpha \in \mathbb{R}$

$$A := \{\omega \in \Omega \mid f(\omega) \leq \alpha\} = \{f \leq \alpha\}.$$

Dann gilt wegen $\mathtt{T}f = f$ bzw. $(\mathtt{T}f)(\omega) = f(\omega)$ für $P$-fast-alle $\omega \in \Omega$

$$\begin{aligned} T^{-1}(A) &= \{\omega \in \Omega \mid T(\omega) \in A\} = \{\omega \in \Omega \mid f(T(\omega)) \leq \alpha\} \\ &= \{\omega \in \Omega \mid (\mathtt{T}f)(\omega) \leq \alpha\} \doteq \{\omega \in \Omega \mid f(\omega) \leq \alpha\} \\ &= A, \end{aligned}$$

wobei im Weiteren per Definition $C \doteq D$ genau dann, wenn $C = D$ bis auf eine $P$-Nullmenge gilt. Also haben wir $T^{-1}(A) \doteq A$ und $A$ ist eine $T$-invariante Menge. Wegen der Ergodizität von $(\Omega, \Sigma, T, P)$ gilt also $P(A) = 0$ oder $P(A) = 1$. Das heißt

$$P(\{f \leq \alpha\}) = 0 \qquad \text{oder} \qquad P(\{f \leq \alpha\}) = 1.$$

Damit muss $f$ $P$-fast-überall konstant sein.

**„*(ii)⇒(i)*":** Sei $A$ eine $T$-invariante Menge, das heißt

$$T^{-1}(A) \doteq A.$$

Dann gilt

$$(\mathtt{T}1_A)(\omega) = 1_A(T(\omega)) = 1_{T^{-1}(A)}(\omega) = 1_A(\omega), \qquad P\text{-fast-überall.}$$

Also ist $1_A \in \operatorname{Fix}\mathtt{T}$ und wegen *(ii)* muss dann $1_A(\omega) = 1$ für $P$-fast-alle $\omega \in \Omega$ oder $1_A(\omega) = 0$ für $P$-fast-alle $\omega \in \Omega$ gelten. Damit ist

$$P(A) = \int_\Omega 1_A \mathrm{d}P = 1 \qquad \text{oder} \qquad P(A) = \int_\Omega 1_A \mathrm{d}P = 0,$$

was gerade die Ergodizität von $(\Omega, \Sigma, T, P)$ bedeutet. □

Es folgt nun ein weiterer wichtiger Satz aus der Ergodentheorie, welcher sich als wesentlich für den Beweis des Satzes 3.9 erweisen wird: Seien $(Y, \mathfrak{Y}, \eta)$ und $(S, \mathfrak{S}, \nu)$ Wahrscheinlichkeitsräume. Ferner sei für jedes $s \in S$ $\varphi_s \colon Y \longrightarrow Y$ eine $\eta$-maßerhaltende Abbildung, das heißt $\eta_{\varphi_s} = \eta$. Wir haben also eine Familie $(\varphi_s)_{s\in S}$ von $\eta$-maßerhaltenden Transformationen auf $Y$.

Sei nun

$$\begin{aligned} X &:= Y \times S \times S \times \ldots, \\ \Sigma &:= \mathfrak{Y} \otimes \mathfrak{S} \otimes \mathfrak{S} \otimes \ldots, \\ \mu &:= \eta \otimes \nu \otimes \nu \otimes \ldots. \end{aligned}$$

Ferner sei $T\colon X \longrightarrow X$ definiert gemäß

$$T(y, s_1, s_2, \ldots) \coloneqq (\varphi_{s_1}(y), s_2, s_3, \ldots).$$

Dann erhalten wir folgenden

**Satz 3.8.** *Seien $(X, \Sigma, T, \mu)$ wie gerade definiert und $g \in \mathscr{L}^1(X, \Sigma, \mu)$ $\mu$-fast-überall $T$-invariant, also $g(T(y, s_1, s_2, \ldots)) = g(y, s_1, s_2, \ldots)$ für $\mu$-fast-alle $(y, s_1, s_2, \ldots) \in X$. Dann gibt es eine Funktion $f\colon Y \longrightarrow \mathbb{R}$, so dass $\mu$-fast-überall gilt*

$$g(y, s_1, s_2, \ldots) = f(y).$$

*Insbesondere gilt für $\nu$-fast-alle $s \in S$ und $\eta$-fast-alle $y \in Y$*

$$f(y) = f(\varphi_s(y)).$$

Bevor wir Satz 3.8 beweisen, benötigen wir noch folgendes Hilfsresultat:

**Lemma 3.2.** *Seien $(X, \Sigma, \mu)$ und $(Y, \Sigma', \nu)$ Wahrscheinlichkeitsräume und sei ferner $g \in \mathscr{L}^1(X \times Y, \Sigma \otimes \Sigma', \mu \otimes \nu)$, das heißt*

$$\int_{X \times Y} |g| \,\mathrm{d}(\mu \otimes \nu) < +\infty.$$

*Falls nun für jedes $h \in \mathscr{L}^\infty(Y, \Sigma', \nu)$ mit $\int_Y h(y)\,\nu(\mathrm{d}y) = 0$ gilt*

$$\int_Y g(x, y)\, h(y)\,\nu(\mathrm{d}y) = 0, \qquad \text{für } \mu\text{-fast-alle } x \in X,$$

*so gibt es ein $g^*\colon X \longrightarrow \mathbb{R}$, so dass für $\mu$-fast-alle $x \in X$ und $\nu$-fast-alle $y \in Y$ gilt*

$$g(x, y) = g^*(x).$$

Beweis. Sei $H \in \mathscr{L}^\infty(Y, \Sigma', \nu)$ beliebig. Zunächst gilt

$$\begin{aligned}
\int_Y \left( H(y) - \int_Y H(y')\,\nu(\mathrm{d}y') \right) \nu(\mathrm{d}y) &= \int_Y H(y)\,\nu(\mathrm{d}y) - \int_Y \left( \int_Y H(y')\,\nu(\mathrm{d}y') \right) \nu(\mathrm{d}y) \\
&= \int_Y H(y)\,\nu(\mathrm{d}y) - \int_Y H(y')\,\nu(\mathrm{d}y') \cdot \int_Y 1\nu(\mathrm{d}y) \\
&= \int_Y H(y)\,\nu(\mathrm{d}y) - \int_Y H(y')\,\nu(\mathrm{d}y') = 0.
\end{aligned}$$

Also ist $h(y) \coloneqq H(y) - \int_Y H(y')\,\nu(\mathrm{d}y')$ eine Funktion aus $\mathscr{L}^\infty(Y, \Sigma', \nu)$ mit $\int_Y h(y)\,\nu(\mathrm{d}y) = 0$ und nach Voraussetzung gilt dann

$$\begin{aligned}
0 = \int_Y g(x, y)\, h(y)\,\nu(\mathrm{d}y) &= \int_Y g(x, y) \cdot \left[ H(y) - \int_Y H(y')\,\nu(\mathrm{d}y') \right] \nu(\mathrm{d}y) \\
&= \int_Y g(x, y) \cdot H(y)\,\nu(\mathrm{d}y) - \int_Y g(x, y) \cdot \left( \int_Y H(y')\,\nu(\mathrm{d}y') \right) \nu(\mathrm{d}y) \\
&= \int_Y g(x, y) \cdot H(y)\,\nu(\mathrm{d}y) - \int_Y g(x, y)\,\nu(\mathrm{d}y) \cdot \int_Y H(y')\,\nu(\mathrm{d}y').
\end{aligned}$$

Es gilt daher

$$\int\limits_Y g(x,y)\cdot H(y)\,\nu(\mathrm{d}y) = \int\limits_Y g(x,y)\,\nu(\mathrm{d}y)\cdot\int\limits_Y H(y)\,\nu(\mathrm{d}y).$$

