2 Ergodentheorie

In Kap. 1 haben wir stetige Transformationen auf kompakten metrischen Räumen untersucht, jedoch bereits in mehreren Beispielen invariante Wahrscheinlichkeitsmaße auf der Borel-σ-Algebra des kompakten Raumes gefunden. Damit lassen sich derartige Transformationen (und v. a. ihr asymptotisches Verhalten) auch mit maßtheoretischen Methoden beschreiben und untersuchen. In diesem Kapitel lassen wir die Topologie beiseite und betrachten Transformationen auf abstrakten Maßräumen. Die Untersuchung derartiger Transformationen ist Gegenstand der *Ergodentheorie*. Auf den historischen Ursprung des Namens „Ergodentheorie" werden wir im Anschluss an die Definition 2.12 der Ergodizität eingehen.

2.1 Ergodensätze

Wir geben zunächst noch einmal die Definition eines invariantes Maßes, jedoch in einer etwas allgemeineren Form.

Es seien X eine Menge, $\mathcal{P}(X)$ die Menge aller Teilmengen von X und $\mathcal{S} \subset \mathcal{P}(X)$ eine σ-Algebra. Dann nennt man $(X, \mathcal{S})$ einen *messbaren Raum*. Wenn μ ein Wahrscheinlichkeitsmaß auf $\mathcal{S}$ ist, so stellt $(X, \mathcal{S}, \mu)$ einen *Wahrscheinlichkeitsraum* dar. Wenn $(Y, \mathcal{T})$ ein zweiter messbarer Raum ist, so heißt eine Abbildung $\phi: X \longrightarrow Y$ *messbar* (oder $(\mathcal{S}, \mathcal{T})$-*messbar*), wenn $\phi^{-1}(\mathcal{T}) \subset \mathcal{S}$ gilt. Wenn es eine Bijektion $\phi: X \longrightarrow Y$ mit $\phi(\mathcal{S}) = \mathcal{T}$ gibt, so nennt man die Räume $(X, \mathcal{S})$ und $(Y, \mathcal{T})$ *isomorph* und ϕ ist ein *messbarer Isomorphismus* von $(X, \mathcal{S})$ und $(Y, \mathcal{T})$.

Definition 2.1 (Bildmaß und maßtreue Transformation)

Es seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum, $(Y, \mathcal{T})$ ein messbarer Raum und $T: X \longrightarrow Y$ eine messbare Transformation. Das durch $\nu(B) = \mu(T^{-1}(B))$, $B \in \mathcal{T}$, gegebene Wahrscheinlichkeitsmaß ν auf der σ-Algebra $\mathcal{T}$ heißt *Bildmaß* von μ unter T und wird mit $T_*\mu$ oder μT^{-1} bezeichnet.

M. Einsiedler, K. Schmidt, *Dynamische Systeme*, Mathematik Kompakt,
DOI 10.1007/978-3-0348-0634-3_2, © Springer Basel 2014

Wenn $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T: X \longrightarrow X$ eine messbare (d. h. $(\mathcal{S}, \mathcal{S})$-messbare) Transformation ist, für die $T_* \mu = \mu$ gilt, so nennt man T *maßtreu* oder *maßerhaltend* und das Maß μ T-*invariant*.

Eine maßtreue Transformation $T: X \longrightarrow X$ ist *invertierbar*, wenn T^{-1} existiert und messbar ist. Wenn T invertierbar ist, so ist natürlich auch T^{-1} maßtreu.

Die Invarianz von Maßen lässt sich auch mit Hilfe von Integralen ausdrücken.

Für $p \geq 1$ wird die Menge aller messbaren reellwertigen Funktionen f auf X, die die Bedingung $\int |f|^p \, d\mu < \infty$ erfüllen, mit $\mathcal{L}_p(\mu)$ oder $\mathcal{L}_p(X, \mathcal{S}, \mu)$ bezeichnet. Der Raum der messbaren beschränkten Funktionen $f: X \longrightarrow \mathbb{R}$ wird mit $\mathcal{L}_\infty(\mu)$ oder $\mathcal{L}_\infty(X, \mathcal{S}, \mu)$ bezeichnet. Die Menge der Äquivalenzklassen der Funktionen in $\mathcal{L}_p(\mu)$ wird mit $L_p(\mu) = \{[f] : f \in \mathcal{L}^p(\mu)\}$ bezeichnet.[1]

Für $[f] \in L_p(\mu)$ bezeichnet man mit $\|f\|_p = \left(\int |f|^p \, d\mu\right)^{1/p}$ die L_p-*Norm* der Äquivalenzklasse $[f]$ (die natürlich nicht von der Wahl des Repräsentanten $f \in [f]$ abhängt). Bezüglich dieser Norm ist $L_p(\mu)$ ein Banachraum [3, Kapitel VI].

Zumeist betrachten wir $\mathcal{L}_p$-Räume reellwertiger Funktionen. Falls wir doch komplexwertige Funktionen zulassen müssen, verwenden wir die Bezeichnungen $\mathcal{L}_p(\mu, \mathbb{C})$ oder $\mathcal{L}_p(X, \mathcal{S}, \mu, \mathbb{C})$ (bzw. $L_p(\mu, \mathbb{C})$ oder $L_p(X, \mathcal{S}, \mu, \mathbb{C})$ für die entsprechenden Räume von Äquivalenzklassen).

Obwohl man im Prinzip sorgfältig zwischen den Funktionenräumen $\mathcal{L}_p(\mu)$ und den Räumen $L_p(\mu)$ der Äquivalenzklassen unterscheiden muss, ist es sowohl in der Maßtheorie als auch in der Ergodentheorie üblich, diesen Unterschied *nicht* in der Notation zum Ausdruck zu bringen. So spricht man von Elementen von $L_p(\mu)$ als „Funktionen“ und schreibt $f \in L_p(\mu)$, auch wenn man eigentlich die Äquivalenzklasse $[f] \in L_p(\mu)$ meint. Wenn es wirklich nötig ist, *Funktionen* und nicht *Äquivalenzklassen* zu betrachten, dann nennt man manchmal eine Funktion $f \in \mathcal{L}_p(\mu)$ eine *Version* der Äquivalenzklasse $[f] \in L_p(\mu)$.

Satz 2.2 *Seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum, $T: X \longrightarrow X$ eine messbare Transformation und $\nu = T_* \mu$ das Bildmaß von μ unter T. Ist $f: X \longrightarrow \mathbb{R}$ messbar, dann liegt $f \circ T$ in $\mathcal{L}_1(\mu)$ genau dann, wenn f in $\mathcal{L}_1(\nu)$ liegt, und es gilt $\int f \circ T \, d\mu = \int f \, d\nu$.*

Beweis Es gilt $\int g \circ T \, d\mu = \int g \, d\nu$ für jede messbare nichtnegative Funktion $g: X \longrightarrow \mathbb{R}$ aufgrund der Transformationsformel in [3, Kapitel IV]. (Für Elementarfunktionen ist dies direkt aus den Definitionen ersichtlich. Für messbare nichtnegative Funktionen ergibt sich diese Formel aus dem Satz über monotone Konvergenz.)

Wenn nun $f: X \longrightarrow \mathbb{R}$ eine beliebige messbare Funktion ist, dann setzen wir $f^+ = \max(f, 0)$ und $f^- = \max(-f, 0)$ und erhalten so zwei nichtnegative messbare Funktionen

[1] Wir erinnern daran, dass zwei messbare Funktionen f, g auf X als *äquivalent* (genauer: als *äquivalent* $(\mathrm{mod}\, \mu)$) bezeichnet werden (in Symbolen: $f = g \pmod{\mu}$ oder $f = g$ μ-*f.ü.*), wenn $\mu(\{x : f(x) \neq g(x)\}) = 0$ ist.

mit $f = f^+ - f^-$. Wir haben gesehen, dass $\int f^+ \circ T\,d\mu = \int f^+\,d\nu$ und $\int f^- \circ T\,d\mu = \int f^-\,d\nu$ gilt. Daher sind die Integrale $\int f^+ \circ T\,d\mu$ und $\int f^- \circ T\,d\mu$ genau dann endlich, wenn die Integrale $\int f^+\,d\nu$ und $\int f^-\,d\nu$ endlich sind, d. h., $f \circ T$ liegt in $\mathcal{L}_1(\mu)$ genau dann, wenn f in $\mathcal{L}_1(\nu)$ liegt. Außerdem erhalten wir $\int f \circ T\,d\mu = \int f^+ \circ T\,d\mu - \int f^- \circ T\,d\mu = \int f\,d\nu - \int f^-\,d\nu = \int f\,d\nu$. □

Korollar 2.3 (Invarianz des Integrals) Es sei $T\colon X \longrightarrow X$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Liegt f in $\mathcal{L}_1(\mu)$, dann auch $f \circ T$, und es gilt $\int f \circ T\,d\mu = \int f\,d\mu$.

Beweis Da T maßtreu ist, ist $T_*\mu = \mu$. Somit folgt das Korollar aus Satz 2.2. □

Wir untersuchen jetzt Eigenschaften der Bahnen einer maßtreuen Transformation. Der folgende Satz gibt Auskunft über die Rückkehr typischer Bahnen in eine vorgegebene Menge.

Satz 2.4 (Poincaréscher Rekurrenzsatz) *Sei $T\colon X \longrightarrow X$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Sei weiterhin $E \in \mathcal{S}$ und $\mu(E) > 0$. Für fast jedes $x \in E$ existiert eine unendliche Folge $n_1 < n_2 < \dots$ in $\mathbb{N}$ mit $T^{n_j}(x) \in E$ für $j \geq 1$.*

Beweis Wir setzen $A_m = \bigcup_{k=m}^{\infty} T^{-k}(E)$ für $m \geq 0$ und $A = \bigcap_{m=0}^{\infty} A_m$. Dann ist A die Menge aller $x \in X$ mit der folgenden Eigenschaft: Für jedes $m \geq 0$ existiert ein $k \geq m$ mit $T^k(x) \in E$. Somit existiert für jedes $x \in A$ eine Folge $n_1 < n_2 < \dots$ in $\mathbb{N}$ mit $T^{n_j}(x) \in E$ für $j \geq 1$. Es genügt also $\mu(E \setminus A) = 0$ zu zeigen, was bedeutet, dass fast alle $x \in E$ zur Menge A gehören.

Wegen $T^{-1}(A_m) = A_{m+1}$ und der Maßtreue von T muss $\mu(A_m) = \mu(A_{m+1})$ gelten und daher auch $\mu(A_m) = \mu(A_0)$ für alle $m \geq 1$. Des Weiteren gilt $A_m \supset A_{m+1}$ für $m \geq 0$, woraus wir $\mu(A) = \lim_{m\to\infty} \mu(A_m) = \mu(A_0)$ erhalten. Wegen $E \subset A_0$ und $A \subset A_0$ folgt schließlich $\mu(E \setminus A) \leq \mu(A_0 \setminus A) = \mu(A_0) - \mu(A) = 0$. □

Nun stellt sich die Frage nach der *Häufigkeit*, mit der die Bahn eines Punktes $x \in X$ eine vorgegebene Menge $E \in \mathcal{S}$ besucht. Die Anzahl dieser Besuche in den ersten n Schritten ist $\sum_{j=0}^{n-1} 1_E(T^j(x))$. Wir interessieren uns für die *durchschnittliche Häufigkeit* dieser Besuche, also für $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_E(T^j(x))$, falls dieser Grenzwert existiert. Derartige Fragen werden von den sogenannten *Ergodensätzen* beantwortet, die die Mittelwerte von Funktionen $f\colon X \longrightarrow \mathbb{R}$ entlang der Bahnen „typischer" Punkte im Maßraum behandeln, d. h. $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x))$ für typische $x \in X$. Solche Grenzwerte werden oft auch *Zeitmittel* genannt.

Zur Charakterisierung dieser Mittel benötigen wir einen wichtigen Begriff aus der Wahrscheinlichkeitstheorie.

Definition 2.5 (Bedingte Erwartung)
Es seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum, $f \in \mathcal{L}_1(\mu)$ und $\mathcal{T}$ eine Teil-σ-Algebra von $\mathcal{S}$. Eine Funktion $g \in \mathcal{L}_1(\mu)$, die messbar bezüglich $\mathcal{T}$ ist und $\int 1_A g \, d\mu = \int 1_A f \, d\mu$ für alle $A \in \mathcal{T}$ erfüllt, heißt *bedingte Erwartung von* f*, gegeben* $\mathcal{T}$. Sie wird mit $E(f \,|\, \mathcal{T})$ oder $E_\mu(f \,|\, \mathcal{T})$ bezeichnet.

Mit Hilfe des Satzes von Radon-Nikodym zeigt man, dass für jede Funktion $f \in \mathcal{L}_1(\mu)$ und jede Teil-σ-Algebra $\mathcal{T} \subset \mathcal{S}$ eine bedingte Erwartung $E(f \,|\, \mathcal{T})$ existiert, die in $\mathcal{L}_1(\mu)$ liegt (vgl. [3, Kapitel IX]). Außerdem ist diese bedingte Erwartung fast sicher eindeutig bestimmt: Wenn zwei Funktionen g und g' in $\mathcal{L}_1(\mu)$ liegen und die Definition einer bedingten Erwartung von f erfüllen, dann gilt $g = g' \pmod{\mu}$ (d. h. $g = g'$ *μ-f.ü.*)

Die in Definition 2.5 eingeführte *bedingte Erwartung* $E_\mu(\cdot \,|\, \mathcal{T}) = E(\cdot \,|\, \mathcal{T})$ hat folgende Eigenschaften (vgl. z. B. [12, Kapitel 6] oder [18, Kapitel 14]):

(1) Die Abbildung $f \mapsto E(f \,|\, \mathcal{T})$ ist linear $(\operatorname{mod} \mu)$.
(2) Wenn $f \geq 0 \pmod{\mu}$ ist, so ist auch $E(f \,|\, \mathcal{T}) \geq 0 \pmod{\mu}$.
(3) Für $f \in \mathcal{L}_1(X, \mathcal{S}, \mu)$ und $g \in \mathcal{L}^\infty(X, \mathcal{T}, \mu)$ ist $E(gf \,|\, \mathcal{T}) = gE(f, \mathcal{T}) \pmod{\mu}$.
(4) Die Einschränkung der bedingten Erwartung $E(\cdot \,|\, \mathcal{T})$ auf den in $L_1(X, \mathcal{S}, \mu)$ enthaltenen Hilbertraum $L_2(X, \mathcal{S}, \mu)$ ist der Projektionsoperator auf den abgeschlossenen linearen Teilraum $L_2(X, \mathcal{T}, \mu) \subset L_2(X, \mathcal{S}, \mu)$.

Lemma 2.6 (Jensensche Ungleichung) Es sei $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum. Wenn $J \subset [0, \infty) = \mathbb{R}_+$ ein Intervall und $\phi: J \longrightarrow \mathbb{R}$ eine konvexe[2] Funktion ist, so gilt

$$E(\phi \circ f \,|\, \mathcal{T}) \geq \phi(E(f \,|\, \mathcal{T})).$$

Insbesondere gilt für $p \in [1, \infty]$

$$\|f\|_p \geq \|E(|f| \,|\, \mathcal{T})\|_p \geq \|E(f \,|\, \mathcal{T})\|_p$$

für alle $f \in \mathcal{L}_p(\mu)$. Die bedingte Erwartung ist also auf jedem L_p mit $1 \leq p \leq \infty$ eine Kontraktion.

Aufgabe 2.7
Beweisen Sie Lemma 2.6.

Hinweis: Ersetzen Sie das Integral im Beweis von [3, Satz V.7] durch die bedingte Erwartung. Verwenden Sie anschließend die konvexe Funktion $f(t) = t^p$ für $p \geq 1$. Für $p = \infty$ verwenden Sie die obige Positivitätseigenschaft (2).

Für den Ergodensatz ist die *σ-Algebra $\mathcal{S}^T$ der T-invarianten messbaren Teilmengen* aus dem folgenden Lemma wichtig.

[2] Eine Funktion $\phi: J \longrightarrow \mathbb{R}$ ist *konvex*, wenn $\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$ für $0 \leq t \leq 1$ und $x < y$ in J gilt.

Lemma 2.8 Seien $(X, \mathcal{S})$ ein messbarer Raum und $T: X \longrightarrow X$ eine messbare Transformation. Dann ist $\mathcal{S}^T = \{A \in \mathcal{S} : T^{-1}(A) = A\}$ eine σ-Algebra. Eine Funktion $f: X \longrightarrow \mathbb{R}$ ist genau dann messbar bezüglich $\mathcal{S}^T$, wenn sie messbar bezüglich $\mathcal{S}$ ist und $f \circ T = f$ erfüllt.

Beweis Da T^{-1} mit Mengenoperationen vertauschbar ist, gilt $T^{-1}(\varnothing) = \varnothing$, $T^{-1}(A^c) = (T^{-1}(A))^c$ und $T^{-1}(\bigcup_{n=0}^{\infty} A_n) = \bigcup_{n=0}^{\infty} T^{-1}(A_n)$ für alle Teilmengen A, A_n von X. Daraus folgt, dass $\mathcal{S}^T$ die Eigenschaften einer σ-Algebra erfüllt.

Um die zweite Aussage zu zeigen, sei $f: X \longrightarrow \mathbb{R}$ eine beliebige Funktion. Für $t \in \mathbb{R}$ definieren wir die Mengen $A_t = \{x \in X : f(x) < t\}$.

Ist f messbar bezüglich $\mathcal{S}^T$, dann gilt $A_t \in \mathcal{S}^T$ für alle $t \in \mathbb{R}$. Würde ein $x \in X$ existieren mit $f(T(x)) \neq f(x)$, dann könnte man ein $t \in \mathbb{R}$ finden, sodass A_t einen der Punkte x oder $T(x)$ enthält, den anderen aber nicht. Daraus folgt $T^{-1}(A_t) \neq A_t$, im Widerspruch zur Annahme dass $A_t \in \mathcal{S}^T$ ist. Damit ist $f \circ T = f$ gezeigt. Die Messbarkeit von f bezüglich $\mathcal{S}$ folgt wegen $\mathcal{S}^T \subset \mathcal{S}$.

Ist nun f eine messbare Funktion, die $f \circ T = f$ erfüllt, dann gilt $A_t \in \mathcal{S}$ und $T^{-1}(A_t) = A_t$ und daher auch $A_t \in \mathcal{S}^T$ für alle $t \in \mathbb{R}$. Das zeigt, dass f $\mathcal{S}^T$-messbar ist. □

Da man in der Maßtheorie im Allgemeinen nicht zwischen Mengen unterscheidet, die sich nur um eine Nullmenge (also eine Menge vom Maß null) unterscheiden, kann man auch den Begriff der Invarianz von Mengen unter T abschwächen und Mengen $A \in \mathcal{S}$ betrachten, die sich nur um eine Nullmenge von $T^{-1}(A)$ unterscheiden. Das folgende Lemma besagt, dass das keine wesentliche Änderung des Invarianzbegriffs ist.

Lemma 2.9 Es sei $T: X \longrightarrow X$ eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Dann ist $\mathcal{S}_\mu^T = \{A \in \mathcal{S} : \mu(A \vartriangle T^{-1}(A)) = 0\}$ eine σ-Algebra. Weiterhin existiert zu jedem $A \in \mathcal{S}_\mu^T$ eine Menge $B \in \mathcal{S}^T$ mit $\mu(A \vartriangle B) = 0$.

Beweis Für jede Folge $(A_n)_{n \geq 1}$ in $\mathcal{S}_\mu^T$ ist

$$T^{-1}\Big(\bigcup_{n\geq 1} A_n\Big) \vartriangle \Big(\bigcup_{n\geq 1} A_n\Big) \subset \bigcup_{n\geq 1}\big(T^{-1}(A_n) \vartriangle A_n\big)$$

und daher auch $\bigcup_{n\geq 1} A_n \in \mathcal{S}_\mu^T$. Des Weiteren ist $T^{-1}(X \setminus A) \vartriangle (X \setminus A) = T^{-1}(A) \vartriangle A$ und somit $X \setminus A \in \mathcal{S}_\mu^T$ für jedes $A \in \mathcal{S}_\mu^T$. Daher ist $\mathcal{S}_\mu^T$ abgeschlossen bezüglich Komplementen und abzählbaren Vereinigungen. Da $\mathcal{S}_\mu^T$ die leere Menge enthält, ist $\mathcal{S}_\mu^T$ eine σ-Algebra.

Es sei nun $A \in \mathcal{S}$ mit $\mu(T^{-1}(A) \vartriangle A) = 0$. Da T maßtreu ist, ist

$$\mu(T^{-n}(A) \vartriangle A) \leq \sum_{j=0}^{n-1} \mu(T^{-j-1}(A) \vartriangle T^{-j}(A)) = \sum_{j=0}^{n-1} \mu(T^{-j}(T^{-1}(A) \vartriangle A)) = 0$$

für jedes $n \geq 1$. Damit ist auch

$$\mu\Big(A \triangle \bigcup_{j \geq n} T^{-j}(A)\Big) \leq \sum_{j \geq 0} \mu(A \triangle T^{-j}(A)) = 0$$

für jedes $n \geq 0$. Wir setzen $A_n = \bigcup_{j \geq n} T^{-j}(A)$ und $A_\infty = \bigcap_{n \geq 0} A_n = T^{-1}(A_\infty)$. Da die Folge $(A_n)_{n \geq 0}$ monoton abnehmend ist, erhalten wir aus der Stetigkeit von μ, dass $\mu(A \triangle A_\infty) = \lim_{n \to \infty} \mu(A \triangle A_n) = 0$ ist. Damit haben wir eine Menge $B = A_\infty \in \mathcal{S}^T$ gefunden, die $\mu(A \triangle B) = 0$ erfüllt. □

Der folgende Satz wird als *punktweiser* oder *individueller* Ergodensatz bezeichnet, da er die fast sichere punktweise Konvergenz der Zeitmittel zum Inhalt hat. Er wurde im Jahr 1931 von Birkhoff[3] bewiesen. Wir geben hier einen neueren Beweis aus [9], der sich durch seine Kürze auszeichnet.

Satz 2.10 (Individueller Ergodensatz) *Seien $T: X \longrightarrow X$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ und $\mathcal{S}^T$ die σ-Algebra der T-invarianten messbaren Teilmengen von X. Für jedes $f \in \mathcal{L}_1(\mu)$ konvergiert das „ergodische Mittel“ $\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j$ für $n \to \infty$ fast sicher gegen $E(f \mid \mathcal{S}^T)$.*

Beweis Ist $h: X \longrightarrow \mathbb{R}$ eine Funktion, dann definieren wir $S_k h = \sum_{j=0}^{k-1} h \circ T^j$ für $k \geq 1$ und $S_0 h = 0$. Sei $f \in \mathcal{L}_1(\mu)$, $\varepsilon > 0$ und $g = f - E(f \mid \mathcal{S}^T) - \varepsilon$. Es gilt $g \in \mathcal{L}_1(\mu)$, da mit f auch $E(f \mid \mathcal{S}^T)$ in $\mathcal{L}_1(\mu)$ liegt.

Wir wollen zeigen, dass $\limsup_{n\to\infty} \frac{1}{n} S_n g$ fast sicher ≤ 0 ist. Da aus der Bedingung $\limsup_{n\to\infty} \frac{1}{n} S_n g > 0$ auch $\limsup_{n\to\infty} S_n g = \infty$ folgt, genügt es zu zeigen, dass $A = \{x \in X : \sup_{k \geq 1} S_k g(x) = \infty\}$ eine Nullmenge ist.

Es gilt $S_k g(T(x)) = S_{k+1} g(x) - g(x)$ für $k \geq 0$. Daraus folgt, dass $T(x) \in A$ genau dann gilt, wenn $x \in A$ gilt, womit $A \in \mathcal{S}^T$ gezeigt ist. Für $n \geq 1$ setzen wir $M_n = \max_{0 \leq k \leq n} S_k g$. Dann ist $M_{n+1}(x) = \max\{0, g(x) + M_n(T(x))\}$ und daher auch

$$M_{n+1}(x) - M_n(T(x)) = \max\{-M_n(T(x)), g(x)\}.$$

Daraus folgt, dass $\lim_{n\to\infty} M_{n+1}(x) - M_n(T(x)) = g(x)$ für alle $x \in A$ gilt und dass $g(x) \leq M_{n+1}(x) - M_n(T(x)) \leq \max\{0, g(x)\}$ für alle $n \geq 1$ und $x \in X$ erfüllt ist. Nun liegt g in $\mathcal{L}_1(\mu)$, sodass $\lim_{n\to\infty} \int 1_A (M_{n+1} - M_n \circ T)\, d\mu = \int 1_A g \, d\mu$ aus dem Satz über dominierte Konvergenz folgt.

Oben wurde $T^{-1}(A) = A$ gezeigt, d. h. $1_A \circ T = 1_A$. Für $n \geq 1$ ist dann $\int 1_A (M_n \circ T)\, d\mu = \int (1_A \circ T)(M_n \circ T)\, d\mu = \int 1_A M_n \, d\mu$ wegen Korollar 2.3, woraus wir $\int 1_A (M_{n+1} -$

[3] George David Birkhoff (1884–1944) war ein einflussreicher amerikanischer Mathematiker, der v. a. für seinen individuellen Ergodensatz bekannt ist.

$M_n \circ T)\, d\mu = \int 1_A M_{n+1}\, d\mu - \int 1_A M_n\, d\mu \geq 0$ erhalten, da $M_{n+1} \geq M_n$ unmittelbar aus der Definition folgt. Damit ist $\int 1_A g\, d\mu \geq 0$ gezeigt.

Da A in $\mathcal{S}^T$ liegt, folgt $\int 1_A f\, d\mu = \int 1_A E(f\,|\,\mathcal{S}^T)\, d\mu$ aus den Eigenschaften der bedingten Erwartung. Die Definition von g ergibt dann $\int 1_A g\, d\mu = -\varepsilon\mu(A)$, woraus $\mu(A) = 0$ folgt.

Es sei nun $U_\varepsilon = X \setminus A$. Dann gilt $\mu(U_\varepsilon) = 1$ und $\sup_{k\geq 1} S_k g(x) < \infty$ für alle $x \in U_\varepsilon$. Da $E(f\,|\,\mathcal{S}^T)$ messbar ist bezüglich $\mathcal{S}^T$, gilt $E(f\,|\,\mathcal{S}^T) \circ T = E(f\,|\,\mathcal{S}^T)$ nach Lemma 2.8. Es folgt $\frac{1}{n} S_n E(f\,|\,\mathcal{S}^T) = E(f\,|\,\mathcal{S}^T)$ für alle $n \geq 1$. Für $x \in U_\varepsilon$ ist $\limsup_{n\to\infty} \frac{1}{n} S_n g(x) \leq 0$. Nach Definition von g folgt daraus $\limsup_{n\to\infty} \frac{1}{n} S_n f(x) \leq E(f\,|\,\mathcal{S}^T)(x) + \varepsilon$.

Dasselbe kann man für $-f$ anstelle von f beweisen: Es existiert eine Menge $V_\varepsilon \subset X$ mit $\mu(V_\varepsilon) = 1$ und $\limsup_{n\to\infty} -\frac{1}{n} S_n f(x) \leq -E(f\,|\,\mathcal{S}^T)(x) + \varepsilon$ für alle $x \in V_\varepsilon$. Es folgt $\liminf_{n\to\infty} \frac{1}{n} S_n f(x) \geq E(f\,|\,\mathcal{S}^T)(x) - \varepsilon$ für alle $x \in V_\varepsilon$ wegen der Identität $E(-f\,|\,\mathcal{S}^T) = -E(f\,|\,\mathcal{S}^T)$. Setzt man jetzt $G = \bigcap_{m=1}^\infty U_{1/m} \cap \bigcap_{m=1}^\infty V_{1/m}$, dann ist $\mu(G) = 1$ und $\lim_{n\to\infty} \frac{1}{n} S_n f(x) = E(f\,|\,\mathcal{S}^T)(x)$ für alle $x \in G$. □

Als Folgerung aus dem individuellen Ergodensatz erhalten wir den L_p-Ergodensatz, der im Jahr 1930 von John von Neumann[4] bewiesen wurde.

Satz 2.11 (**L_p-Ergodensatz**) *Sei $T\colon X \longrightarrow X$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ und $\mathcal{S}^T \subset \mathcal{S}$ die σ-Algebra der invarianten Teilmengen. Sei weiterhin $p \geq 1$. Für alle $f \in \mathcal{L}_p(\mu)$ liegt dann $E(f\,|\,\mathcal{S}^T)$ ebenfalls in $\mathcal{L}_p(\mu)$, und es gilt* $\lim_{n\to\infty} \|\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - E(f\,|\,\mathcal{S}^T)\|_p = 0$.

Beweis Es sei $f \in \mathcal{L}_p(\mu)$. Aus Lemma 2.6 wissen wir, dass $E(f\,|\,\mathcal{S}^T) \in \mathcal{L}_p(\mu)$ ist.