Sei nun noch $f \in \mathscr{L}^\infty(X,\Sigma,\mu)$. Im Folgenden schreiben wir zur Vereinfachung für $\mu(\mathrm{d}x)$ kurz $\mathrm{d}\mu$ und für $\nu(\mathrm{d}y)$ kurz $\mathrm{d}\nu$, da klar ist, dass die Integration in $X$ bezüglich $\mu$ erfolgt und die Integration in $Y$ bezüglich $\nu$ erfolgt. Dann ergibt sich nun mittels des Satzes von Fubini

$$\begin{aligned}
0 &= \int\limits_X 0\cdot f(x)\,\mathrm{d}\mu = \int\limits_X \left[\int\limits_Y g(x,y)\,H(y)\,\mathrm{d}\nu - \int\limits_Y g(x,y)\,\mathrm{d}\nu\cdot\int\limits_Y H(y)\,\mathrm{d}\nu\right] f(x)\,\mathrm{d}\mu\\
&= \int\limits_X\left(\int\limits_Y g(x,y)\,H(y)\,\mathrm{d}\nu\right) f(x)\,\mathrm{d}\mu - \int\limits_X\left(\int\limits_Y g(x,y)\,\mathrm{d}\nu\cdot\int\limits_Y H(y)\,\mathrm{d}\nu\right) f(x)\,\mathrm{d}\mu\\
&= \int\limits_X\left(\int\limits_Y g(x,y)\,H(y)\,\mathrm{d}\nu\right) f(x)\,\mathrm{d}\mu - \int\limits_X\left(\int\limits_Y\left[\int\limits_Y g(x,y')\,\mathrm{d}\nu\right] H(y)\,\mathrm{d}\nu\right) f(x)\,\mathrm{d}\mu\\
&= \int\limits_X\int\limits_Y g(x,y)\,f(x)\,H(y)\,\mathrm{d}\mu\mathrm{d}\nu - \int\limits_X\int\limits_Y\left(\int\limits_Y g(x,y')\,\mathrm{d}\nu\right) f(x)\,H(y)\,\mathrm{d}\mu\mathrm{d}\nu\\
&= \int\limits_X\int\limits_Y\left[g(x,y) - \int\limits_Y g(x,y')\,\mathrm{d}\nu\right] f(x)\,H(y)\,\mathrm{d}\mu\mathrm{d}\nu.
\end{aligned}$$

Da $f$ und $H$ beliebig gewählt waren, folgt $g(x,y) - \int_Y g(x,y')\,\nu(\mathrm{d}y') = 0$ $\mu\otimes\nu$-fast-überall, also

$$g(x,y) = \int\limits_Y g(x,y)\,\nu(\mathrm{d}y), \qquad \mu\otimes\nu\text{-fast-überall.}$$

Mit $g^*(x) \coloneqq \int_Y g(x,y)\,\nu(\mathrm{d}y)$ ist die Behauptung gezeigt. □

Nun kommen wir zum Beweis des Satzes 3.8:

Beweis. Sei also $g \in \mathscr{L}^1(X,\Sigma,\mu)$ $\mu$-fast-überall $T$-invariant. Mit den Bezeichnungen aus Satz 3.7 können wir dies schreiben als $g \in \mathrm{Fix}\mathtt{T} = \left\{f \in \mathscr{L}^1(X,\Sigma,\mu)\,\middle|\,\mathtt{T}f = f\right\}$. In [9] wird gezeigt, dass $\mathscr{L}^\infty(X,\Sigma,\mu)\cap\mathrm{Fix}\mathtt{T}$ dicht in $\mathrm{Fix}\mathtt{T}$ ist, so dass wir ohne Beschränkung der Allgemeinheit $g \in \mathscr{L}^\infty(X,\Sigma,\mu)\cap\mathrm{Fix}\mathtt{T}$ annehmen können. Sei nun $h \in \mathscr{L}^\infty\left(S^{\mathbb{N}}, \bigotimes_{n=1}^\infty \mathfrak{S}, \bigotimes_{n=1}^\infty \nu\right)$ mit $|h(s_1,s_2,\ldots)| \le M$ $\bigotimes_{n=1}^\infty \nu$-fast-überall und

$$\int\limits_{S^{\mathbb{N}}} h(s_1,s_2,\ldots)\,\nu(\mathrm{d}s_1)\,\nu(s_2)\ldots = 0.$$

Nach Lemma 3.2 reicht es zu zeigen, dass für $H\colon Y \longrightarrow \mathbb{R}$ definiert gemäß

$$H(y) \coloneqq \int\limits_{S^{\mathbb{N}}} g(y,s_1,s_2,\ldots)\,h(s_1,s_2,\ldots)\,\nu(\mathrm{d}s_1)\,\nu(\mathrm{d}s_2)\ldots,$$

für $\eta$-fast-alle $y \in Y$ gilt

$$H(y) = 0.$$

Wegen $g(\varphi_{s_1}(y), s_2, s_3,\ldots) = g(y,s_1,s_2,\ldots)$, folgt induktiv für alle $n\in\mathbb{N}$ und $\nu$-fast-alle $s_1,\ldots,s_{n-1}\in S$

$$g\left(\left(\varphi_{s_{n-1}}\circ\varphi_{s_{n-2}}\circ\ldots\circ\varphi_{s_1}\right)(y), s_n, s_{n+1},\ldots\right) = g(y,s_1,s_2,\ldots). \tag{3.4}$$

Für $n \in \mathbb{N}$ sei nun $H_n \colon X \longrightarrow \mathbb{R}$ definiert gemäß

$$H_n(y, s_1, s_2, \ldots) := \int\limits_{S^{\mathbb{N}}} g(y, s_1, \ldots, s_{n-1}, t_n, t_{n+1}, \ldots) \, h(t_n, t_{n+1}, \ldots) \, \nu(\mathrm{d}t_n) \, \nu(\mathrm{d}t_{n+1}) \ldots .$$

Nun ist wegen (3.4)

$$\begin{aligned}
& H_n(y, s_1, s_2, \ldots) \\
=& \int\limits_{S^{\mathbb{N}}} g(y, s_1, \ldots, s_{n-1}, t_n, t_{n+1}, \ldots) \, h(t_n, t_{n+1}, \ldots) \, \nu(\mathrm{d}t_n) \, \nu(\mathrm{d}t_{n+1}) \ldots \\
=& \int\limits_{S^{\mathbb{N}}} g\left(\left(\varphi_{s_{n-1}} \circ \ldots \circ \varphi_{s_1}\right)(y), t_n, t_{n+1}, \ldots\right) h(t_n, t_{n+1}, \ldots) \, \nu(\mathrm{d}t_n) \, \nu(\mathrm{d}t_{n+1}) \ldots \\
=& H\left(\left(\varphi_{s_{n-1}} \circ \ldots \circ \varphi_{s_1}\right)(y)\right).
\end{aligned} \tag{3.5}$$

Sei $\varepsilon > 0$. Dann gibt es wegen

$$(X, \Sigma, \mu) = (Y \times S \times S \times \ldots, \mathfrak{Y} \otimes \mathfrak{S} \otimes \mathfrak{S} \otimes \ldots, \eta \otimes \nu \otimes \nu \otimes \ldots)$$

ein $N \in \mathbb{N}$ und eine Funktion $g_\varepsilon \colon Y \times \bigtimes_{k=1}^{N} S \longrightarrow \mathbb{R}$ mit

$$\int\limits_X \left| g(y, s_1, s_2, \ldots) - g_\varepsilon(y, s_1, s_2, \ldots, s_N) \right| \mu(\mathrm{d}y) \, \nu(\mathrm{d}s_1) \, \nu(\mathrm{d}s_2) \ldots < \frac{\varepsilon}{M}.$$

Damit ist für $n > N$

$$\begin{aligned}
& \int\limits_{S^{\mathbb{N}}} \left(g(y, s_1, s_2, \ldots) - g_\varepsilon(y, s_1, s_2, \ldots, s_N)\right) h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots \\
=& \int\limits_{S^{\mathbb{N}}} g(y, s_1, s_2, \ldots) \, h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots \\
& - \int\limits_{S^{\mathbb{N}}} g_\varepsilon(y, s_1, s_2, \ldots, s_N) \, h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots \\
=& \int\limits_{S^{\mathbb{N}}} g(y, s_1, s_2, \ldots) \, h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots \\
& - g_\varepsilon(y, s_1, \ldots, s_N) \cdot \underbrace{\int\limits_{S^{\mathbb{N}}} h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots}_{=0} \\
=& \int\limits_{S^{\mathbb{N}}} g(y, s_1, s_2, \ldots) \, h(s_n, s_{n+1}, \ldots) \, \nu(\mathrm{d}s_n) \, \nu(\mathrm{d}s_{n+1}) \ldots \\
=& H_n(y, s_1, s_2, \ldots).
\end{aligned}$$