Für $m \geq 1$ sei $C_m = \{x \in X : |f(x)| \leq m\}$ und $f_m = f 1_{C_m}$. Dann ist $\lim_{m\to\infty} |f(x) - f_m(x)|^p = 0$ für alle $x \in X$. Des Weiteren gilt $|f - f_m|^p \leq |f|^p$ für alle $m \geq 1$ und $\int |f|^p d\mu < \infty$, sodass man aufgrund des Satzes über dominierte Konvergenz die Gleichung $\lim_{m\to\infty} \int |f - f_m|^p d\mu = 0$ erhält. Damit ist gezeigt, dass $\lim_{m\to\infty} \|f - f_m\|_p = 0$ ist.

Für $n \geq 1$ und eine Funktion $h\colon X \longrightarrow \mathbb{R}$ sei $S_n h = \sum_{j=0}^{n-1} h \circ T^j$ wie im letzten Beweis. Sei $\varepsilon > 0$ vorgegeben. Es existiert ein $m \geq 1$ mit $\|f - f_m\|_p < \frac{\varepsilon}{4}$. Wegen Korollar 2.3 gilt $\|f \circ T^j - f_m \circ T^j\|_p = \|f - f_m\|_p < \frac{\varepsilon}{4}$ für alle $j \geq 1$, woraus $\|\frac{1}{n} S_n f - \frac{1}{n} S_n f_m\|_p \leq \frac{1}{n} \sum_{j=0}^{n-1} \|f \circ T^j - f_m \circ T^j\|_p < \frac{\varepsilon}{4}$ für alle $n \geq 1$ mit Hilfe der Dreiecksungleichung folgt. Da die bedingte Erwartung eine Kontraktion auf $\mathcal{L}_p(\mu)$ ist (Lemma 2.6), gilt auch $\|E(f\,|\,\mathcal{S}^T) - E(f_m\,|\,\mathcal{S}^T)\|_p = \|E(f - f_m\,|\,\mathcal{S}^T)\|_p \leq \|f - f_m\|_p < \frac{\varepsilon}{4}$.

[4] John von Neumann (1903–1957), geboren in Budapest, arbeitete ab 1930 in den USA. Seine Beiträge reichten von der Mathematik (z. B. Mengenlehre, Funktionalanalysis, Ergodentheorie, Geometrie, numerische Analysis), Physik (z. B. Quantenmechanik, Hydrodynamik, Strömungsdynamik), Ökonomie (Spieltheorie), Informatik (u. a. lineares Programmieren, Computerarchitektur, theoretische Computerwissenschaften) bis hin zur Statistik. Er gilt als einer der bedeutendsten Mathematiker der jüngeren Geschichte.

Wegen $|f_m| \le m$ ist auch $|E(f_m\,|\,\mathcal{S}^T)| \le m$ und $|f_m \circ T^j| \le m$ für $j \ge 1$, woraus wir $|\frac{1}{n}S_n f_m - E(f_m\,|\,\mathcal{S}^T)|^p \le (2m)^p$ für alle $n \ge 1$ erhalten. Ebenfalls wegen $|f_m| \le m$ liegt f_m in $\mathcal{L}_1(\mu)$, sodass $\lim_{n\to\infty} \frac{1}{n}S_n f_m = E(f_m\,|\,\mathcal{S}^T)$ fast sicher nach Satz 2.10 gilt. Nun folgt $\lim_{n\to\infty} \int |\frac{1}{n}S_n f_m - E(f_m\,|\,\mathcal{S}^T)|^p\,d\mu = 0$ aus dem Satz über dominierte Konvergenz und daraus dann $\lim_{n\to\infty} \|\frac{1}{n}S_n f_m - E(f_m\,|\,\mathcal{S}^T)\|_p = 0$. Es existiert also ein n_0, sodass $\|\frac{1}{n}S_n f_m - E(f_m\,|\,\mathcal{S}^T)|_p < \frac{\varepsilon}{2}$ für alle $n \ge n_0$ gilt. Mit Hilfe der Dreiecksungleichung folgt dann $\|\frac{1}{n}S_n f - E(f\,|\,\mathcal{S}^T)\|_p < \varepsilon$ für alle $n \ge n_0$. Damit ist $\lim_{n\to\infty} \|\frac{1}{n}S_n f - E(f\,|\,\mathcal{S}^T)\|_p = 0$ gezeigt. □

Für $p = 2$ ist der Raum $L_p(\mu,\mathbb{C})$ in Satz 2.11 ein Hilbertraum mit Skalarprodukt $\langle\!\langle [f],[g]\rangle\!\rangle = \int f\bar{g}\,d\mu$. Die durch

$$U_T[f] = [f \circ T], \quad f \in \mathcal{L}_2(\mu) \tag{2.1}$$

gegebene lineare Abbildung $U_T\colon L_2(\mu) \longrightarrow L_2(\mu)$ ist wohldefiniert (da $f_1 \circ T = f_2 \circ T \pmod{\mu}$ ist, wenn $f_1 = f_2 \pmod{\mu}$ ist), und erfüllt die Bedingung $\langle\!\langle U_t[f_1], U_T[f_2]\rangle\!\rangle = \langle\!\langle [f_1],[f_2]\rangle\!\rangle$ und $\|U_T[f_1]\|_2 = \|[f_1]\|_2$ für alle $f_1, f_2 \in \mathcal{L}_2(\mu)$.

2.2 Mischungseigenschaften

Die Ergodensätze besagen, dass die Zeitmittel einer Funktion $f \in \mathcal{L}_1(\mu)$ unter einer maßtreuen Transformation $T\colon X \longrightarrow X$ auf einem Wahrscheinlichkeitsraum $(X,\mathcal{S},\mu)$ fast sicher gegen die bedingte Erwartung $E(f\,|\,\mathcal{S}^T)$ konvergieren. Diese bedingte Erwartung ist von besonders einfacher Form, wenn die σ-Algebra der T-invarianten Teilmengen $\mathcal{S}^T \subset \mathcal{S}$ in der σ-Algebra

$$\mathcal{N}_\mu = \{B \in \mathcal{S} : \mu(B) \in \{0,1\}\} \tag{2.2}$$

enthalten ist (vgl. Definition 1.78):

Definition 2.12 (Ergodizität)
Sei $(X,\mathcal{S},\mu)$ ein Wahrscheinlichkeitsraum. Eine maßtreue Transformation $T\colon X \longrightarrow X$ heißt *ergodisch*, wenn für jedes $A \in \mathcal{S}^T$ entweder $\mu(A) = 0$ oder $\mu(A) = 1$ gilt.

▶ **Bemerkung 2.13 (Bemerkung zum Wort „Ergodizität")** Ist $T\colon X \longrightarrow X$ eine maßtreue ergodische Transformation auf einem Wahrscheinlichkeitsraum $(X,\mathcal{S},\mu)$, so gilt $\mathcal{S}^T \subset \mathcal{N}_\mu$. Für eine Funktion $f \in \mathcal{L}_1(\mu)$ ist dann laut dem folgenden Satz 2.14 die bedingte Erwartung $E(f\,|\,\mathcal{S}^T)$ fast sicher gleich dem Integral $\int f\,d\mu$, d. h., das „Zeitmittel" $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j$ von f entlang der Bahn von x stimmt für fast jedes $x \in X$ mit dem „Raummittel" $\int f\,d\mu$ von f überein.

Ludwig Boltzmann formulierte das Prinzip „*Zeitmittel = Raummittel*“ für eine Klasse von mechanischen Systemen, die von äußeren Einflüssen isoliert sind und sich im Gleichgewichtszustand befinden. Unter der Voraussetzung, dass jedes Teilchen des Systems alle im System unter Erhaltung der Energie möglichen Zustände durchläuft (Boltzmann bezeichnete derartige Systeme als „*Ergoden*“) besagt Boltzmanns Postulat, dass die Wahrscheinlichkeit eines bestimmten Zustands (oder einer Menge von Zuständen) eines Teilchens proportional zum Prozentsatz der Zeit ist, die das Teilchen in diesem Zustand (oder in dieser Menge von Zuständen) verbringt.

Boltzmanns Begriff der Ergode wurde schon bei seinem Erscheinen im Jahr 1884 als zu restriktiv kritisiert und wird heute durch den Begriff der *Ergodizität* im Sinne der Definition 1.78 ersetzt. Der individuelle Ergodensatz ist die exakte Interpretation von Boltzmanns Postulat für ergodische Systeme – allerdings eben nur für *fast jedes* Teilchen des Systems und nicht für *jedes* Teilchen.

Zur Frage der Gültigkeit von Boltzmanns Postulat für *jedes* Teilchen des Systems (d. h. für *jeden* Punkt unseres Raumes X) verweisen wir auf Abschn. 2.3.1.

Für eine ergodische Transformation ist also der in den Ergodensätzen auftretende Grenzwert bereits durch das Integral bestimmt. Die folgenden Sätze 2.14 und 2.18 charakterisieren wichtige Eigenschaften ergodischer Transformationen.

Satz 2.14 (Erste Charakterisierung der Ergodizität) *Sei $T: X \longrightarrow X$ eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Die folgenden Aussagen sind äquivalent.*

(1) T ist ergodisch.
(2) Für $A \in \mathcal{S}$ mit $\mu(T^{-1}(A) \triangle A) = 0$ gilt entweder $\mu(A) = 0$ oder $\mu(A) = 1$.
(3) Ist $f: X \longrightarrow \mathbb{R}$ eine messbare Funktion und $f \circ T = f \pmod{\mu}$, dann ist f fast sicher konstant.
(4) Ist $f: X \longrightarrow \mathbb{R}$ eine beschränkte messbare Funktion und $f \circ T = f \pmod{\mu}$, dann ist f fast sicher konstant.

Beweis Die Implikation (1) ⇒ (2) folgt aus Lemma 2.9.

Um (2) ⇒ (3) zu zeigen, seien $f: X \longrightarrow \mathbb{R}$ messbar und $f \circ T = f \pmod{\mu}$. Für $a \in \mathbb{R}$ sei $C_a = \{x \in X : f(x) < a\}$. Wenn f nicht konstant $(\operatorname{mod} \mu)$ ist, dann gibt es ein $a \in \mathbb{R}$ mit $\mu(C_a) > 0$ und $\mu(X \setminus C_a) > 0$. Wegen $f \circ T = f \pmod{\mu}$ gilt jedoch $\mu(T^{-1}(C_a) \triangle C_a) = 0$, woraus wegen (2) entweder $\mu(C_a) = 0$ oder $\mu(C_a) = 1$ folgt. Dieser Widerspruch beweist (3).

Die Implikation (3) ⇒ (4) ist offensichtlich. Wenn schließlich (4) gilt und A in $\mathcal{S}^T$ liegt, dann gilt $1_A \circ T = 1_A$ und daher (wegen (4)) $1_A = 0$ fast sicher oder $1_A = 1$ fast sicher. Daraus folgt $\mu(A) = 0$ oder $\mu(A) = 1$, womit (1) gezeigt ist. □

Wir wenden Satz 2.14 auf Transformationen des n-dimensionalen Torus an.

Satz 2.15 *Seien* $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$ *und* A *eine* $d \times d$*-Matrix mit Koeffizienten in* $\mathbb{Z}$ *und Determinante* 1 *oder* −1*. Für jedes* $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{Z}^d \smallsetminus \{\mathbf{0}\}$ *sei entweder die Menge* $\{\mathbf{m}A^k : k \in \mathbb{Z}\}$ *unendlich, oder es gelte*

$$\sum_{j=0}^{k-1} \langle \mathbf{m}A^j, \boldsymbol{\alpha} \rangle \notin \mathbb{Z},$$

wenn $\mathbf{m}A^k = \mathbf{m}$ *für ein* $k \geq 1$ *ist. Dann ist die Transformation* $T(\mathbf{x}) = A\mathbf{x} + \boldsymbol{\alpha} \pmod 1$ *auf dem Wahrscheinlichkeitsraum* $(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d)$ *ergodisch.*

Beweis Im Sinne der Bemerkungen auf Seite 48 unterscheiden wir nicht zwischen Funktionen und Äquivalenzklassen in $L_2(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d, \mathbb{C})$.

Für jedes $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{Z}^d$ betrachten wir die bereits in (1.32) eingeführte Funktion $f_{\mathbf{m}}(\mathbf{x}) = e^{2\pi i \sum_{j=1}^d m_j x_j}$, $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{T}^d$. Für $\mathbf{m}, \mathbf{m}' \in \mathbb{Z}^d$ gilt

$$\langle\!\langle f_{\mathbf{m}}, f_{\mathbf{m}'} \rangle\!\rangle = \begin{cases} 1 & \text{für } \mathbf{m} = \mathbf{m}', \\ 0 & \text{für } \mathbf{m} \neq \mathbf{m}', \end{cases}$$

wobei $\langle\!\langle f, g \rangle\!\rangle = \int f \bar{g} \, d\lambda^d$ das Skalarprodukt in $L_2(\lambda^d, \mathbb{C})$ ist. Damit bilden diese Funktionen ein Orthonormalsystem in $L_2(\lambda^d, \mathbb{C})$. Wie wir in Aufgabe 1.35 gesehen haben, liegen die endlichen komplexen Linearkombinationen dieser Funktionen dicht in $C(\mathbb{T}^d)$ und daher auch in $L_2(\lambda^d, \mathbb{C})$. Daraus folgt, dass $\{f_{\mathbf{m}} : \mathbf{m} \in \mathbb{Z}^d\}$ eine Basis (genauer gesagt: ein *vollständiges* Orthonormalsystem) des Hilbertraums $L_2(\lambda^d, \mathbb{C})$ bilden. Wie man leicht nachrechnet, ist $U_T(f_{\mathbf{m}}) = f_{\mathbf{m}} \circ T = e^{2\pi i \langle \mathbf{m}, \boldsymbol{\alpha} \rangle} f_{\mathbf{m}A}$ für jedes $\mathbf{m} \in \mathbb{Z}^d$.