Im Folgenden schreiben wir zur Abkürzung $\mathrm{d}s_k$ für $\nu(\mathrm{d}s_k)$, $\mathrm{d}t_k$ für $\nu(\mathrm{d}t_k)$ und $\mathrm{d}y$ für $\mu(\mathrm{d}y)$. Für $n > N$ haben wir dann mittels des Satzes von Fubini

$$\begin{aligned}
&\int\limits_X |H_n(y, s_1, s_2, ...)| \,\mathrm{d}y\mathrm{d}s_1\mathrm{d}s_2 ... \\
\leq &\int\limits_X \left( \int\limits_{S^{\mathbb{N}}} |g(y, s_1, ..., s_{n-1}, t_n, ...) - g_\varepsilon(y, s_1, s_2, ..., s_N)| \, |h(t_n, t_{n+1}, ...)| \,\mathrm{d}t_n ... \right) \mathrm{d}y\mathrm{d}s_1 ... \\
\leq M &\int\limits_X \left( \int\limits_{S^{\mathbb{N}}} |g(y, s_1, ..., s_{n-1}, t_n, ...) - g_\varepsilon(y, s_1, s_2, ..., s_N)| \,\mathrm{d}t_n ... \right) \mathrm{d}y\mathrm{d}s_1 ... \\
= M &\int\limits_{S^{\mathbb{N}}} \left( \int\limits_X |g(y, s_1, ..., s_{n-1}, t_n, ...) - g_\varepsilon(y, s_1, s_2, ..., s_N)| \,\mathrm{d}y\mathrm{d}s_1 ... \mathrm{d}s_{n-1}\mathrm{d}t_n ... \right) \mathrm{d}s_n ... \\
< M &\int\limits_{S^{\mathbb{N}}} \frac{\varepsilon}{M} \mathrm{d}s_n\mathrm{d}s_{n+1} ... = M \cdot \frac{\varepsilon}{M} = \varepsilon.
\end{aligned}$$

Da die Abbildung $\varphi_s$ für jedes $s \in S$ $\eta$-maßerhaltend ist, folgt dann nach dem Transformationssatz für die Integration bezüglich Bildmaßen

$$\int\limits_Y |H(y)| \,\eta(\mathrm{d}y) = \int\limits_Y |H(y)| \,\eta_{\varphi_s}(\mathrm{d}y) = \int\limits_Y |H(\varphi_s(y))| \,\eta(\mathrm{d}y).$$

Induktiv folgt also für jedes $n \in \mathbb{N}$ und alle $s_1, \ldots, s_{n-1} \in S$

$$\int\limits_Y |H(y)| \,\eta(\mathrm{d}y) = \int\limits_Y \left| H\left( \left( \varphi_{s_{n-1}} \circ \varphi_{s_{n-2}} \ldots \circ \varphi_{s_1} \right)(y) \right) \right| \eta(\mathrm{d}y)$$

und daher nach dem Satz von Fubini auch

$$\begin{aligned}
\int\limits_Y |H(y)| \,\eta(\mathrm{d}y) &= \left[ \int\limits_{S^{\mathbb{N}}} 1 \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots \right] \cdot \int\limits_Y |H(y)| \,\eta(\mathrm{d}y) \\
&= \left[ \int\limits_{S^{\mathbb{N}}} 1 \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots \right] \cdot \int\limits_Y \left| H\left( \left( \varphi_{s_{n-1}} \circ \varphi_{s_{n-2}} \ldots \circ \varphi_{s_1} \right)(y) \right) \right| \eta(\mathrm{d}y) \\
&= \int\limits_Y \left( \int\limits_{S^{\mathbb{N}}} \left| H\left( \left( \varphi_{s_{n-1}} \circ \varphi_{s_{n-2}} \ldots \circ \varphi_{s_1} \right)(y) \right) \right| \eta(\mathrm{d}y) \right) \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots \\
&= \int\limits_X \left| H\left( \left( \varphi_{s_{n-1}} \circ \varphi_{s_{n-2}} \ldots \circ \varphi_{s_1} \right)(y) \right) \right| \eta(\mathrm{d}y) \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots .
\end{aligned}$$

Daraus ergibt sich schließlich für $n > N$ wegen (3.5)

$$\begin{aligned}
\int\limits_Y |H(y)| \,\eta(\mathrm{d}y) &= \int\limits_X \left| H\left( \left( \varphi_{s_{n-1}} \circ \ldots \circ \varphi_{s_1} \right)(y) \right) \right| \eta(\mathrm{d}y) \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots \\
&= \int\limits_X |H_n(y, s_1, s_2, \ldots)| \,\eta(\mathrm{d}y) \nu(\mathrm{d}s_1) \nu(\mathrm{d}s_2) \ldots < \varepsilon.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig war, gilt also $H(y) = 0$ für $\eta$-fast-alle $y \in Y$. □

Nun haben wir alle Vorbereitungen für das Hauptresultat des vorliegenden Kapitels getroffen und erhalten

**Satz 3.9.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, P)$, so dass $X_1$ nicht nur abzählbar viele Werte annimmt, das heißt also $P(\{X_1 \in C\}) < 1$ für jedes abzählbare $C \in \mathfrak{B}$. Dann gilt:*

(a) *Die Folge $\left(\prod_{k=1}^n X_k\right)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen das Benfordsche Gesetz.*

(b) $P\left(\left\{\left(\prod_{k=1}^n X_k\right)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1.$

*Bemerkung.* Bevor wir Satz 3.9 beweisen, treffen wir noch eine kleine Vorbereitung zur Vereinfachung des Beweises. Bezeichne im Weiteren $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ die Menge der natürlichen Zahlen mit Null. Sei für $n \in \mathbb{N}$

$$Y_n := \log S(X_n) \in [0, 1).$$

Dann ist die Folge $(Y_n)_{n\in\mathbb{N}}$ immer noch unabhängig und identisch verteilt. Wir wollen nun zeigen, dass wir ohne Beschränkung der Allgemeinheit eine Zufallsgröße $Y_0$ auf $(\Omega, \Sigma, P)$ wählen können, so dass die Folge $(Y_n)_{n\in\mathbb{N}_0}$ unabhängig ist und $P_{Y_0} = \lambda\!\!\lambda_{0,1}$.

Nach Lemma 2.5 gibt es eine Zufallsgröße $Y_0$ auf einem Wahrscheinlichkeitsraum $(\Omega', \Sigma', P')$ mit $P'_{Y_0} = \lambda\!\!\lambda_{0,1}$. Wir betrachten nun den Wahrscheinlichkeitsraum $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{P})$ definiert gemäß $\boldsymbol{\Omega} := \Omega' \times \Omega$, $\boldsymbol{\Sigma} := \Sigma' \otimes \Sigma$ und $\boldsymbol{P} := P' \otimes P$. Mit den Abbildungen $\mathcal{Y}_0 \colon \Omega' \times \Omega \longrightarrow \mathbb{R}$ und $\mathcal{Y}_n \colon \Omega' \times \Omega \longrightarrow \mathbb{R}$, welche wir für $n \in \mathbb{N}$ gemäß

$$\mathcal{Y}_0(\omega', \omega) := Y_0(\omega') \qquad \text{und} \qquad \mathcal{Y}_n(\omega', \omega) := Y_n(\omega)$$

definieren. Dann sind die Abbildungen $\mathcal{Y}_0, \mathcal{Y}_1, \mathcal{Y}_2, \ldots$ messbar, denn $\{\mathcal{Y}_0 \le t\} = \{Y_0 \le t\} \times \Omega$ und $\{\mathcal{Y}_n \le s\} = \Omega' \times \{Y_n \le s\}$. Damit sind $\mathcal{Y}_0, \mathcal{Y}_1, \mathcal{Y}_2, \ldots$ Zufallsgrößen auf $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{P})$. Außerdem haben wir

$$\begin{aligned} \boldsymbol{P}_{\mathcal{Y}_0}((-\infty, t]) &= \boldsymbol{P}(\{\mathcal{Y}_0 \le t\}) = \boldsymbol{P}(\{Y_0 \le t\} \times \Omega) = P'(\{Y_0 \le t\}) \cdot P(\Omega) \\ &= P'(\{Y_0 \le t\}) = \lambda\!\!\lambda_{0,1}((-\infty, t]), \end{aligned}$$

also $\boldsymbol{P}_{\mathcal{Y}_0} = \lambda\!\!\lambda_{0,1}$ und analog $\boldsymbol{P}_{\mathcal{Y}_n} = P_{Y_n} = P_{Y_1}$. Schließlich gilt

$$\begin{aligned} &\boldsymbol{P}(\{\mathcal{Y}_0 \le t\} \cap \{\mathcal{Y}_n \le s\}) \\ =&\boldsymbol{P}((\{Y_0 \le t\} \times \Omega) \cap (\Omega' \times \{Y_n \le s\})) = \boldsymbol{P}(\{Y_0 \le t\} \times \{Y_n \le s\}) \\ =&P'(\{Y_0 \le t\}) P(\{Y_n \le s\}) = [P'(\{Y_0 \le t\}) \cdot P(\Omega)] [P'(\Omega') \cdot P(\{Y_n \le s\})] \\ =&\boldsymbol{P}(\{Y_0 \le t\} \times \Omega) \boldsymbol{P}(\Omega' \times \{Y_n \le s\}) \boldsymbol{P}(\{\mathcal{Y}_0 \le t\}) \boldsymbol{P}(\{\mathcal{Y}_n \le s\}), \end{aligned} \tag{3.6}$$

so dass $\mathcal{Y}_0$ und $\mathcal{Y}_n$ für jedes $n \in \mathbb{N}$ unabhängig (bezüglich $\boldsymbol{P}$) sind. Schließlich gilt für alle $m \in \mathbb{N}$ und jede Wahl von paarweise verschiedenen $k_1, \ldots, k_m \in \mathbb{N}$ und $s_1, \ldots, s_m \in \mathbb{R}$ dann noch wegen der Unabhängigkeit der Folge $(Y_n)_{n\in\mathbb{N}}$ (bezüglich $P$)

$$\begin{aligned} &\boldsymbol{P}(\{\mathcal{Y}_{k_1} \le s_1\} \cap \ldots \cap \{\mathcal{Y}_{k_m} \le s_m\}) \\ =&\boldsymbol{P}((\Omega' \times \{Y_{k_1} \le s_1\}) \cap \ldots \cap (\Omega' \times \{Y_{k_m} \le s_1\})) \\ =&\boldsymbol{P}(\Omega' \times (\{Y_{k_1} \le s_1\} \cap \ldots \cap \{Y_{k_m} \le s_m\})) \\ =&P'(\Omega') P(\{Y_{k_1} \le s_1\} \cap \ldots \cap \{Y_{k_m} \le s_m\}) \\ =&P'(\Omega') P(\{Y_{k_1} \le s_1\}) \cdot \ldots \cdot P(\{Y_{k_m} \le s_m\}) \\ =&[P'(\Omega') \cdot P(\{Y_{k_1} \le s_1\})] \cdot \ldots \cdot [P'(\Omega') \cdot P(\{Y_{k_m} \le s_m\})] \\ =&\boldsymbol{P}(\Omega' \times \{Y_{k_1} \le s_1\}) \cdot \ldots \cdot \boldsymbol{P}(\Omega' \times \{Y_{k_m} \le s_m\}) \\ =&\boldsymbol{P}(\{\mathcal{Y}_{k_1} \le s_1\}) \cdot \ldots \cdot \boldsymbol{P}(\{\mathcal{Y}_{k_m} \le s_m\}), \end{aligned} \tag{3.7}$$

so dass die Folge $(\mathcal{Y}_n)_{n\in\mathbb{N}}$ unabhängig (bezüglich $\boldsymbol{P}$) ist. Schließlich kann man mit (3.6) und (3.7) zeigen, dass auch die Folge $(\mathcal{Y}_n)_{n\in\mathbb{N}_0}$ unabhängig bezüglich $\boldsymbol{P}$ ist. Wir können also stets einen gemeinsamen Wahrscheinlichkeitsraum $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{P})$ und eine Folge unabhängiger

Zufallsgrößen $(\mathcal{Y}_n)_{n\in\mathbb{N}_0}$ auf $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{P})$ konstruieren mit $\boldsymbol{P}_{\mathcal{Y}_0}=\lambda\!\!\!\lambda_{0,1}$ und $\boldsymbol{P}_{\mathcal{Y}_n}=P_{Y_n}=P_{Y_1}$ für alle $n\in\mathbb{N}$.

Um die Aussage des Satzes 3.9 zu beweisen, wollen wir zeigen, dass

$$\boldsymbol{P}\left(\left\{\left(\sum_{k=1}^{n}\mathcal{Y}_k\right)_{n\in\mathbb{N}}\text{ ist gleichverteilt modulo 1}\right\}\right)=1$$

gilt. Angenommen wir hätten dies bereits gezeigt. Dann ist die Zahlenfolge

$$\left(\sum_{k=1}^{n}\mathcal{Y}_k\left(\omega',\omega\right)\right)_{n\in\mathbb{N}}=\left(\sum_{k=1}^{n}Y_k\left(\omega\right)\right)_{n\in\mathbb{N}}$$

für $\boldsymbol{P}$-fast-alle $(\omega',\omega)\in\Omega'\times\Omega$ gleichverteilt modulo 1. Dann muss die Aussage jedoch wegen $\boldsymbol{P}=P'\otimes P$ für $P$-fast-alle $\omega\in\Omega$ gelten, das heißt $\left(\sum_{k=1}^{n}Y_k\left(\omega\right)\right)_{n\in\mathbb{N}}$ ist für $P$-fast-alle $\omega\in\Omega$ gleichverteilt modulo 1, also $P\left(\left\{\left(\sum_{k=1}^{n}Y_k\right)_{n\in\mathbb{N}}\text{ ist gleichverteilt modulo 1}\right\}\right)=1$. Haben wir also die Aussage für die modifizierten Zufallsgrößen $\mathcal{Y}_1,\mathcal{Y}_2,\ldots$ bezüglich $\boldsymbol{P}$ gezeigt, so gilt sie auch für $Y_1,Y_2,\ldots$ bezüglich $P$. Zusammenfassend können wir damit ohne Beschränkung der Allgemeinheit $(\Omega,\Sigma,P)=(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{P})$ annehmen. Wir können also ohne Einschränkung davon ausgehen, dass es eine Zufallsgröße $Y_0$ auf $(\Omega,\Sigma,P)$ gibt, so dass $(Y_n)_{n\in\mathbb{N}_0}$ unabhängig ist und $P_{Y_0}=\lambda\!\!\!\lambda_{0,1}$. ♣

Wir kommen nun zum Beweis des Satzes 3.9:

Beweis. Teil *(a)* wurde bereits in Satz 2.6 bewiesen. Wir müssen also nur Teil *(b)* beweisen. Dazu betrachten wir für $n\in\mathbb{N}$

$$Y_n:=\log S\left(X_n\right)=\langle\log|X_n|\rangle\in[0,1)\,,$$

wobei die letzte Gleicheit schon im Beweis von Teil *(a)* von Satz 3.4 gezeigt wurde. Die Folge $(Y_n)_{n\in\mathbb{N}}$ ist weiterhin unabhängig, identisch verteilt und $Y_1$ nimmt nicht nur abzählbar viele Werte an. Nach letzter Bemerkung können wir nun ohne Beschränkung der Allgemeinheit eine Zufallsgröße $Y_0$ auf $(\Omega,\Sigma,P)$ wählen, so dass $(Y_n)_{n\in\mathbb{N}_0}$ unabhängig ist und $P_{Y_0}=\lambda\!\!\!\lambda_{0,1}$. Ferner sei

$$\begin{aligned}\mathrm{supp}P_{Y_1}&:=\{y\in[0,1)\,|\,\forall\varepsilon>0\colon P\left(\{|Y_1-y|<\varepsilon\}\right)>0\}\\&=\{y\in[0,1)\,|\,\forall\varepsilon>0\colon P_{Y_1}\left(U_\varepsilon\left(y\right)\right)>0\}\,,\end{aligned}$$

wobei $U_\varepsilon\left(y\right):=\{x\in[0,1)\,|\,|x-y|<\varepsilon\}$.

Wir zeigen nun zunächst, dass $[0,1)\setminus\mathrm{supp}\,P_{Y_1}$ offen ist. Sei dazu $y\in[0,1)\setminus\mathrm{supp}\,P_{Y_1}$. Dann gibt es ein $\varepsilon>0$ mit $P_{Y_1}\left(U_\varepsilon\left(y\right)\right)=0$. Wir müssen zeigen, dass es dann ein $\varepsilon'>0$ gibt, so dass $U_{\varepsilon'}\left(y\right)\subseteq[0,1)\setminus\mathrm{supp}\,P_{Y_1}$. Sei dazu $y'\in U_\varepsilon\left(y\right)$, das heißt es gilt $\delta:=|y-y'|<\varepsilon$. Offenbar gilt dann wegen der Dreiecksungleichung

$$U_{\varepsilon-\delta}\left(y'\right)\subseteq U_\varepsilon\left(y\right).$$

Damit ist $0\le P_{Y_1}\left(U_{\varepsilon-\delta}\left(y'\right)\right)\le P_{Y_1}\left(U_\varepsilon\left(y\right)\right)=0$ und somit $y'\in[0,1)\setminus\mathrm{supp}\,P_{Y_1}$, das heißt $U_\varepsilon\left(y\right)\subseteq[0,1)\setminus\mathrm{supp}\,P_{Y_1}$. Also ist $[0,1)\setminus\mathrm{supp}\,P_{Y_1}$ offen.