Es sei nun $g \colon \mathbb{T}^d \longrightarrow \mathbb{R}$ eine beschränkte messbare Funktion, die $g \circ T = g \pmod{\lambda^d}$ erfüllt. Dann liegt g in $L_2(\lambda^d) \subset L_2(\lambda^d, \mathbb{C})$ und kann daher auf eindeutige Weise in der Basis $\{f_{\mathbf{m}} : \mathbf{m} \in \mathbb{Z}^d\}$ entwickelt werden:

$$g = \sum_{\mathbf{m} \in \mathbb{Z}^d} c_{\mathbf{m}} f_{\mathbf{m}}, \tag{2.3}$$

wobei $c_{\mathbf{m}} = \langle\!\langle g, f_{\mathbf{m}} \rangle\!\rangle = \int g \overline{f_{\mathbf{m}}} \, d\lambda^d = \int g f_{-\mathbf{m}} \, d\lambda^d$ für jedes $\mathbf{m} \in \mathbb{Z}^d$ ist und wobei die Summe in (2.3) in der L_2-Norm konvergiert und die Bedingung $\|g\|_2^2 = \sum_{\mathbf{m} \in \mathbb{Z}^n} |c_{\mathbf{m}}|^2$ erfüllt. Die Darstellung (2.3) ist natürlich nichts anderes als die Entwicklung von g als *Fourierreihe*. Da U_T eine Isometrie auf $L_2(\lambda^d, \mathbb{C})$ ist, gilt

$$g = \sum_{\mathbf{m} \in \mathbb{Z}^d} c_{\mathbf{m}} f_{\mathbf{m}} = g \circ T = \sum_{\mathbf{m} \in \mathbb{Z}^d} c_{\mathbf{m}} f_{\mathbf{m}} \circ T = \sum_{\mathbf{m} \in \mathbb{Z}^d} c_{\mathbf{m}} e^{2\pi i \langle \mathbf{m}, \boldsymbol{\alpha} \rangle} f_{\mathbf{m}A}.$$

Wegen der Eindeutigkeit der Darstellung (2.3) gilt $c_{\mathbf{m}A} = c_{\mathbf{m}} e^{2\pi i \langle \mathbf{m}, \boldsymbol{\alpha} \rangle}$ für alle $\mathbf{m} \in \mathbb{Z}^d$.

Wenn nun $\mathbf{m} \in \mathbb{Z}^d \smallsetminus \{\mathbf{0}\}$ ist, so gilt $|c_{\mathbf{n}}| = |c_{\mathbf{m}}|$ für alle $\mathbf{n} \in \{\mathbf{m}A^k : k \in \mathbb{Z}\}$. Wenn die Menge $\{\mathbf{m}A^k : k \in \mathbb{Z}\}$ unendlich ist, dann folgt $c_{\mathbf{m}} = 0$ wegen $\sum_{\mathbf{m} \in \mathbb{Z}^d} |c_{\mathbf{m}}|^2 < \infty$.

Wenn $\mathbf{m}A^k = \mathbf{m}$ für ein $k \geq 1$ und $\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}$ ist, so ist laut Voraussetzung $c_{\mathbf{m}A} = c_{\mathbf{m}} e^{2\pi i \langle \mathbf{m}, \boldsymbol{\alpha} \rangle}$ und $c_{\mathbf{m}A^2} = c_{\mathbf{m}} e^{2\pi i \langle \mathbf{m}, \boldsymbol{\alpha} \rangle} e^{2\pi i \langle \mathbf{m}A, \boldsymbol{\alpha} \rangle}$. Nach Voraussetzung und mittels Induktion ergibt sich also $c_{\mathbf{m}} = c_{\mathbf{m}A^k} = c_{\mathbf{m}} e^{2\pi i \sum_{j=0}^{k-1} \langle \mathbf{m}A^j, \boldsymbol{\alpha} \rangle}$ und $e^{2\pi i \sum_{j=0}^{k-1} \langle \mathbf{m}A^j, \boldsymbol{\alpha} \rangle} \neq 1$ und daher wieder $c_{\mathbf{m}} = 0$. Somit gilt $c_{\mathbf{m}} = 0$ für alle $\mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$, womit gezeigt ist, dass g fast sicher konstant ist. Nach Satz 2.14 ist T ergodisch. □

Beispiel 2.16 (Rotationen auf $\mathbb{T}^d$)
Seien $\boldsymbol{\alpha} \in \mathbb{R}^d$ und $R_{\boldsymbol{\alpha}}: \mathbb{T}^d \longrightarrow \mathbb{T}^d$ durch $R_{\boldsymbol{\alpha}}(\mathbf{x}) = \mathbf{x} + \boldsymbol{\alpha} \pmod 1$ definiert. Nach Beispiel 1.62 ist $R_{\boldsymbol{\alpha}}$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d)$. Wählt man in Satz 2.15 für A die Einheitsmatrix, dann erhält man, dass $R_{\boldsymbol{\alpha}}$ ergodisch ist, wenn $\mathbf{n}\boldsymbol{\alpha} \notin \mathbb{Z}$ für alle $\mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ ist. Das ist äquivalent zu der schon aus Aufgabe 1.35 bekannten Bedingung, dass $1, \alpha_1, \ldots, \alpha_d$ linear unabhängig über $\mathbb{Q}$ sind.

Beispiel 2.17 (Automorphismen von $\mathbb{T}^d$)
Sei A eine $d \times d$-Matrix mit Koeffizienten in $\mathbb{Z}$ und Determinante 1 oder −1. Die Transformation $T_A: \mathbb{T}^d \longrightarrow \mathbb{T}^d$ sei definiert durch $T_A(\mathbf{x}) = A\mathbf{x} \pmod 1$. Nach Beispiel 1.62 ist T_A wiederum eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d)$. Wir zeigen, dass die Transformation T_A ergodisch ist, wenn keiner der Eigenwerte von A eine Einheitswurzel ist. Das folgt aus Satz 2.15, wenn wir zeigen, dass unter dieser Voraussetzung die Menge $\{\mathbf{n}A^k : k \in \mathbb{Z}\}$ für jedes $\mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ unendlich ist.

Nehmen wir an, die Menge $\{\mathbf{n}A^k : k \in \mathbb{Z}\}$ sei für ein $\mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ endlich. Dann existieren k und l in $\mathbb{Z}$ mit $k < l$ und $\mathbf{n}A^k = \mathbf{n}A^l$. Für $m = l - k \in \mathbb{N}$ gilt dann $\mathbf{n}A^m = \mathbf{n}$. Damit hat A^m den linken Eigenvektor $\mathbf{n} \neq 0$ zum Eigenwert 1. Da die Mengen $\mathrm{EV}(A)$ und $\mathrm{EV}(A^m)$ der Eigenwerte von A und A^m die Bedingung $\mathrm{EV}(A^m) = \mathrm{EV}(A)^m = \{\gamma^m : \gamma \in \mathrm{EV}(A)\}$ erfüllen, gibt es ein $\gamma \in \mathrm{EV}(A)$ mit $\gamma^m = 1$, im Widerspruch zu unserer Voraussetzung.

Wenn also keiner der Eigenwerte von A eine Einheitswurzel ist, dann ist die Menge $\{\mathbf{n}A^k : k \in \mathbb{Z}\}$ für jedes $\mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ unendlich und T_A ist nach Satz 2.15 ergodisch.

Für topologische dynamische Systeme haben wir in Kap. 1 den Begriff der topologischen Transitivität eingeführt, der der Ergodizität maßerhaltender Transformationen auf Wahrscheinlichkeitsräumen entspricht. Wie in der topologischen Dynamik gibt es auch in der Ergodentheorie stärkere Formen der Transitivität, die ebenfalls als „Mischungseigenschaften“ bezeichnet werden. Bevor wir diese formulieren, charakterisieren wir im folgenden Satz Ergodizität als eine quantitative Aussage über die „Mischung“ von vorgegebenen Mengen $A, B \in \mathcal{S}$ durch T, d. h. über die Zahlen $\mu(T^{-n}(A) \cap B)$, $n \geq 1$.

Satz 2.18 (Zweite Charakterisierung der Ergodizität) *Seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T: X \longrightarrow X$ eine maßtreue Transformation. Die folgenden Aussagen sind äquivalent.*

(1) T ist ergodisch.
(2) Es gilt $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) = \mu(A)\mu(B)$ für alle $A, B \in \mathcal{S}$.
(3) Es gilt $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap A) = \mu(A)^2$ für alle $A \in \mathcal{S}$.

Beweis Um (1) $\Rightarrow$ (2) zu zeigen, nehmen wir zwei Mengen A und B in $\mathcal{S}$. Da wir (1) voraussetzen, folgt $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A \circ T^j = \int 1_A \, d\mu = \mu(A)$ fast sicher aus dem Ergodensatz. Wenn wir mit 1_B multiplizieren und den Satz über dominierte Konvergenz anwenden, erhalten wir $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) = \mu(A)\mu(B)$. Damit ist (2) gezeigt.

Die Implikation (2) $\Rightarrow$ (3) ist offensichtlich. Um (3) $\Rightarrow$ (1) zu zeigen, sei $A \in \mathcal{S}^T$ beliebig. Da wir (3) voraussetzen, ist $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap A) = \mu(A)^2$ erfüllt. Da A in $\mathcal{S}^T$ liegt, gilt $T^{-j}(A) \cap A = A$ für alle $j \geq 0$, sodass $\mu(A) = \mu(A)^2$ folgt. Also ist $\mu(A)$ entweder 0 oder 1, und (1) ist gezeigt. □

Definition 2.19 (Mischungseigenschaften)

Es sei $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum. Eine maßtreue Transformation $T\colon X \longrightarrow X$ heißt *schwach mischend*, wenn

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)| = 0 \tag{2.4}$$

für alle $A, B \in \mathcal{S}$ gilt, und *stark mischend*, wenn

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \tag{2.5}$$

für alle $A, B \in \mathcal{S}$ gilt.

Eine stark mischende Transformation ist schwach mischend und eine schwach mischende Transformation ist ergodisch wegen Satz 2.18.

Wir erinnern daran, dass im Sinne der Wahrscheinlichkeitstheorie zwei messbare Mengen A, B in einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ *unabhängig* genannt werden, wenn $\mu(A \cap B) = \mu(A)\mu(B)$ gilt. Die Gleichung (2.5) beschreibt also die *asymptotische Unabhängigkeit* von $T^{-n}(A)$ und B.

Auch wenn die Definition der schwachen Mischung in (2.4) auf den ersten Blick einen künstlichen Eindruck erweckt, so werden wir in Satz 2.24 sehen, dass dieser Begriff sehr natürliche äquivalente Formulierungen besitzt. In der Tat ist der Begriff der schwachen Mischung für einen großen Teil der Ergodentheorie weitaus wichtiger als der Begriff der starken Mischung.

Satz 2.20 *Seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T\colon X \longrightarrow X$ eine maßtreue Transformation. Weiterhin sei $\mathcal{E} \subset \mathcal{S}$ ein Semiring, der die σ-Algebra $\mathcal{S}$ erzeugt. Wenn $\lim_{n\to\infty} \mu(T^{-n}(U) \cap V) = \mu(U)\mu(V)$ für alle $U, V \in \mathcal{E}$ gilt, dann ist T stark mischend. Wenn $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(U) \cap V) = \mu(U)\mu(V)$ (bzw. Bedingung (2.4)) für alle $U, V \in \mathcal{E}$ gilt, dann ist T ergodisch (bzw. schwach mischend).*

Beweis Seien $V \in \mathcal{E}$ und $\mathcal{D}_V = \{A \in \mathcal{S} : \lim_{n\to\infty} \mu(T^{-n}(A) \cap V) = \mu(A)\mu(V)\}$. Wir zeigen, dass $\mathcal{D}_V$ eine σ-Algebra ist, die $\mathcal{E}$ enthält und die daher mit $\mathcal{S}$ übereinstimmt.

Für $A \in \mathcal{D}_V$ gilt $\mu(T^{-n}(A^c) \cap V) = \mu((T^{-n}(A))^c \cap V) = \mu(V) - \mu(T^{-n}(A) \cap V)$ für jedes $n \geq 0$. Damit erhalten wir, dass $\lim_{n\to\infty} \mu(T^{-n}(A^c) \cap V) = \mu(V) - \lim_{n\to\infty} \mu(T^{-n}(A) \cap V) = (1 - \mu(A))\mu(V) = \mu(A^c)\mu(V)$, sodass $\mathcal{D}_V$ abgeschlossen bezüglich Komplementbildung ist.

Wenn $A = \bigcup_{j=1}^m A_j$ eine Vereinigung paarweise disjunkter Mengen in $\mathcal{E}$ ist, so gilt $\mu(T^{-n}(A) \cap V) = \sum_{j=1}^m \mu(T^{-n}(A_j) \cap V)$ für $n \geq 1$ und daher

$$\begin{aligned}\lim_{n\to\infty} \mu(T^{-n}(A) \cap V) &= \lim_{n\to\infty} \sum_{j=1}^m \mu(T^{-n}(A_j) \cap V)\\ &= \sum_{j=1}^m \mu(A_j)\mu(V) = \mu(A)\mu(V).\end{aligned}$$

Damit enthält $\mathcal{D}_V$ alle endlichen Vereinigungen disjunkter Mengen in $\mathcal{E}$. Da sich jede endliche Vereinigung von Mengen in $\mathcal{E}$ aufgrund der Eigenschaften eines Semirings auch als endliche Vereinigung disjunkter Mengen in $\mathcal{E}$ schreiben lässt und da $\mathcal{D}_V$ unter Komplementbildung abgeschlossen ist, enthält $\mathcal{D}_V$ die von $\mathcal{E}$ erzeugte Mengenalgebra.

Sei nun $(A_j)_{j\geq 1} \in \mathcal{D}_V$ eine monotone Folge, sodass entweder $A_j \subset A_{j+1}$ für alle $j \geq 1$ oder $A_j \supset A_{j+1}$ für alle $j \geq 1$ gilt. Wir behaupten, dass dann auch $\bigcup_{j\geq 1} A_j \in \mathcal{D}_V$ beziehungsweise $\bigcap_{j\geq 1} A_j \in \mathcal{D}_V$ gilt. Da wir schon wissen, dass $\mathcal{D}_V$ unter Komplementbildung abgeschlossen ist, genügt es, den aufsteigenden Fall $A_j \subset A_{j+1}$ für alle $j \geq 1$ zu betrachten. Sei $A = \bigcup_{j\geq 1} A_j$ und $\varepsilon > 0$, dann gibt es ein $j \geq 1$ mit $\mu(A \setminus A_j) < \varepsilon/3$. Daher folgt $|\mu(A \cap V) - \mu(A_j \cap V)| < \varepsilon/3$ und auch $|\mu(T^{-n}A \cap V) - \mu(T^{-n}A_j \cap V)| < \varepsilon/3$ für alle $n \geq 1$. Da $A_j \in \mathcal{D}_V$, gibt es ein n_0 mit $|\mu(T^{-n}A_j \cap V) - \mu(A_j)\mu(V)| < \varepsilon/3$ für alle $n \geq n_0$, was $|\mu(T^{-n}A \cap V) - \mu(A)\mu(V)| < \varepsilon$ zur Folge hat. Daher gilt $A \in \mathcal{D}_V$. Wir haben somit gesehen, dass $\mathcal{D}_V$ eine monotone Klasse[5] ist. Wegen des Satzes über monotone Klassen (Aufgabe 13) gilt daher, dass $\mathcal{D}_V = \mathcal{S}$ ist.