Wir zeigen nun

$$[0,1)\setminus\mathrm{supp}\,P_{Y_1}=\bigcup_{q\in\mathbb{Q}\cap\left([0,1)\setminus\mathrm{supp}\,P_{Y_1}\right)}U_{\varepsilon(q)}\left(q\right),$$

wobei für $q\in\mathbb{Q}\cap\left([0,1)\setminus\mathrm{supp}\,P_{Y_1}\right)$ dann $\varepsilon\left(q\right):=\sup\left\{\varepsilon>0\,|\,U_\varepsilon\left(q\right)\subseteq[0,1)\setminus\mathrm{supp}\,P_{Y_1}\right\}$. So ein $\varepsilon\left(q\right)$ gibt es immer, da $[0,1)\setminus\mathrm{supp}\,P_{Y_1}$ offen ist. Sei zunächst

$$x\in\bigcup_{q\in\mathbb{Q}\cap\left([0,1)\setminus\mathrm{supp}\,P_{Y_1}\right)}U_{\varepsilon(q)}\left(q\right).$$

Dann gibt es ein $q \in \mathbb{Q} \cap ([0,1) \setminus \operatorname{supp} P_{Y_1})$ mit $x \in U_{\varepsilon(q)}(q)$. Per Definition von $\varepsilon(q)$ gilt $U_{\varepsilon(q)} \subseteq [0,1) \setminus \operatorname{supp} P_{Y_1}$, also $x \in [0,1) \setminus \operatorname{supp} P_{Y_1}$. Ist nun umgekehrt $x \in [0,1) \setminus \operatorname{supp} P_{Y_1}$, so gibt es wegen der Offenheit von $[0,1) \setminus \operatorname{supp} P_{Y_1}$ ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subseteq [0,1) \setminus \operatorname{supp} P_{Y_1}$. Wegen der Dichtheit der rationalen Zahlen in $\mathbb{R}$, gibt es dann ein $q \in \mathbb{Q} \cap ([0,1) \setminus \operatorname{supp} P_{Y_1})$ mit

$$|x - q| < \frac{\varepsilon}{3}.$$

Für jedes $x' \in U_{\frac{\varepsilon}{2}}(q)$ gilt dann wegen $|x - x'| \leq |x - q| + |q - x'| < \frac{\varepsilon}{3} + \frac{\varepsilon}{2} < \varepsilon$ auch $x' \in U_\varepsilon(x)$, das heißt es gilt $x \in U_{\frac{\varepsilon}{2}}(q) \subseteq U_\varepsilon(x) \subseteq [0,1) \setminus \operatorname{supp} P_{Y_1}$. Per Definition von $\varepsilon(q)$ heißt das $\frac{\varepsilon}{2} \leq \varepsilon(q)$. Daher gilt wegen $U_{\frac{\varepsilon}{2}}(q) \subseteq U_{\varepsilon(q)}(q)$ auch $x \in U_{\varepsilon(q)}(q)$ und damit

$$x \in \bigcup_{q \in \mathbb{Q} \cap \left([0,1) \setminus \operatorname{supp} P_{Y_1}\right)} U_{\varepsilon(q)}(q).$$

Damit ist die Gleichheit gezeigt. Insbesondere lässt sich $[0,1) \setminus \operatorname{supp} P_{Y_1}$ also als höchstens abzählbare Vereinigung

$$[0,1) \setminus \operatorname{supp} P_{Y_1} = \bigcup_{n=1}^{\infty} U_{\varepsilon_n}(q_n),$$

schreiben, wobei für alle $n \in \mathbb{N}$ dann $U_{\varepsilon_n}(q_n) \subseteq [0,1) \setminus \operatorname{supp} P_{Y_1}$ und $q_n \in [0,1) \setminus \operatorname{supp} P_{Y_1}$. Es folgt nun

$$P_{Y_1}([0,1) \setminus \operatorname{supp} P_{Y_1}) = P_{Y_1}\left(\bigcup_{n=1}^{\infty} U_{\varepsilon_n}(q_n)\right) \leq \sum_{n=1}^{\infty} \underbrace{P_{Y_1}(U_{\varepsilon_n}(q_n))}_{=0} = 0,$$

also $P_{Y_1}(\operatorname{supp} P_{Y_1}) = 1$. Ferner enthält $\operatorname{supp} P_{Y_1}$ mindestens eine irrationale Zahl $\alpha$, denn würde $\operatorname{supp} P_{Y_1}$ nur rationale Zahlen enthalten, so wäre

$$1 = P_{Y_1}(\operatorname{supp} P_{Y_1}) \leq P_{Y_1}(\mathbb{Q} \cap [0,1)) = P(\{Y_1 \in \mathbb{Q} \cap [0,1)\}) = 0,$$

was ein Widerspruch ist.

Definiere nun

$$\mathbb{S}_{Y_1} := \{\langle y_1 + y_2 + \ldots + y_n \rangle \mid n \in \mathbb{N},\ y_k \in \operatorname{supp} P_{Y_1}\} \subseteq [0,1).$$

Da $\operatorname{supp} P_{Y_1}$ mindestens eine irrationale Zahl $\alpha$ enthält, ist für jedes $n \in \mathbb{N}$ die Zahl $\langle n\alpha \rangle \in \mathbb{S}_{Y_1}$. Nach Teil *(a)* von Lemma 2.2 ist die Folge $(n\alpha)_{n \in \mathbb{N}}$ gleichverteilt modulo 1 und daher ist die Menge $\{\langle n\alpha \rangle \mid n \in \mathbb{N}\}$ dicht in $[0,1)$. Damit ist auch $[0,1) \supseteq \mathbb{S}_{Y_1} \supseteq \{\langle n\alpha \rangle \mid n \in \mathbb{N}\}$ dicht in $[0,1)$.

Der weitere Beweis beruht auf der Anwendung des Birkhoffschen Ergodensatzes. Wir definieren uns ein maßtheoretisches dynamisches System durch

$$X := [0,1)^{\mathbb{N}_0}$$

und

$$\Sigma := \bigotimes_{n=0}^{\infty} \mathfrak{B}[0,1).$$

Wir betrachten dann das Produktmaß

$$\mu := \lambda\!\!\lambda_{0,1} \otimes P_{Y_1} \otimes P_{Y_1} \otimes P_{Y_1} \otimes \ldots.$$

Das heißt für jede Rechteckmenge $B_0 \times B_1 \times \ldots \times B_m \times [0,1) \times \ldots$ mit $B_0, B_1, \ldots, B_m \in \mathfrak{B}[0,1)$ ist

$$\begin{aligned} \mu(B_0 \times B_1 \times \ldots \times B_m \times [0,1) \times \ldots) &= \lambda\!\!\lambda_{0,1}(B_0)\, P_{Y_1}(B_1) \cdot \ldots \cdot P_{Y_1}(B_m) \\ &= \lambda\!\!\lambda_{0,1}(B_0)\, P_{Y_1}(B_1) \cdot \ldots \cdot P_{Y_m}(B_m) \\ &= P(\{Y_0 \in B_0, Y_1 \in B_1, \ldots, Y_m \in B_m\}), \end{aligned}$$

wobei die Unabhängigkeit und die identische Verteiltheit ausgenutzt wurde. Wir definieren nun eine Transformation $T\colon X \longrightarrow X$ gemäß

$$T(s_0, s_1, s_2, \ldots) := (\langle s_0 + s_1 \rangle, s_2, \ldots).$$

Da $Y_0$ gleichverteilt modulo 1 und $Y_1$ unabhängig von $Y_0$ ist liefert Teil *(a)* von Satz 2.5, dass auch $\langle Y_0 + Y_1 \rangle$ gleichverteilt modulo 1 ist, das heißt $P_{\langle Y_0+Y_1\rangle} = \lambda\!\!\lambda_{0,1}$. Damit folgt, dass $T$ $\mu$-maßerhaltend ist, denn wegen

$$\begin{aligned}
&\mu\left(T^{-1}\left(B_0 \times B_1 \times \ldots \times B_m \times [0,1) \times \ldots\right)\right)\\
=&P\left(\{\langle Y_0 + Y_1 \rangle \in B_0, Y_2 \in B_1, \ldots, Y_{m-1} \in B_m\}\right)\\
=&P_{\langle Y_0+Y_1\rangle}(B_0) \cdot P_{Y_2}(B_1) \cdot \ldots \cdot P_{Y_{m-1}}(B_m)\\
=&\lambda\!\!\lambda_{0,1}(B_0) \cdot P_{Y_1}(B_1) \cdot \ldots \cdot P_{Y_1}(B_m)\\
=&\mu\left(B_0 \times B_1 \times \ldots \times B_m \times [0,1) \times \ldots\right),
\end{aligned}$$

ist $T$ auf einem Erzeuger von $\Sigma$ $\mu$-maßerhaltend und damit auch auf ganz $\Sigma$. Wir wollen nun zeigen, dass $(X, \Sigma, T, \mu)$ sogar ergodisch ist. Dabei wenden wir insbesondere Satz 3.7 und 3.8 an. Wir zeigen die Ergodizität mittels Satz 3.7 durch das Zeigen der Aussage

*(I)* Fix $\mathtt{T}$ besteht nur aus $\mu$-fast-überall konstanten Funktionen.

Sei also $g \in \operatorname{Fix} \mathtt{T}$, das heißt $(\mathtt{T}g)(x) = g(x)$ für $\mu$-fast-alle $x \in X$ bzw.

$$g\left(T(s_0, s_1, s_2, \ldots)\right) = g(s_0, s_1, s_2, \ldots), \qquad \text{für } \mu\text{-fast-alle } (s_0, s_1, s_2, \ldots) \in X.$$

Mit den Bezeichnungen aus Satz 3.8 sei nun $(Y, \mathfrak{Y}, \eta) := ([0,1), \mathfrak{B}[0,1), \lambda\!\!\lambda_{0,1})$ und $(S, \mathfrak{S}, \nu) := ([0,1), \mathfrak{B}[0,1), P_{Y_1})$ zusammen mit der Transformation $\varphi_s(y) := \langle y + s \rangle$. Zunächst gilt wegen der Translationsinvarianz des Lebesgue-Maßes dann

$$\lambda\!\!\lambda_{0,1}\left(\varphi_s^{-1}(B)\right) = \lambda\!\!\lambda_{0,1}(\langle B - s \rangle) = \lambda\!\!\lambda_{0,1}(B), \qquad \text{für alle } B \in \mathfrak{B},$$

so dass $\varphi_s$ für jedes $s \in [0,1)$ eine $\lambda\!\!\lambda_{0,1}$-maßerhaltende Transformation ist. Nach Satz 3.8 gibt es dann ein $f\colon [0,1) \longrightarrow \mathbb{R}$, so dass $\mu$-fast-überall

$$g(s_0, s_1, s_2, \ldots) = f(s_0). \tag{3.8}$$

Zudem gilt für $\lambda\!\!\lambda_{0,1}$-fast-alle $s_0 \in [0,1)$ und $P_{Y_1}$-fast-alle $s \in [0,1)$

$$f(s_0) = f(\langle s_0 + s \rangle).$$

Das heißt es gibt ein $B \in \mathfrak{B}[0,1)$ mit $P_{Y_1}(B) = P(\{Y_1 \in B\}) = 1$, so dass $f(s_0) = f(\langle s_0 + s \rangle)$ für alle $s \in B$ und $\lambda\!\!\lambda_{0,1}$-fast-alle $s_0 \in [0,1)$ gilt. Wegen

$$P_{Y_1}(\operatorname{supp} P_{Y_1} \cap B) = P_{Y_1}(\operatorname{supp} P_{Y_1}) + P_{Y_1}(B) - P_{Y_1}(\operatorname{supp} P_{Y_1} \cup B) = 1 + 1 - 1 = 1,$$

können wir ohne Beschränkung der Allgemeinheit $B = \operatorname{supp} P_{Y_1}$ annehmen. Daher gilt die Aussage $f(s_0) = f(\langle s_0 + s \rangle)$ insbesondere für $\lambda\!\!\lambda_{0,1}$-fast-alle $s_0 \in [0,1)$ und *alle* $s \in \operatorname{supp} P_{Y_1}$. Es gilt also auch für $s_1, s_2 \in \operatorname{supp} P_{Y_1}$

$$f(s_0) = f(\langle s_0 + s_1 \rangle) = f(\langle \langle s_0 + s_1 \rangle + s_2 \rangle) = f(\langle s_0 + s_1 + s_2 \rangle),$$

und induktiv daher für alle $n \in \mathbb{N}$ und alle $s_1, \ldots, s_n \in \operatorname{supp} P_{Y_1}$

$$f(s_0) = f(\langle s_0 + s_1 + \ldots + s_n \rangle),$$

jeweils für $\lambda\!\!\lambda_{0,1}$-fast-alle $s_0 \in [0,1)$.

Betrachte nun für $\alpha \in \mathbb{R}$ die Menge $A := \{s_0 \in [0,1) \,|\, f(s_0) \le \alpha\} = \{f \le \alpha\}$. Mit $s \in \operatorname{supp} P_{Y_1}$ gilt

$$\begin{aligned}
\varphi_s^{-1}(A) &= \{s_0 \in [0,1) \,|\, \langle s_0 + s \rangle \in A\} = \{s_0 \in [0,1) \,|\, f(\langle s_0 + s \rangle) \le \alpha\}\\
&\doteq \{s_0 \in [0,1) \,|\, f(s_0) \le \alpha\} = A,
\end{aligned}$$

wobei wie im Beweis von Satz 3.7 „$\doteq$" die Gleichheit bis auf eine $\lambda\!\!\lambda_{0,1}$ Nullmenge bedeuten soll. Damit ist $A$ eine $\varphi_s$-invariante Menge und somit

$$\lambda\!\!\lambda_{0,1}\left(A \triangle \varphi_s^{-1}(A)\right) = 0.$$

Wir wollen nun zeigen, dass dies nur dann möglich ist, wenn $\lambda\!\!\lambda_{0,1}(A) = 0$ oder $\lambda\!\!\lambda_{0,1}(A) = 1$ gilt. Angenommen es gilt $\lambda\!\!\lambda_{0,1}(A) > 0$. Dann definiere durch

$$\nu(B) := \frac{\lambda\!\!\lambda_{0,1}(B \cap A)}{\lambda\!\!\lambda_{0,1}(A)} \tag{3.9}$$

ein Wahrscheinlichkeitsmaß auf $([0,1), \mathfrak{B}[0,1))$. Dann ist für $B \in \mathfrak{B}[0,1)$

$$\begin{aligned}
\lambda\!\!\lambda_{0,1}(A) \cdot |\nu(B) - \nu(\langle B+s\rangle)| &= |\lambda\!\!\lambda_{0,1}(A \cap B) - \lambda\!\!\lambda_{0,1}(A \cap \langle B+s\rangle)| \\
&= |\lambda\!\!\lambda_{0,1}(A \cap B) - \lambda\!\!\lambda_{0,1}(\langle A-s\rangle \cap B)| \\
&= \left| \int_0^1 \left(\mathbf{1}_{A\cap B} - \mathbf{1}_{\langle A-s\rangle \cap B}\right) \mathrm{d}\lambda\!\!\lambda_{0,1} \right| \\
&= \left| \int_0^1 \mathbf{1}_B \left(\mathbf{1}_A - \mathbf{1}_{\langle A-s\rangle}\right) \mathrm{d}\lambda\!\!\lambda_{0,1} \right| \\
&\le \int_0^1 \left|\mathbf{1}_A - \mathbf{1}_{\langle A-s\rangle}\right| \mathrm{d}\lambda\!\!\lambda_{0,1} = \int_0^1 \mathbf{1}_{A\triangle\langle A-s\rangle} \mathrm{d}\lambda\!\!\lambda_{0,1} \\
&= \lambda\!\!\lambda_{0,1}(A \triangle \langle A-s\rangle) = \lambda\!\!\lambda_{0,1}\left(A \triangle \varphi_s^{-1}(A)\right) \\
&= 0.
\end{aligned}$$