Seien jetzt $A \in \mathcal{S}$ und $\mathcal{D}'_A = \{B \in \mathcal{S} : \lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)\}$. Ganz ähnlich wie oben zeigt man, dass $\mathcal{D}'_A$ eine monotone Klasse ist, die wiederum die von $\mathcal{E}$ erzeugte Algebra enthält. Nach dem Satz über monotone Klassen ist somit $\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$ für alle A und B in $\mathcal{S}$ gezeigt, d. h., T ist stark mischend.

Der Beweis der zweiten Aussage über Ergodizität und schwache Mischung ist völlig analog. □

Beispiel 2.21 (Multiplikation mit 2)
Sei $T: \mathbb{T} \longrightarrow \mathbb{T}$, wie in den Beispiel 1.3 durch $T(x) = 2x \pmod 1$ definiert. Laut Beispiel 1.61 (2) ist T eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(\mathbb{T}, \mathcal{B}_{\mathbb{T}}, \lambda)$. Wir zeigen mit Hilfe von Satz 2.20, dass T stark mischend ist.

Seien $\mathcal{E}_k = \{[\frac{j}{2^k}, \frac{j+1}{2^k}) : 0 \leq j \leq 2^k - 1\}$ für $k \geq 1$ und $\mathcal{E} = \bigcup_{k=1}^\infty \mathcal{E}_k$. Dann ist $\mathcal{E}$ ein Semiring, der die Borel-σ-Algebra $\mathcal{B}_{\mathbb{T}}$ erzeugt. Seien U und V in $\mathcal{E}$ und k und m so, dass $U \in \mathcal{E}_m$ und $V \in \mathcal{E}_k$ gilt. Sei $n \geq k$. Dann ist V die disjunkte Vereinigung von 2^{n-k} Intervallen $V_1, V_2, \ldots, V_{2^{n-k}}$ aus $\mathcal{E}_n$. Nach

[5] Eine Mengenfamilie $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ ist eine *monotone Klasse*, wenn mit jeder monoton nicht abnehmenden Folge (A_j) in $\mathcal{M}$ auch $\bigcup_{j\geq 1} A_j \in \mathcal{M}$ ist und mit jeder monoton nicht zunehmenden Folge (A_j) auch $\bigcap_{j\geq 1} A_j \in \mathcal{M}$ ist.

Beispiel 1.3 bildet T^n jedes dieser Intervalle linear und bijektiv auf $[0,1)$ ab. Daher ist $T^{-n}(U) \cap V_i$ für $1 \leq i \leq 2^{n-k}$ ein Intervall der Länge $\frac{1}{2^{m+n}}$, und $T^{-n}(U) \cap V$ ist die disjunkte Vereinigung von 2^{n-k} Intervallen der Länge $\frac{1}{2^{m+n}}$. Für $n \geq k$ erhalten wir also $\lambda(T^{-n}(U) \cap V) = 2^{n-k}\frac{1}{2^{m+n}} = \frac{1}{2^{m+k}} = \lambda(U)\lambda(V)$. Nach Satz 2.20 ist T stark mischend.

Sind $(X_1, \mathcal{S}_1, \mu_1)$ und $(X_2, \mathcal{S}_2, \mu_2)$ Wahrscheinlichkeitsräume, dann ist deren Produkt $(X_1 \times X_2, \mathcal{S}_1 \otimes \mathcal{S}_2, \mu_1 \otimes \mu_2)$ wieder ein Wahrscheinlichkeitsraum (vgl. [3, Kapitel VIII]). Sind $T_1 \colon X_1 \longrightarrow X_1$ und $T_2 \colon X_2 \longrightarrow X_2$ maßtreue Transformationen, dann ist auch $T_1 \times T_2 \colon X_1 \times X_2 \longrightarrow X_1 \times X_2$, definiert durch $(T_1 \times T_2)(x_1, x_2) = (T_1(x_1), T_2(x_2))$, eine maßtreue Transformation.

Um schwache Mischung in einer anschaulicheren Form zu charakterisieren, benötigen wir noch ein weiteres wichtiges Konzept.

Definition 2.22 (Eigenfunktionen)

Es sei T eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Eine messbare Funktion $f \colon X \longrightarrow \mathbb{C}$, die nicht $= 0 \pmod{\mu}$ ist, stellt eine *Eigenfunktion* von T mit Eigenwert $\gamma \in \mathbb{C}$ dar, wenn $f \circ T = \gamma f \pmod{\mu}$ ist.

Aufgaben 2.23 (Eigenschaften von Eigenfunktionen)
Es sei T eine maßtreue ergodische Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Zeigen Sie Folgendes:

(1) Es sei $f \colon X \longrightarrow \mathbb{C}$ eine Eigenfunktion von T mit Eigenwert γ. Dann ist $|\gamma| = 1$ und $|f|$ ist konstant $(\operatorname{mod} \mu)$.
(2) Wenn f_1 und f_2 Eigenfunktionen von T mit demselben Eigenwert γ sind, so ist f_2 ein konstantes Vielfaches von f_1 mod μ).
(3) Wenn f_1 und f_2 Eigenfunktionen von T mit verschiedenen Eigenwerten sind, so ist $\langle\langle f_1, f_2 \rangle\rangle = \int f_1 \overline{f_2}\, d\mu = 0$.

Satz 2.24 (Charakterisierung der schwachen Mischung) *Sei $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T \colon X \longrightarrow X$ eine maßtreue Transformation. Die folgenden Aussagen sind äquivalent.*

(1) T ist schwach mischend.
(2) Die Transformation $T \times T$ auf $(X \times X, \mathcal{S} \otimes \mathcal{S}, \mu \otimes \mu)$ ist ergodisch.
(3) Die Transformation $T \times T$ auf $(X \times X, \mathcal{S} \otimes \mathcal{S}, \mu \otimes \mu)$ ist schwach mischend.
(4) T ist ergodisch und jede Eigenfunktion von T ist konstant $(\operatorname{mod} \mu)$.

Beweis Wir schreiben S für $T \times T$ und $\mathcal{C}$ für $\mathcal{S} \otimes \mathcal{S}$. Das Produktmaß $\mu \otimes \mu$ bezeichnen wir mit ν.

Wir zeigen (1) $\Rightarrow$ (3). Sei $\mathcal{G} = \{A \times B : A, B \in \mathcal{S}\}$. Dann ist $\mathcal{G}$ ein Semiring, der die σ-Algebra $\mathcal{C}$ erzeugt. Für Mengen $U = A \times B$ und $V = C \times D$ in $\mathcal{G}$ und für $j \geq 0$ gilt dann $\nu(S^{-j}(U) \cap V) = \mu(T^{-j}(A) \cap C)\mu(T^{-j}(B) \cap D)$. Verwendet man auch, dass $\nu(U) = \mu(A)\mu(B)$ und $\nu(V) = \mu(C)\mu(D)$ ist, dann erhält man

$$\frac{1}{n}\sum_{j=0}^{n-1} |\nu(S^{-j}(U) \cap V) - \nu(U)\nu(V)| \le \frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(A) \cap C) - \mu(A)\mu(C)|\mu(T^{-j}(B) \cap D) + \frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(B) \cap D) - \mu(B)\mu(D)|\mu(A)\mu(C)$$

für jedes $n \ge 1$. Nun gilt wegen (1), dass $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(A) \cap C) - \mu(A)\mu(C)| = 0$ und $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(B) \cap D) - \mu(B)\mu(D)| = 0$. Daher ist

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} |\nu(S^{-j}(U) \cap V) - \nu(U)\nu(V)| = 0.$$

Wegen Satz 2.20 ist die Transformation S auf $(X \times X, \mathcal{C}, \nu)$ schwach mischend, womit (3) gezeigt ist.

Da schwach mischende Transformationen immer ergodisch sind, folgt (3) $\Rightarrow$ (2). Wir zeigen jetzt (2) $\Rightarrow$ (1). Seien $A, B \in \mathcal{S}$. Dann können wir (wegen der Annahme in (2)) Satz 2.18 für $U = A \times X$ und $V = B \times X$ und für die Transformation S anwenden, woraus wir

$$\frac{1}{n}\sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) \to \mu(A)\mu(B)$$

für $n \to \infty$ erhalten. Wir können aber auch $U = A \times A$ und $V = B \times B$ setzen, woraus sich

$$\frac{1}{n}\sum_{j=0}^{n-1} \mu(T^{-j}A \cap B)^2 \to \mu(A)^2\mu(B)^2$$

für $n \to \infty$ ergibt. Damit erhalten wir

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} (\mu(T^{-j}A \cap B) - \mu(A)\mu(B))^2 \\ &\quad = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} (\mu(T^{-j}A \cap B)^2 - 2\mu(T^{-j}A \cap B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2) = 0.\end{aligned}$$

Wir setzen $c_j = |\mu(T^{-j}A \cap B) - \mu(A)\mu(B)| \le 1$ und wählen n groß genug, sodass $\frac{1}{n}\sum_{j=0}^{n-1} c_n^2 < \varepsilon^3$ gilt. Dann ist $|\{j : 0 \le j < n \text{ und } c_j^2 \ge \varepsilon^2\}| < \varepsilon n$, und daher gilt

$$\frac{1}{n}\sum_{j=0}^{n-1} c_j = \frac{1}{n}\sum_{c_j < \varepsilon} c_j + \frac{1}{n}\sum_{c_j \ge \varepsilon} c_j < \varepsilon + \varepsilon.$$

Nach Definition von c_n und da $A, B \in \mathcal{S}$ beliebig waren, folgt nun (1).

Wir müssen noch zeigen, dass (4) äquivalent zu den ersten drei Bedingungen ist. Wenn nun (2) gilt, aber T eine nichtkonstante Eigenfunktion $f: X \longrightarrow \mathbb{C}$ mit Eigenwert $\gamma \in \mathbb{C}$ hat, so ist die durch $\phi(x_1, x_2) = f(x_1)\overline{f(x_2)}$ gegebene messbare Funktion $\phi: X_1 \times X_2 \longrightarrow \mathbb{C}$ invariant unter $T \times T$, aber nicht konstant (mod $\mu \otimes \mu$), im Widerspruch zur Annahme (2). Die Ergodizität von T folgt unmittelbar aus der Ergodizität von $T \times T$, da ja mit jeder T-invarianten Menge $A \in \mathcal{S}$ die Menge $A \times X_2$ invariant unter $T \times T$ wäre. Damit ist (2) $\Rightarrow$ (4) gezeigt.

Um jetzt noch (4) $\Rightarrow$ (2) zu zeigen, argumentieren wir indirekt und verwenden die Spektraltheorie kompakter selbstadjungierter Operatoren auf Hilberträumen (vgl. [4, §10]). Sei also $F \in L_2(X \times X, \mathcal{S} \otimes \mathcal{S}, \nu, \mathbb{C})$ eine nichtkonstante S-invariante Funktion auf $X \times X$, wobei wir ohne Beschränkung der Allgemeinheit $\int_{X \times X} F d\nu = 0$ annehmen dürfen. Wenn wir $F_1(x, y) = \frac{1}{2}(F(x, y) + \overline{F(y, x)})$ und $F_2(x, y) = \frac{1}{2i}(F(x, y) - \overline{F(y, x)})$ setzen, so gilt $F = F_1 + iF_2$ und $\overline{F_k(y, x)} = F_k(x, y)$ für $k = 1, 2$. Damit können wir also ohne Beschränkung der Allgemeinheit fordern, dass F auch noch die Bedingung $\overline{F(y, x)} = F(x, y)$ erfüllt. Wir definieren nun einen Integraloperator $\mathcal{K}: L_2(X, \mathcal{S}, \mu, \mathbb{C}) \longrightarrow L_2(X, \mathcal{S}, \mu, \mathbb{C})$ durch die Formel

$$\mathcal{K}(f)(x) = \int_X F(x, y) f(y)\, d\mu(y).$$

Aus dem Satz von Fubini [3, Satz VIII.1] folgt, dass für μ-*f.a.* $x \in X$ die Funktion $y \mapsto F(x, y)$ im $L_2(X, \mathcal{S}, \mu, \mathbb{C})$ liegt. Daher ist $\mathcal{K}(f)(x)$ für μ-*f.a.* $x \in X$ wohldefiniert. Laut [15, Chapter IV, Exercise 15] oder [21, Chapter 3.2] ist die Abbildung $\mathcal{K}: L_2(X, \mathcal{S}, \mu, \mathbb{C}) \longrightarrow L_2(X, \mathcal{S}, \mu, \mathbb{C})$ ein kompakter linearer Operator.

Wir behaupten, dass $\mathcal{K} \neq 0$ ist. Da man jede Funktion in $L_2(X \times X, \mathcal{S} \otimes \mathcal{S}, \nu, \mathbb{C})$ in der L_2-norm beliebig gut durch endliche komplexe Linearkombinationen von Indikatorfunktionen $1_{B_1 \times B_2}$ messbarer Rechtecke $B_1 \times B_2 \subset X \times X$ approximieren kann, gibt es eine solche Linearkombination $G: X \times X \longrightarrow \mathbb{C}$ mit $\iint F(x, y)\overline{G(x, y)}\, d\mu(y)\, d\mu(x) > 0$. Aufgrund des Satzes von Fubini ist die Funktion $x \to \int F(x, y)\overline{G(x, y)}\, d\mu(y)$ eine nichttriviale L_1-Funktion auf X. Wegen der speziellen Form von G folgt daraus, dass es eine messbare Menge $B \subset X$ mit $\mu(B) > 0$ gibt, sodass $G(x_1, y) = G(x_2, y)$ für alle $x_1, x_2 \in B$ und alle $y \in X$ gilt und für die $\int_X F(x, y)\overline{G(x, y)}\, d\mu(y) \neq 0$ für alle x in einer Teilmenge $B' \subset B$ mit $\mu(B') > 0$ ist. Wir wählen ein $x \in B'$, setzen $f(y) = G(x, y)$ für alle $y \in X$ und erhalten $\mathcal{K}(f)(x) \neq 0$ für $x \in B'$. Damit ist $\mathcal{K} \neq 0$ bewiesen.