Also gilt $\nu(B) = \nu(\langle B+s\rangle)$ für alle $B \in \mathfrak{B}[0,1)$ und alle $s \in \operatorname{supp} P_{Y_1}$. Für die Fourier-Koeffizienten von $\nu$ folgt daraus

$$\begin{aligned}
\widehat{\nu}(k) &= \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\, \nu(\mathrm{d}t) = \int_0^1 \mathrm{e}^{\mathrm{i}2\pi k\langle t+s\rangle}\, \nu(\mathrm{d}t) \\
&= \int_0^1 \mathrm{e}^{\mathrm{i}2\pi k(t+s)}\, \mathrm{e}^{-\mathrm{i}2\pi k\lfloor t+s\rfloor}\, \nu(\mathrm{d}t) = \int_0^1 \mathrm{e}^{\mathrm{i}2\pi k(t+s)}\, \nu(\mathrm{d}t) \\
&= \mathrm{e}^{\mathrm{i}2\pi ks} \cdot \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\, \nu(\mathrm{d}t) = \mathrm{e}^{\mathrm{i}2\pi ks}\, \widehat{\nu}(k).
\end{aligned}$$

Da $\mathbb{S}_{Y_1}$ dicht in $[0,1)$ liegt, gibt es nun für jedes $k \neq 0$ eine Folge $(s_{k;n})_{n\in\mathbb{N}}$ aus $\mathbb{S}_{Y_1}$ mit $\lim_{n\to\infty} k \cdot s_{k;n} = \frac{1}{2}$. Damit folgt für die Fourier-Koeffizienten

$$\widehat{\nu}(k) = \mathrm{e}^{\mathrm{i}2\pi k s_{k;n}}\, \widehat{\nu}(k) \xrightarrow{n\to+\infty} \mathrm{e}^{\mathrm{i}\pi}\, \widehat{\nu}(k) = -\widehat{\nu}(k),$$

also $\widehat{\nu}(k) = 0$ für alle $k \neq 0$. Folglich hat $\nu$ dieselben Fourier-Koeffizienten wie $\lambda\!\!\lambda_{0,1}$ und es gilt

$$\nu = \lambda\!\!\lambda_{0,1}.$$

Damit folgt aus (3.9)

$$\lambda\!\!\lambda_{0,1}(A) = \frac{\lambda\!\!\lambda_{0,1}(A \cap A)}{\lambda\!\!\lambda_{0,1}(A)} = 1.$$

Wir haben damit $\lambda\!\!\lambda_{0,1}(A) = 0$ oder $\lambda\!\!\lambda_{0,1}(A) = 1$ gezeigt.

Wegen $A = \{f \leq \alpha\}$ folgt wie im Beweis von Satz 3.7, dass dann $f$ $\lambda\!\!\lambda_{0,1}$-fast-überall konstant sein muss. Also ist $g \in \operatorname{Fix} \mathtt{T}$ wegen (3.8) auch $\mu$-fast-überall konstant. Daher ist Aussage *(I)* gezeigt, woraus sich wegen Satz 3.7 die Ergodizität von $(X, \Sigma, T, \mu)$ ergibt.

Wir wenden nun den komplexen Birkhoffschen Ergodensatz (vergleiche Folgerung 3.3) für $g\colon X \longrightarrow \mathbb{C}$ aus $\mathscr{L}^1(X, \Sigma, \mu)$ definiert gemäß

$$g(s_0, s_1, s_2, \ldots) := f(s_0)$$

mit einem $f\colon [0,1) \longrightarrow \mathbb{C}$ aus $\mathscr{L}^1([0,1), \mathfrak{B}[0,1), \lambda\!\!\lambda_{0,1})$ an. Dann folgt wegen der Ergodizität von $(X, \Sigma, T, \mu)$

$$\frac{1}{N}\sum_{n=0}^{N-1} g \circ T^n \xrightarrow{N \to +\infty} \int_X g \mathrm{d}\mu, \qquad \mu\text{-fast-überall.}$$

Nun ist aber

$$\begin{aligned}(g \circ T^n)(s_0, s_1, s_2, \ldots) &= g(\langle s_0 + s_1 + \ldots + s_n\rangle, s_{n+1}, s_{n+2}, \ldots)\\ &= f(\langle s_0 + s_1 + \ldots + s_n\rangle).\end{aligned}$$

Ferner gilt nach dem Satz von Fubini wegen $\mu = \lambda\!\!\lambda_{0,1} \otimes P_{Y_1} \otimes P_{Y_2} \otimes \ldots$

$$\int_X g\mathrm{d}\mu = \int_0^1\int_0^1 \cdots f(s_0)\, \lambda\!\!\lambda_{0,1}(\mathrm{d}s_0)\, P_{Y_1}(\mathrm{d}s_1)\, P_{Y_1}(\mathrm{d}s_2) \ldots = \int_0^1 f(s_0)\, \lambda\!\!\lambda_{0,1}(\mathrm{d}s_0).$$

Damit gilt für $\mu$-fast-alle $(s_0, s_1, s_2, \ldots) \in X$

$$\frac{1}{N}\sum_{n=0}^{N-1} f(\langle s_0 + s_1 + s_2 + \ldots + s_n\rangle) \xrightarrow{N \to +\infty} \int_0^1 f \mathrm{d}\lambda\!\!\lambda_{0,1}.$$

Ferner haben wir

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N} f(\langle s_0 + s_1 + \ldots + s_n\rangle)\\
&= \lim_{N-1\to\infty}\frac{1}{N-1}\sum_{n=0}^{N-1} f(\langle s_0 + s_1 + \ldots + s_n\rangle)\\
&= \lim_{N-1\to\infty}\frac{N}{N-1}\cdot\frac{1}{N}\sum_{n=0}^{N-1} f(\langle s_0 + s_1 + \ldots + s_n\rangle)\\
&= \lim_{N-1\to\infty}\frac{N}{N-1}\cdot\lim_{N-1\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(\langle s_0 + s_1 + \ldots + s_n\rangle)\\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(\langle s_0 + s_1 + \ldots + s_n\rangle).\end{aligned}$$

Daher gilt auch

$$\frac{1}{N}\sum_{n=0}^{N} f(\langle s_0 + s_1 + s_2 + \ldots + s_n\rangle) \xrightarrow{N \to +\infty} \int_0^1 f \mathrm{d}\lambda\!\!\lambda_{0,1}, \qquad \mu\text{-fast-überall.}$$

Wegen $\mu = \lambda\!\!\lambda_{0,1} \otimes P_{Y_1} \otimes P_{Y_1} \otimes \ldots = P_{Y_0} \otimes P_{Y_1} \otimes P_{Y_2} \otimes \ldots$ gilt für $P$-fast-alle $\omega \in \Omega$

$$\frac{1}{N}\sum_{n=0}^{N} f(\langle Y_0(\omega) + Y_1(\omega) + \ldots + Y_n(\omega)\rangle) \xrightarrow{N \to +\infty} \int_0^1 f \mathrm{d}\lambda\!\!\lambda_{0,1}. \tag{3.10}$$

Wir betrachten nun speziell für $k \in \mathbb{Z}\setminus\{0\}$ die Funktion $f(t) := \mathrm{e}^{\mathrm{i}2\pi kt}$. Dann ist

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N} f\left(\langle Y_0(\omega)+Y_1(\omega)+\ldots+Y_n(\omega)\rangle\right) &= \frac{1}{N}\sum_{n=0}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_0(\omega)+Y_1(\omega)+\ldots+Y_n(\omega))}\\ &= \frac{1}{N}\sum_{n=0}^{N} \mathrm{e}^{\mathrm{i}2\pi kY_0(\omega)}\,\mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))}\\ &= \mathrm{e}^{\mathrm{i}2\pi kY_0(\omega)}\,\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))}.\end{aligned}$$

Nach (3.10) gilt für $P$-fast-alle $\omega \in \Omega$ dann

$$\mathrm{e}^{\mathrm{i}2\pi kY_0(\omega)}\,\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))} \xrightarrow{N\to+\infty} \int_0^1 \mathrm{e}^{\mathrm{i}2\pi kt}\,\lambda\!\!\lambda_{0,1}(\mathrm{d}t) = 0.$$

Da aber $\left|\mathrm{e}^{\mathrm{i}2\pi kY_0(\omega)}\right| = 1$ muss dann auch

$$\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))} \xrightarrow{N\to+\infty} 0, \qquad \text{für } P\text{-fast-alle } \omega\in\Omega$$

gelten. Das heißt es gibt ein $\Omega_k \in \Sigma$ mit $P(\Omega_k) = 1$ mit

$$\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))} \xrightarrow{N\to+\infty} 0, \qquad \text{für alle } \omega\in\Omega_k. \tag{3.11}$$

Da der abzählbare Durchschnitt von Mengen mit Maß 1 immer noch Maß 1 hat, gilt $P\left(\bigcap_{k\in\mathbb{Z}\setminus\{0\}}\Omega_k\right) = 1$ und (3.11) gilt für alle $\omega \in \bigcap_{k\in\mathbb{Z}\setminus\{0\}}\Omega_k$. Damit gilt

$$\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k(Y_1(\omega)+\ldots+Y_n(\omega))} \xrightarrow{N\to+\infty} 0, \qquad \text{für } P\text{-fast-alle } \omega\in\Omega$$

und alle $k \in \mathbb{Z}\setminus\{0\}$. Mit dem Kriterium von Weyl (vergleiche Satz 2.3) bedeutet das aber, dass die Folge $\left(\sum_{k=1}^{n} Y_k(\omega)\right)_{n\in\mathbb{N}}$ für $P$-fast-alle $\omega\in\Omega$ gleichverteilt modulo 1 ist. Wegen $Y_k(\omega) = \log S(X_k)$ ist das nach Teil *(a)* von Satz 2.2 äquivalent dazu, dass die Folge $\left(\prod_{k=1}^{n} X_k(\omega)\right)_{n\in\mathbb{N}}$ für $P$-fast-alle $\omega\in\Omega$ Benford ist, also

$$P\left(\left\{\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1.$$

□

**Folgerung 3.4.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängig und identisch verteilter Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega,\Sigma,P)$, so dass $F_{X_1}$ stetig ist. Dann gilt:*

(a) *Die Folge $\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen das Benfordsche Gesetz.*

(b) $P\left(\left\{\left(\prod_{k=1}^{n} X_k\right)_{n\in\mathbb{N}} \text{ ist Benford}\right\}\right) = 1.$

Beweis. Falls $F_{X_1}$ stetig ist gilt für alle $t\in\mathbb{R}$ $P(\{X_1 = t\}) = 0$, so dass also $P(\{X_1\in C\}) = P\left(\bigcup_{t\in C}\{X_1 = t\}\right) = 0$ für jede abzählbare Menge $C \in \mathfrak{B}$. Damit folgt die Behauptung aus Satz 3.9. □

# Zusammenfassung

Ziel der vorliegenden Arbeit war es das von Newcomb und Benford aufgestellte Benfordsche Gesetz mathematisch zu präzisieren und Kriterien für die Gültigkeit zu erhalten. Dazu wurden zunächst die wichtigsten Definitionen in Form der Signifikanten und der signifikanten $\sigma$-Algebra getätigt. Es folgte die für die weiteren Untersuchungen wesentliche Definition der Benford-Eigenschaft für Zahlenfolgen, Funktionen, Zufallsgrößen und Wahrscheinlichkeitsmaße. Dabei wurde zudem die Benford-Verteilung $\mathbb{B}$ auf $(\mathbb{R}^+, \mathcal{S})$ eingeführt.

Mithilfe dieser Begriffsbildungen folgte als erstes Kriterium Satz 2.2, welcher die Benford-Eigenschaft für die verschiedenen Objekte mittels der Gleichverteilung modulo 1 charakterisiert. Als eine erste Anwendung wurde in Beispiel 2.4 dann gezeigt, dass die Folge der Fibonacci-Zahlen die Benford-Eigenschaft besitzt, jedoch die Folge der Primzahlen nicht. Im Hinblick auf das Grenzverhalten von Folgen von Zufallsgrößen wurde die Konvergenz in Verteilung gegen das Benfordsche Gesetz eingeführt. Satz 2.6 zeigt dann, dass falls eine Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ unabhängig und identisch verteilt ist, so dass die Zufallsgröße $X_1$ nicht nur abzählbar viele Werte annimmt, das Produkt $\prod_{k=1}^{n} X_k$ in Verteilung gegen das Benfordsche Gesetz konvergiert. Als zweite große Charakterisierung der Benford-Eigenschaft folgte dann Satz 2.7. Dieser besagt, dass die Benford-Eigenschaft von Wahrscheinlichkeitsmaßen äquivalent zu der Eigenschaft von skaleninvarianten signifikanten Dezimalziffern ist. Ferner ließ sich die Benford-Verteilung mittels des sogenannten Multiplikationsspiels in Satz 2.8 charakterisieren. Schließlich wurde in Satz 2.9 gezeigt, dass ein Wahrscheinlichkeitsmaß genau dann Basis-invariante signifikante Ziffern besitzt, wenn es sich als eine Konvexkombination aus dem Dirac-Maß $\delta_1$ und der Benford-Verteilung $\mathbb{B}$ schreiben lässt.

Das letzte Kapitel widmete sich dann der asymptotischen Betrachtung von Produkten von Zufallsgrößen. Zunächst wurde in Satz 3.2 gezeigt, dass falls die Zufallsgröße $X$ eine Riemann-integrierbare Dichte besitzt, die Folge $(X^n)_{n\in\mathbb{N}}$ in Verteilung gegen das Benfordsche Gesetz konvergiert und die Folge $(X^n)_{n\in\mathbb{N}}$ mit Wahrscheinlichkeit 1 Benford ist. Im anschließenden Beispiel wurde zudem die Konvergenzgeschwindigkeit gegen das Benfordsche Gesetz für eine auf $[0, 1)$ gleichverteilte Zufallsgröße bestimmt. Abschließend wurde dann das Produkt $\prod_{k=1}^{n} X_k$ einer unabhängig und identisch verteilten Folge von Zufallsgrößen untersucht. In Satz 3.9 wurde mithilfe des Birkhoffschen Ergodensatzes gezeigt, dass die Folge $(\prod_{k=1}^{n} X_k)_{n\in\mathbb{N}}$ mit Wahrscheinlichkeit 1 Benford ist, falls $X_1$ nicht nur abzählbar viele Werte annimmt.

# Literaturverzeichnis


[1] Benford, Frank: The Law of Anomalous Numbers. In: *Proceedings of the American Philosophical Society* (1938)

[2] Berg, Christian: *Fourier-Analysis*. `http://www.math.ku.dk/kurser/2009-10/blok1/fouan`. Version: 2009. – Accessed: 28.12.2015

[3] Berger, Arno ; Hill, Theodore: *An Introduction to Benford's Law*. Princeton University Press, 2015. – ISBN 9780691163062

[4] Berger, Arno ; Hill, Theodore P.: A basic theory of Benford's Law. In: *Probability Surveys* 8 (2011)

[5] Berger, Arno ; Hill, Theodore P. ; Kaynar, Bahar ; Ridder, Ad: Finite-state Markov Chains Obey Benford's Law. In: *SIAM J. Matrix Analysis & Applications* 32 (2011), S. 665–684

[6] Capraro, Valerio ; Morrison, Kent E.: *Existence of optimal strategies for the Operation Game on amenable semigroups*. `http://arxiv.org/abs/1104.3098v2`. Version: 2014. – Accessed: 28.12.2015

[7] Chow, Yuan S. ; Teicher, Henry: *Probability Theory. Independence, interchangeability, martingales*. Third edition. Springer, 1997. – ISBN 9780387406077

[8] Diaconis, Persi: The Distribution of Leading Digits and Uniform Distribution mod 1. In: *The Annals of Probability* 5 (1977), S. 72–81

[9] Eisner, Tatjana ; Farkas, Bálint ; Haase, Markus ; Nagel, Rainer: *Operator Thereotic Aspects of Ergodic Theory*. Springer, 2015

[10] Hill, Theodore: Base-invariance implies Benfords law. In: *Proceedings of the American Mathematical Society* 123 (1995), S. 887–895

[11] Hill, Theodore P.: A statistical derivation of the significant-digit law. In: *Statistical Science* 10 (1995)

[12] Hungerbühler, Norbert: *Benfords Gesetz über führende Ziffern: Wie die Mathematik Steuersündern das Fürchten lehrt*. `http://www.educ.ethz.ch/unt/um/mathe/ana/benford`. Version: 2007. – Accessed: 22.12.2015

[13] Klenke, Achim: *Wahrscheinlichkeitstheorie*. 3. Auflage. Springer, 2013. – ISBN 9783642360176

[14] Kupiers, Lauwerens ; Niederreiter, Harald: *Uniform distribution of sequences*. John Wiley & Sons, 1974

[15] Morrison, Kent E.: The Multiplication Game. In: *Mathematics Magazine* 83 (2012), S. 100–110

[16] Newcomb, Simon: Note on the Frequency of the Use of different Digits in Natural Numbers. In: *American Journal of Mathematics* (1881)

[17] Petersen, Karl: *Ergodic Theory*. Cambridge University Press, 1989

[18] Ryll-Nardzewski, C.: On the ergodic theorems III. In: *'Studia Mathematica* 14 (1954), S. 299–301

# Eigenständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.

Leipzig, den 19. Januar 2016