Da

$$\begin{aligned}\langle\!\langle \mathcal{K} f_1, f_2 \rangle\!\rangle &= \iint F(x, y) f_1(y)\, d\mu(y)\overline{f_2(x)}\, d\mu(x) \\ &= \iint f_1(y)\overline{F(y, x) f_2(x)}\, d\mu(x)\, d\mu(y) = \langle\!\langle f_1, \mathcal{K} f_2 \rangle\!\rangle\end{aligned}$$

für alle $f_1, f_2 \in L_2(X, \mathcal{S}, \mu, \mathbb{C})$ gilt, ist der Operator $\mathcal{K}$ selbstadjungiert. Es folgt nun aus der Spektraltheorie kompakter selbstadjungierter Operatoren (siehe [15, S. 103 ff.]), dass $\mathcal{K}$ einen Eigenwert $\gamma \neq 0$ und einen zugehörigen endlich-dimensionalen Eigenraum $V_\gamma =$

$\ker(\mathcal{K} - \gamma)$ besitzt. Für die konstante Funktion $\mathbf{1} = 1_X$ gilt $\langle\!\langle \mathcal{K}\mathbf{1}, \mathbf{1} \rangle\!\rangle = \int_{X \times X} F \, d\nu = 0$, daher ist $1 \notin V_\gamma$.

Wir behaupten nun, dass $U_T(V_\gamma) \subseteq V_\gamma$ gilt. Sei also $f \in V_\gamma$. Dann ist

$$\begin{aligned}\mathcal{K}(U_T(f))(x) &= \int_X F(x,y) f(Ty) \, d\mu(y) = \int_X F(Tx, Ty) f(Ty) \, d\mu(y) \\ &= \int_X F(Tx, y) f(y) \, d\mu(y) = U_T(\mathcal{K}(f))(x) = \gamma U_T(f)(x)\end{aligned}$$

für μ-*f.a.* $x \in X$ nach der Definition von $\mathcal{K}$ und U_T, der S-Invarianz von F und Lemma 2.3. Daher ist die Einschränkung von U_T auf den Raum V_γ eine unitäre Abbildung auf einem endlich-dimensionalen Vektorraum und besitzt auf diesem Raum zumindest einen Eigenvektor $f \in V_\gamma$. Da die konstante Funktion nicht zu V_γ gehört, haben wir hiermit eine nichtkonstante Eigenfunktion von T gefunden. Dieser Widerspruch zu (4) beendet den Beweis. □

Beispiel 2.25 (Irrationale Rotationen)

Es seien $\alpha \in \mathbb{R}$ eine irrationale reelle Zahl und $T_\alpha : \mathbb{T} \longrightarrow \mathbb{T}$ die in Beispiel 1.2 definierte Rotation, die laut Beispiel 1.61 (1) maßerhaltend auf dem Wahrscheinlichkeitsraum $(\mathbb{T}, \mathcal{B}_\mathbb{T}, \lambda)$ ist. In Beispiel 1.2 sahen wir, dass die in (1.1) definierte Metrik d auf $\mathbb{T}$ invariant unter T_α ist. Für jedes $c > 0$ ist daher die Menge $D_c = \{(x,y) \in \mathbb{T}^2 : \mathsf{d}(x,y) < c\}$ invariant unter der Transformation $S = T_\alpha \times T_\alpha$ auf dem Produktraum $\mathbb{T} \times \mathbb{T}$. Da $(\lambda \otimes \lambda)(D_c) = c$ ist, hat die Transformation S auf $(\mathbb{T} \times \mathbb{T}, \mathcal{B}_{\mathbb{T}^2}, \lambda \otimes \lambda)$ invariante Borelmengen von beliebigem Maß $c \in (0,1]$, sodass S nicht ergodisch und T_α nicht schwach mischend ist.

Man kann auch Eigenfunktionen verwenden, um zu zeigen, dass irrationale Rotationen nicht schwach mischend sind: Für jedes $m \in \mathbb{Z}$ erfüllt die Funktion $f_m(t) = e^{2\pi i m t}$, $t \in \mathbb{T}$, die Gleichung $f_m \circ R_\alpha = e^{2\pi i m \alpha} f_m$. Für $m \neq 0$ ist f_m also eine nichtkonstante Eigenfunktion, womit R_α wegen Satz 2.24 nicht schwach mischend sein kann.

Der Beweis des nächsten Satzes illustriert nochmals den Zusammenhang zwischen schwacher Mischung und Eigenfunktionen.

Satz 2.26 *Es sei T eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Wenn T schwach mischend ist, so ist für jedes irrationale $\alpha \in \mathbb{R}$ die Transformation $T \times R_\alpha$ auf $(X \times \mathbb{T}, \mathcal{S} \otimes \mathcal{B}_\mathbb{T}, \mu \otimes \lambda)$ ergodisch.*

Beweis Wir setzen $Y = X \times \mathbb{T}$, $\mathcal{T} = \mathcal{S} \otimes \mathcal{B}_\mathbb{T}$, $\nu = \mu \otimes \lambda$ und $S = T \times R_\alpha$. Wenn S nicht ergodisch auf $(Y, \mathcal{T}, \nu)$ ist, so gibt es wegen Satz 2.14 eine S-invariante, beschränkte Funktion $g : Y \longrightarrow \mathbb{R}$. Für jedes $x \in X$ entwickeln wir die Funktion $g(x, \cdot) : \mathbb{T} \longrightarrow \mathbb{R}$ als Fourierreihe und erhalten die Darstellung

$$g(x,t) = \sum_{k \in \mathbb{Z}} g_k(x) e^{2\pi i k t}, \quad \text{mit} \quad g_k(x) = \int_0^1 g(x,t) e^{-2\pi i k t} d\lambda(t)$$

für jedes $k \in \mathbb{Z}$. Da $g(x,t) = g \circ S(x,t) = g(Tx, t+\alpha)$ ist, folgt aus der Eindeutigkeit der Fourierentwicklung der Funktionen $g(x,\cdot)$, dass für jedes $k \in \mathbb{Z}$ fast sicher

$$g_k \circ T = g_k \cdot e^{-2\pi i k\alpha}$$

gilt. Damit ist also g_k eine Eigenfunktion von T mit Eigenwert $e^{2\pi i k\alpha}$. Wenn T als schwach mischend vorausgesetzt ist, ist laut Satz 2.24 (4) $g_k = 0$ für $k \neq 0$ und $g_0 = \int g_0 \, d\mu \pmod{\mu}$. Somit ist g konstant $\pmod{\nu}$ und S ist ergodisch. □

Korollar 2.27 Es seien T eine schwach mischende maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ und $f \in \mathcal{L}_1(\mu)$. Dann gilt für jedes irrationale $\alpha \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i k\alpha} f(T^k x) = 0 \quad \text{für} \quad \mu\text{-f.a. } x \in X. \tag{2.6}$$

Beweis Wir definieren $(Y, \mathcal{T}, \nu) = (X \times \mathbb{T}, \mathcal{S} \otimes \mathcal{B}_{\mathbb{T}}, \mu \otimes \lambda)$ und $S = T \times R_\alpha : Y \longrightarrow Y$ wie im Beweis von Satz 2.26 und setzen $g(x,t) = f(x)e^{2\pi i t}$, $(x,t) \in Y$. Da S wegen Satz 2.26 ergodisch ist, folgt aus dem individuellen Ergodensatz 2.10, dass

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k x, t + k\alpha) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i (t+k\alpha)} f(T^k x) \\ &= e^{2\pi i t} \cdot \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i k\alpha} f(T^k x) = 0\end{aligned}$$

$(\mu \otimes \lambda)$–*f.ü.* auf Y ist. Damit ist (2.6) bewiesen. □

Ebenso wie in der topologischen Dynamik spielen auch in der Ergodentheorie Begriffe wie Faktorabbildung und Isomorphie eine wichtige Rolle. Zunächst aber führen wir den Begriff eines *maßerhaltenden dynamischen Systems* ein, in Analogie zu den in Kap. 1 besprochenen topologischen dynamischen Systemen (X, T).

Definition 2.28 (Maßerhaltende Systeme)

Wenn $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T: X \longrightarrow X$ eine maßtreue Transformation ist, so bezeichnet man mit $(X, \mathcal{S}, \mu, T)$ das durch T auf $(X, \mathcal{S}, \mu)$ definierte *maßerhaltende dynamische System*. Das System $(X, \mathcal{S}, \mu, T)$ ist *ergodisch*, schwach mischend oder stark mischend, wenn T ergodisch, schwach mischend oder stark mischend ist.

Die Definition der *Invertierbarkeit* eines maßerhaltenden dynamischen Systems $(X, \mathcal{S}, \mu, T)$ ist etwas allgemeiner als wir es von den topologischen dynamischen Systemen her kennen. Da T messbar ist, gilt natürlich immer $T^{-1}\mathcal{S} \subset \mathcal{S}$. Im Sinne der Maßtheorie nennt man nun das System $(X, \mathcal{S}, \mu, T)$ *invertierbar*, wenn $T^{-1}\mathcal{S} = \mathcal{S} \pmod{\mu}$ gilt.

In Satz 5.13 werden wir diese Definition der Invertierbarkeit anhand von Shifträumen konkretisieren.

Definition 2.29 (Faktoren)

Wenn $(X, \mathcal{S}, \mu, T)$, $(X', \mathcal{S}', \mu', T')$ maßerhaltende dynamische Systeme und $\phi\colon X \longrightarrow X'$ eine messbare Abbildung mit den Eigenschaften $\phi_* \mu = \mu'$ und $\phi \circ T = T' \circ \phi \pmod{\mu}$ ist, so nennt man ϕ eine *messbare Faktorabbildung* und $(X', \mathcal{S}', \mu', T')$ einen (*messbaren*) *Faktor* des Systems $(X, \mathcal{S}, \mu, T)$.[6] Ein Faktor $(X', \mathcal{S}', \mu', T')$ von $(X, \mathcal{S}, \mu, T)$ ist *trivial*, wenn $\mu'(B) \in \{0, 1\}$ für jedes $B \in \mathcal{S}'$ gilt.

Wenn die Faktorabbildung $\phi\colon X \longrightarrow X'$ die Eigenschaft hat, dass $\phi^{-1}(\mathcal{S}') = \mathcal{S} \pmod{\mu}$ ist, so nennt man ϕ einen (*messbaren*) *Isomorphismus* der Systeme $(X, \mathcal{S}, \mu, T)$ und $(X', \mathcal{S}', \mu', T')$, die man dann auch als *isomorph* bezeichnet.[7]

Definition 2.30 (Invariante σ-Algebren)

Wenn $(X, \mathcal{S}, \mu, T)$ ein maßerhaltendes dynamisches System ist, so heißt eine Teil-σ-Algebra $\mathcal{T} \subset \mathcal{S}$ *invariant*, wenn $T^{-1}\mathcal{T} \subset \mathcal{T} \pmod{\mu}$ gilt. Die σ-Algebra $\mathcal{T} \subset \mathcal{S}$ ist *strikt invariant*, wenn $T^{-1}\mathcal{T} = \mathcal{T} \pmod{\mu}$ gilt.

Beispiele 2.31 (Durch invariante σ-Algebren definierte Faktoren)

(1) Es sei $(X, \mathcal{S}, \mu, T)$ ein maßerhaltendes dynamisches System, und $\mathcal{T} \subset \mathcal{S}$ sei eine invariante Teil-σ-Algebra (Definition 2.30). Dann ist $(X, \mathcal{T}, \mu, T)$ ein messbarer Faktor von $(X, \mathcal{S}, \mu, T)$. Als Faktorabbildung können wir hier die Identitätsabbildung auf X nehmen.

(2) Es seien $(X, \mathcal{S}, \mu, T)$ ein maßerhaltendes dynamisches System und $\mathcal{P}$ eine Zerlegung von X (Definition 3.1). Die von $\bigcup_{j=0}^{\infty}(\bigvee_{i=0}^{j} T^{-i}\mathcal{P})$ erzeugte σ-Algebra $\mathcal{T}_{\mathcal{P}}^{-}$ ist invariant und definiert einen messbaren Faktor $(X, \mathcal{T}_{\mathcal{P}}^{-}, \mu, T)$ von $(X, \mathcal{S}, \mu, T)$.

2.3 Anwendungen der Ergodensätze

2.3.1 Eindeutige Ergodizität und Gleichverteilung

Generische Punkte

Sei μ ein T-invariantes Wahrscheinlichkeitsmaß auf der Borel-σ-Algebra $\mathcal{B}_X$ eines topologischen dynamischen Systems (X, T). Wir nennen μ ein *ergodisches Maß*, wenn die Trans-

---

[6] Die Bedingung $\phi \circ T = T' \circ \phi \pmod{\mu}$ wird als *Äquivarianz* $\pmod{\mu}$ (oder einfach als *Äquivarianz*) von ϕ bezeichnet.

[7] Es gibt einen allgemeineren Isomorphiebegriff zwischen maßerhaltenden dynamischen Systemen $(X, \mathcal{S}, \mu, T)$, $(X', \mathcal{S}', \mu', T')$, der nur die Existenz eines äquivarianten Isomorphismus der Maßalgebren $\mathcal{S}_\mu$ und $\mathcal{S}'_{\mu'}$ voraussetzt. Wir werden auf diese Unterschiede nicht weiter eingehen, verweisen aber auf Aufgabe 23.

formation T auf dem Wahrscheinlichkeitsraum $(X, \mathcal{B}_X, \mu)$ ergodisch ist. Ist $f \in C(X)$ und $n \geq 1$, so schreiben wir $M_n f$ für $\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$.

Definition 2.32 (Generische Punkte)

Seien (X, T) ein topologisches dynamisches System und μ ein T-invariantes ergodisches Maß auf $_X$ von X. Ein Punkt $x \in X$ heißt *generisch* für μ, wenn $\lim_{k\to\infty} M_n f(x) = \int f\, d\mu$ für alle $f \in C(X)$ gilt. Wir bezeichnen die Menge aller μ-generischen Punkte mit $G(\mu)$.

Satz 2.33 (Die Menge der generischen Punkte) *Seien (X, T) ein topologisches dynamisches System und μ ein ergodisches T-invariantes Maß auf $\mathcal{B}_X$. Dann gilt $G(\mu) \in \mathcal{B}_X$ und $\mu(G(\mu)) = 1$.*

Beweis Für $g \in C(X)$ sei $G_g(\mu) = \{x \in X : \lim_{n\to\infty} M_n g(x) = \int g\, d\mu\}$. Es gilt $G_g(\mu) \in \mathcal{B}_X$, und aus dem Ergodensatz folgt $\mu(G_g(\mu)) = 1$. Sei $\mathcal{D}$ eine abzählbare dichte Teilmenge von $C(X)$, die wegen Lemma 1.51 existiert, und sei $G_0(\mu) = \bigcap_{h\in\mathcal{D}} G_h(\mu)$. Dann gilt auch $G_0(\mu) \in \mathcal{B}_X$ und $\mu(G_0(\mu)) = 1$.

Seien nun $f \in C(X)$ und $x \in G_0(\mu)$ beliebig. Für jedes $\varepsilon > 0$ existiert ein $h \in \mathcal{D}$ mit $\|f - h\|_\infty < \frac{\varepsilon}{3}$, woraus $\|M_n f - M_n h\|_\infty < \frac{\varepsilon}{3}$ für alle $n \geq 1$ folgt. Wegen $x \in G_0(\mu)$ und $h \in \mathcal{D}$ gilt $x \in G_h(\mu)$ und somit $\lim_{n\to\infty} M_n h(x) = \int h\, d\mu$. Es existiert ein n_0 mit $|M_n h(x) - \int h\, d\mu| < \frac{\varepsilon}{3}$ für alle $n \geq n_0$. Für $n \geq n_0$ gilt dann $|M_n f(x) - \int f\, d\mu| \leq |M_n f(x) - M_n h(x)| + |M_n h(x) - \int h\, d\mu| + |\int h\, d\mu - \int f\, d\mu| < \varepsilon$. Damit ist $\lim_{n\to\infty} M_n f(x) = \int f\, d\mu$ gezeigt, d. h. $x \in G_f(\mu)$. Es gilt also $G_0(\mu) = G(\mu)$, womit der Satz bewiesen ist. □

Aufgaben 2.34

(1) Es sei $N \geq 2$ und μ sei ein shiftinvariantes und ergodisches Wahrscheinlichkeitsmaß auf Σ_N (vgl. (1.11) und (1.13)). Zeigen Sie, dass es eine Folge (ν_n) von shiftinvarianten Wahrscheinlichkeitsmaßen auf Σ_N mit den folgenden Eigenschaften gibt:
   (i) Für jedes $n \geq 1$ ist ν_n auf der Bahn eines einzigen periodischen Punktes konzentriert;
   (ii) $\lim_{n\to\infty} \nu_n = \mu$ in der schwachen*-Topologie.
   Hinweis: Es sei x ein generischer Punkt von μ (Definition 2.32 und Satz 2.33). Dann existiert für jedes $n \geq 1$ ein Punkt $x^{(n)} \in \Sigma_N$ mit $\sigma^n(x^{(n)}) = x^{(n)}$ und $x_k^{(n)} = x_k$ für $k = 0, \ldots, n-1$. Wenn $\mathsf{p}_{x^{(n)}}$ das in $x^{(n)}$ konzentrierte Punktmaß ist und $\nu_n = \frac{1}{n} \sum_{k=0}^{n-1} \mathsf{p}_{x^{(n)}} \sigma^{-k}$ das gleichverteilte Maß auf der Bahn von $x^{(n)}$ ist, so hat die Folge $(\nu_n)_{n\geq 1}$ die gewünschten Eigenschaften.

(2) Es seien $\mu_1, \ldots, \mu_m$ ergodische Maße in $\mathcal{M}(\Sigma_N)^\sigma$ und $\nu = \sum_{k=1}^m r_k \mu_k$ eine rationale Konvexkombination der Maße μ_i, $i = 1, \ldots, m$ (d. h., die r_i sind nichtnegativ und rational und $\sum_{k=1}^m r_i = 1$). Zeigen Sie, dass es eine Folge (ν_n) von shiftinvarianten und ergodischen Wahrscheinlichkeitsmaßen auf Σ_N mit den folgenden Eigenschaften gibt:
   (i) Für jedes $n \geq 1$ ist ν_n auf der Bahn eines einzigen periodischen Punktes konzentriert;
   (ii) $\lim_{n\to\infty} \nu_n = \nu$ in der schwachen*-Topologie.

Wir bemerken folgendes Korollar zu obiger Aufgabe: Der Satz von Krein-Milman (eine Erweiterung von Satz 1.76, siehe [15, Theorem 3.23]) besagt (unter Verwendung von Satz 1.80), dass für jedes topologische dynamische System (X, T) die Menge $\mathcal{M}(X)^T$ der T-invarianten Wahrscheinlichkeitsmaße die kleinste schwach*-abgeschlossene konvexe Teilmenge von $\mathcal{M}(X)$ ist, die alle T-invarianten ergodischen Wahrscheinlichkeitsmaße enthält. Daraus folgt, dass die rationalen Konvexkombinationen endlich vieler ergodischer Maße in $\mathcal{M}(\Sigma_N)^\sigma$ schwach*-dicht in $\mathcal{M}(\Sigma_N)^\sigma$ liegen. Wegen Aufgabe (2) liegen dann insbesondere auch die auf der Bahn eines einzigen Punktes konzentrierten Maße (also spezielle Extremalpunkte von $\mathcal{M}(\Sigma_n)^\sigma$) schwach*-dicht in $\mathcal{M}(\Sigma_N)^\sigma$.

Transformationen auf dem Torus

Mithilfe generischer Punkte untersuchen wir nun die eindeutige Ergodizität von Transformationen auf dem Torus und verwenden diese Resultate für Gleichverteilungsaussagen von Folgen.

Wir beginnen mit einem alternativen Zugang zum Beweis der eindeutigen Ergodizität irrationaler Rotationen auf $\mathbb{T}^d$ in Aufgabe 1.35.

Beispiel 2.35 (Eindeutige Ergodizität irrationaler Rotationen auf $\mathbb{T}^d$ mittels generischer Punkte)

Wie in Aufgabe 1.35 sei $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, sodass $1, \alpha_1, \ldots, \alpha_d$ linear unabhängig über $\mathbb{Q}$ sind. Wir definieren $R_{\boldsymbol{\alpha}}: \mathbb{T}^d \longrightarrow \mathbb{T}^d$ durch $R_{\boldsymbol{\alpha}}\mathbf{x} = \mathbf{x} + \boldsymbol{\alpha} \pmod 1$ für jedes $\mathbf{x} \in \mathbb{T}^d$. In Beispiel 2.16 haben wir gesehen, dass $R_{\boldsymbol{\alpha}}$ auf dem Wahrscheinlichkeitsraum $(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d)$ ergodisch ist. Nach Satz 2.33 gibt es dann mindestens einen generischen Punkt $\mathbf{u} \in \mathbb{T}^d$. Das bedeutet, dass $\lim_{n\to\infty} M_n f(\mathbf{u}) = \int f \, d\lambda^d$ für alle $f \in C(\mathbb{T}^d)$ gilt.

Es seien $\mathbf{v} \in \mathbb{T}^d$ und $f \in C(\mathbb{T}^d)$ beliebig. Die Abbildung $S: \mathbb{T}^d \longrightarrow \mathbb{T}^d$, definiert durch $S(\mathbf{x}) = \mathbf{x} + \mathbf{v} - \mathbf{u} \pmod 1$, ist stetig, sodass $g = f \circ S$ in $C(\mathbb{T}^d)$ liegt. Es folgt $\lim_{n\to\infty} M_n g(\mathbf{u}) = \int g \, d\lambda^d$. Nach Beispiel 1.62 ist λ^d invariant unter S, sodass $\int g \, d\lambda^d = \int f \, d\lambda^d$ wegen Korollar 2.3 gilt. Da auch $S(T^j(\mathbf{u})) = T^j(\mathbf{v})$ für alle $j \geq 1$ gilt, ist $\lim_{n\to\infty} M_n f(\mathbf{v}) = \int f \, d\lambda^d$ gezeigt. Wegen Satz 1.74 ist λ^d das einzige $R_{\boldsymbol{\alpha}}$-invariante Wahrscheinlichkeitsmaß auf der Borel-σ-Algebra $\mathcal{B}_{\mathbb{T}^d}$ von $\mathbb{T}^d$. Das topologische dynamische System $(\mathbb{T}^d, R_{\boldsymbol{\alpha}})$ ist somit eindeutig ergodisch.

Nach Weyl[8] [19] heißt eine Folge $(x_n)_{n\geq 0}$ in $\mathbb{T} = [0, 1)$ *gleichverteilt*, wenn $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k) = \int f \, d\lambda$ für jede Funktion $f \in C(\mathbb{T})$ gilt. Allgemeiner nennt man eine Folge $(\mathbf{x}_n)_{n\geq 0}$ in $\mathbb{T}^d$ *gleichverteilt*, wenn $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\mathbf{x}_k) = \int f \, d\lambda^d$ für jede Funktion $f \in C(\mathbb{T}^d)$ gilt. In Beispiel 2.35 wurde gezeigt, dass jede Bahn des topologischen dynamischen Systems $(\mathbb{T}^d, R_{\boldsymbol{\alpha}})$, d. h. jede Folge $(\mathbf{x} + n\boldsymbol{\alpha} \pmod 1)_{n\geq 0}$ mit $\mathbf{x} \in \mathbb{T}^d$, gleichverteilt ist, wenn $\boldsymbol{\alpha}$ die Eigenschaft hat, dass $1, \alpha_1, \ldots, \alpha_d$ linear unabhängig über $\mathbb{Q}$ sind. Als Spezialfall erhält man das klassische Resultat, dass die Folge $(n\alpha \pmod 1)_{n\geq 0}$ in $\mathbb{T} = [0, 1)$ gleichverteilt ist, wenn α irrational ist.

[8] Hermann Weyl (1885–1955) war ein deutscher Mathematiker und mathematischer Physiker, der lange Jahre in Zürich und Princeton arbeitete. Er leistete wichtige Beiträge zu vielen Gebieten der Mathematik und Physik, darunter Analysis, Algebra, Zahlentheorie, Topologie, Differentialgeometrie, Relativitätstheorie, Quantenmechanik und Grundlagen der Mechanik.

Um weitere Beispiele für eindeutig ergodische Systeme und gleichverteilte Folgen zu finden, beweisen wir folgenden Satz von Furstenberg.[9]

Satz 2.36 *Sei (X,T) ein invertierbares topologisches dynamisches System, das eindeutig ergodisch ist mit T-invariantem Wahrscheinlichkeitsmaß μ. Sei λ das Lebesguemaß auf $\mathbb{T}$ und $c\colon X \longrightarrow \mathbb{T}$ eine stetige Abbildung. Wir versehen den Produktraum $Y = X \times \mathbb{T}$ mit dem Produktmaß $\nu = \mu \otimes \lambda$ und definieren die Transformation $S\colon Y \longrightarrow Y$ durch $S(x,t) = (T(x), t + c(x) \pmod 1)$. Dann ist ν invariant unter S. Ist ν außerdem ergodisch, so ist (Y,S) eindeutig ergodisch.*

Beweis Nach dem Satz von Fubini gilt

$$\nu(S^{-1}(A \times B)) = \int 1_A(T(x)) \int 1_B(t + c(x) \pmod 1)\, d\lambda(t)\, d\mu(x)$$

für $A \in \mathcal{B}_X$ und $B \in \mathcal{B}_{\mathbb{T}}$. Da λ für jedes $\alpha \in \mathbb{R}$ invariant unter der Translation $t \mapsto t + \alpha \pmod 1$ und μ invariant unter T ist, erhalten wir $\nu(S^{-1}(A \times B)) = \mu(A)\lambda(B) = \nu(A \times B)$. Nun ist $\{A \times B : A \in \mathcal{B}_X, B \in \mathcal{B}_{\mathbb{T}}\}$ ein Semiring, der die Produkt-σ-Algebra $\mathcal{B}_X \otimes \mathcal{B}_{\mathbb{T}}$ erzeugt. Aus Lemma 1.58 folgt, dass ν invariant unter S ist.

Wir nehmen jetzt an, dass ν ein ergodisches Maß unter der Transformation S ist, und bezeichnen mit $G(\nu) \subset Y$ die Menge der ν-generischen Punkte.

Für $s \in \mathbb{T}$ sei $V_s\colon Y \longrightarrow Y$ definiert durch $V_s(x,t) = (x, t + s \pmod 1)$. Das Maß ν ist dann invariant unter V_s, da V_s ein Spezialfall der Transformation S ist.

Es sei nun $(x,t) \in G(\nu)$. Für alle $f \in C(Y)$ gilt $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^j(x,t)) = \int f\, d\nu$ und wegen $f \circ V_s \in C(Y)$ auch $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ V_s(S^j(x,t)) = \int f \circ V_s\, d\nu$. Da $V_s \circ S = S \circ V_s$ gilt und wegen Korollar 2.3 auch $\int f \circ V_s\, d\nu = \int f\, d\nu$ ist, erhalten wir $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^j \circ V_s(x,t)) = \int f\, d\nu$, womit $V_s(x,t) \in G(\nu)$ gezeigt ist. Wir haben bewiesen, dass $V_s(G(\nu)) \subset G(\nu)$ für jedes $s \in \mathbb{T}$ gilt. Daraus folgt aber, dass $G(\nu)$ von der Form $G(\nu) = C \times \mathbb{T}$ für eine Menge $C \in \mathcal{B}_X$ ist. Wegen Satz 2.33 gilt $1 = \nu(G(\nu)) = (\mu \otimes \lambda)(C \times \mathbb{T}) = \mu(C)$.

Es sei nun ρ ein beliebiges S-invariantes und ergodisches Wahrscheinlichkeitsmaß auf Y. Sei $\pi\colon Y \longrightarrow X$ die Projektion auf die erste Koordinate und ρ' das Bildmaß $\pi_*\rho$. Dann ist ρ' invariant unter T, da ja $T \circ \pi = \pi \circ S$ gilt. Da (X,T) eindeutig ergodisch ist, muss $\rho' = \mu$ gelten. Es folgt $\rho(G(\nu)) = \rho(C \times \mathbb{T}) = \rho'(C) = \mu(C) = 1$. Wegen Satz 2.33 über die Menge der generischen Punkte gilt außerdem $\rho(G(\rho)) = 1$. Also gibt es einen Punkt (x,t), der sowohl für ν als auch für ρ generisch ist. Aus Lemma 1.51 folgt nun $\rho = \nu$, und Korollar 1.82 zeigt, dass S eindeutig ergodisch ist. □

[9] Hillel Furstenberg, geboren 1935 in Berlin, Professor in Jerusalem. Viele der bedeutendsten Themen der Ergodentheorie und ihrer Verbindungen zur Wahrscheinlichkeitstheorie, Kombinatorik und Zahlentheorie entwickelten sich aus seinen Ideen.

Beispiel 2.37
Seien $M(d)$ die $d \times d$-Matrix mit Koeffizienten $M(d)_{ij} = 1$ für $j \le i \le j+1$ und $M(d)_{ij} = 0$ sonst. Seien $\boldsymbol{\alpha} \in \mathbb{R}^d$ und $T = T_{d,\boldsymbol{\alpha}} : \mathbb{T}^d \longrightarrow \mathbb{T}^d$ definiert durch

$$T(\mathbf{x}) = M(d)\mathbf{x} + \boldsymbol{\alpha} = (x_1 + \alpha_1, x_2 + x_1 + \alpha_2, \dots, x_d + x_{d-1} + \alpha_d) \pmod 1.$$

Da $M(d)$ Koeffizienten aus $\mathbb{Z}$ und Determinante 1 hat, ist $(\mathbb{T}^d, T)$ nach Beispiel 1.62 ein invertierbares topologisches dynamisches System, und das Lebesguemaß λ^d ist invariant unter T. Wir zeigen mit Hilfe von Satz 2.15, dass λ^d ein ergodisches Maß ist, wenn der erste Koeffizient α_1 von $\boldsymbol{\alpha}$ irrational ist.

Sei $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$. Sei j maximal, sodass $n_j \neq 0$ ist. Ist $j = 1$, dann ist $\mathbf{n} = (n_1, 0, \dots, 0)$ und daher $\mathbf{n}M(d) = \mathbf{n}$ und $\mathbf{n}\boldsymbol{\alpha} = n_1\alpha_1 \notin \mathbb{Q}$. Ist $j \ge 2$, dann hat $\mathbf{n}M(d)^k$ für $k \ge 1$ an der Stelle $j-1$ den Koeffizienten $n_{j-1} + kn_j$ und an der Stelle j den Koeffizienten n_j, wie man leicht mit Induktion beweist. Daher ist die Menge $\{\mathbf{n}M(d)^k : k \in \mathbb{Z}\}$ unendlich. Nach Satz 2.15 ist T eine ergodische Transformation auf $(\mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d}, \lambda^d)$.

Um die eindeutige Ergodizität von T zu beweisen, verwenden wir die Induktion nach d. Wir wissen bereits, dass $T = T_{d,\boldsymbol{\alpha}}$ für jedes $d \ge 1$ und jedes $\boldsymbol{\alpha} \in \mathbb{R}^d$ mit $\alpha_1 \in \mathbb{R} \setminus \mathbb{Q}$ ergodisch ist.

Für $d = 1$ gilt $T(x) = x + \alpha_1 \pmod 1$. Da wir $\alpha_1 \notin \mathbb{Q}$ voraussetzen, wurde die eindeutige Ergodizität in Beispiel 2.35 gezeigt. Wir nehmen nun an, dass die eindeutige Ergodizität für Dimension d bereits gezeigt ist. Wir wählen ein Element $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{d+1}) \in \mathbb{R}^{d+1}$ mit $\alpha_1 \in \mathbb{R} \setminus \mathbb{Q}$, definieren $T = T_{d+1,\boldsymbol{\alpha}} : \mathbb{T}^{d+1} \longrightarrow \mathbb{T}^{d+1}$ wie oben und betrachten $T' = T_{d,\boldsymbol{\alpha}'} : \mathbb{T}^d \longrightarrow \mathbb{T}^d$ mit $\boldsymbol{\alpha}' = (\alpha_1, \dots, \alpha_d)$. Dann ist $(\mathbb{T}^d, T')$ laut Induktionsvoraussetzung eindeutig ergodisch. Die Transformation $T : \mathbb{T}^d \times \mathbb{T} \longrightarrow \mathbb{T}^d \times \mathbb{T}$ ist von der Form

$$T(\mathbf{y}, t) = (T'(\mathbf{y}), t + y_d + \alpha_{d+1})$$

für $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{T}^d$ und $t \in \mathbb{T}$ und ist ergodisch, wie wir oben bewiesen haben. Wir wenden Satz 2.36 auf das topologische dynamische System $(\mathbb{T}^d, T')$ mit $\mu = \lambda^d$ und $c(\mathbf{y}) = y_d + \alpha_{d+1}$ an und erhalten, dass $(\mathbb{T}^{d+1}, T)$ eindeutig ergodisch ist mit invariantem Maß $\lambda^d \otimes \lambda = \lambda^{d+1}$.

Damit ist die eindeutige Ergodizität von $T = T_{d,\boldsymbol{\alpha}}$ für jedes $d \ge 1$ und jedes $\boldsymbol{\alpha} = (\alpha_1, \dots \alpha_d) \in \mathbb{R}^d$ mit $\alpha_1 \in \mathbb{R} \setminus \mathbb{Q}$ gezeigt.

Wir können Beispiel 2.37 verwenden, um die Gleichverteilung von Polynomen modulo 1 zu zeigen. Dazu beweisen wir ein Lemma.

Lemma 2.38 Für $1 \le k \le d$ sei $\pi_k : \mathbb{T}^d \longrightarrow \mathbb{T}$ die Projektion auf die k-te Koordinate. Ist $(\mathbf{u}_n)_{n\ge0}$ eine gleichverteilte Folge in $\mathbb{T}^d$, dann ist $(\pi_k(\mathbf{u}_n))_{n\ge0}$ eine gleichverteilte Folge in $\mathbb{T}$.

Beweis Sei $f \in C(\mathbb{T})$ beliebig. Da π_k stetig ist, liegt $f \circ \pi_k$ in $C(\mathbb{T}^d)$. Nach Voraussetzung gilt $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ \pi_k(\mathbf{u}_j) = \int f \circ \pi_k \, d\lambda^d = \int f \, d\lambda$. Das bedeutet, dass die Folge $(\pi_k(\mathbf{u}_n))_{n\ge1}$ gleichverteilt in $\mathbb{T}$ ist. □

Die Matrix $M(d)$ und der Vektor $\boldsymbol{\alpha}$ seien wie in Beispiel 2.37, wobei jetzt alle Koeffizienten des Vektors $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ null seien außer α_1, das irrational sei. Mittels Induktion zeigt man, dass $M(d)^n$ für $n \geq 1$ an der Stelle (i, j) den Koeffizienten $\binom{n}{i-j}$ hat, wenn $i \geq j$ gilt, und sonst den Koeffizienten 0 (wobei $\binom{u}{v}$ gleich 0 zu setzen ist, wenn $u < v$ gilt). Sei $T = T_{d,\boldsymbol{\alpha}}$ die Transformation aus Beispiel 2.37 und $\mathbf{x} \in \mathbb{T}^d$. Durch Induktion zeigt man $T^n(\mathbf{x}) = M(d)^n\mathbf{x} + \sum_{m=0}^{n-1} M(d)^m\boldsymbol{\alpha} \pmod 1$ und $\sum_{m=0}^{n-1}\binom{m}{d-1} = \binom{n}{d}$ für alle $n \geq 1$. Daraus ergibt sich $\pi_d(T^n(\mathbf{x})) = \sum_{j=1}^{d}\binom{n}{d-j}x_j + \binom{n}{d}\alpha_1 \pmod 1$. Ist jetzt $p(t) = \sum_{j=0}^{d} c_j t^j$ ein Polynom vom Grad d mit $c_d \notin \mathbb{Q}$, dann können wir beginnend mit $\alpha_1 = d!c_d \notin \mathbb{Q}$ der Reihe nach $x_1, x_2, x_3, \ldots, x_d \in \mathbb{R}$ so bestimmen, dass $\sum_{j=1}^{d}\binom{n}{d-j}x_j + \binom{n}{d}\alpha_1 = p(n)$ für $n \geq 0$ erfüllt ist.

Zu jedem Polynom p vom Grad d mit höchstem Koeffizienten $c_d \notin \mathbb{Q}$ gibt es somit ein $\alpha_1 \notin \mathbb{Q}$ und ein $\mathbf{x} \in \mathbb{T}^d$, sodass $p(n) \pmod 1 = \pi_d(T^n(\mathbf{x}))$ für alle $n \geq 0$ gilt. Da das topologische dynamische System $(\mathbb{T}^d, T)$ nach Beispiel 2.37 eindeutig ergodisch ist mit invariantem Wahrscheinlichkeitsmaß λ^d, folgt aus Satz 1.74 und der Definition der Gleichverteilung, dass jede Bahn $(T^n(\mathbf{x}))_{n\geq 0}$ mit $\mathbf{x} \in \mathbb{T}^d$ gleichverteilt ist. Mit Hilfe von Lemma 2.38 erhalten wir dann folgendes Resultat von Hermann Weyl aus dem Jahr 1916 [19].

Satz 2.39 (Gleichverteilung von Polynomen) *Sei $p(t) = \sum_{j=0}^{d} c_j t^j$ ein Polynom vom Grad $d \geq 1$ mit $c_d \notin \mathbb{Q}$. Dann ist die Folge $(p(n) \pmod 1)_{n\geq 0}$ gleichverteilt in $\mathbb{T}$.*

2.3.2 Ziffernentwicklungen

Wir haben die Ziffernentwicklung einer Zahl zur Basis 2 mit Hilfe der Transformation $T(x) = 2x \pmod 1$ auf dem Intervall $\mathbb{T} = [0,1)$ in Beispiel 1.3 beschrieben. Seien $Z_0 = [0, \frac{1}{2})$ und $Z_1 = [\frac{1}{2}, 1)$. Ist $x \in \mathbb{T}$ und wird $i_k \in \{0,1\}$ für $k \geq 0$ so gewählt, dass $T^k(x) \in Z_{i_k}$ gilt, dann ist $i_0 i_1 i_2 \ldots$ die Ziffernentwicklung von x zur Basis 2. Nach Beispiel 1.61 ist T eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(\mathbb{T}, \mathcal{B}_{\mathbb{T}}, \lambda)$. In Beispiel 2.21 wurde gezeigt, dass T stark mischend und somit auch ergodisch ist. Ist $f = 1_{Z_1}$, dann gibt das Zeitmittel $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x))$ die Frequenz an, mit der die Ziffer 1 in der dyadischen Entwicklung von x vorkommt. Wegen $\int f\, d\lambda = \frac{1}{2}$ besagt der Ergodensatz, dass für λ-fast alle Zahlen in $[0,1)$ die Frequenz der Ziffer 1 gleich $\frac{1}{2}$ ist. Im Folgenden werden wir derartige Fragestellungen für β-Entwicklungen und Kettenbrüche behandeln.

Beta-Entwicklung

Wir untersuchen jetzt Ziffernentwicklungen zu einer Basis $\beta > 1$, die nicht ganzzahlig sein muss. Wir setzen $I = [0,1]$ und definieren die β-Transformation $T_\beta\colon I \longrightarrow I$ folgendermaßen:

$$T_\beta(x) = \begin{cases} \beta x \pmod 1 & \text{für } 0 \leq x < 1, \\ \lim_{t \nearrow 1} T_\beta(t) & \text{für } x = 1. \end{cases} \tag{2.7}$$

Für $\beta \notin \mathbb{N}$ erhält man $T_\beta(x) = \beta x \pmod 1$ für alle $x \in [0,1]$. Für $\beta \in \mathbb{N}$ ist jedoch $T_\beta(1) = 1 \neq \beta \pmod 1$.

Die Transformation T_β ist nicht stetig. Für $\beta \notin \mathbb{N}$ ist sie auch dann nicht stetig, wenn man das Intervall I durch Identifikation seiner Endpunkte als Torus auffasst.

Wenn $b = \lceil \beta - 1 \rceil$ die kleinste ganze Zahl $\geq \beta - 1$ ist, so setzt man $Z_m = [\frac{m}{\beta}, \frac{m+1}{\beta})$ für $0 \leq m < b$ und $Z_b = [\frac{b}{\beta}, 1]$. Dann gilt $T_\beta(x) = \beta x - m$ für $x \in Z_m$.

Sei $x \in I$. Wir verfolgen die Bahn $\{T_\beta^k(x) : k \geq 0\}$ des Punktes x unter T_β und bezeichnen für jedes $k \geq 0$ mit i_k das eindeutige Element von $\{0, 1, \ldots, b\}$, für das $T_\beta^k(x) \in Z_{i_k}$ gilt. Wegen $x \in Z_{i_0}$ haben wir $T_\beta(x) = \beta x - i_0$, d. h. $x = \frac{i_0}{\beta} + \frac{T_\beta(x)}{\beta}$. Wegen $T_\beta(x) \in Z_{i_1}$ haben wir $T_\beta^2(x) = \beta T_\beta(x) - i_1$, woraus $x = \frac{i_0}{\beta} + \frac{i_1}{\beta^2} + \frac{T_\beta^2(x)}{\beta^2}$ folgt. Fährt man so fort, so erhält man $x = \sum_{k=0}^{n-1} \frac{i_k}{\beta^{k+1}} + \frac{T_\beta^n(x)}{\beta^n}$ für $n \geq 1$. Lässt man n gegen ∞ gehen, so ergibt sich $x = \sum_{k=0}^{\infty} \frac{i_k}{\beta^{k+1}}$. Wir nennen dann $\eta(x) = i_0 i_1 i_2 \ldots$ die *β-Entwicklung* der Zahl x.

Die β-Entwicklung der Zahl 1 spielt bei der Untersuchung der β-Transformation eine besondere Rolle: Wir bezeichnen sie mit $e_0 e_1 e_2 \ldots$ Es gilt natürlich wieder $1 = \sum_{k=0}^{\infty} \frac{e_k}{\beta^{k+1}}$.

Sei $\mathcal{B}_I$ die Borel-σ-Algebra auf dem Intervall I und λ das Lebesguemaß auf $\mathcal{B}_I$. Weiterhin sei

$$g = c \sum_{k=0}^{\infty} \beta^{-k} 1_{[0, T_\beta^k(1))}, \tag{2.8}$$

wobei $c > 0$ so gewählt wird, dass $\int g \, d\lambda = 1$ gilt. Wir bezeichnen mit μ_β das Wahrscheinlichkeitsmaß auf $[0, 1]$, das Dichte g bezüglich λ hat.

Lemma 2.40 Das Wahrscheinlichkeitsmaß μ_β ist invariant unter der β-Transformation T_β.

Beweis Sei $J = [0, y)$ mit $0 < y \leq 1$. Wegen $T_\beta^{-1}(J) = \bigcup_{m=0}^{b-1} [\frac{m}{\beta}, \frac{m+y}{\beta}) \cup [\frac{b}{\beta}, \min(\frac{b+y}{\beta}, 1))$ und $\frac{e_k}{\beta} \leq T_\beta^k(1) \leq \frac{e_k+1}{\beta}$ gilt

$$T_\beta^{-1}(J) \cap [0, T_\beta^k(1)) = \bigcup_{m=0}^{e_k-1} \left[\frac{m}{\beta}, \frac{m+y}{\beta}\right) \cup \left[\frac{e_k}{\beta}, \frac{e_k}{\beta} + \min\left(\frac{y}{\beta}, T_\beta^k(1) - \frac{e_k}{\beta}\right)\right).$$

Nun ist aber $\min(\frac{y}{\beta}, T_\beta^k(1) - \frac{e_k}{\beta}) = \frac{1}{\beta}\min(y, \beta T_\beta^k(1) - e_k) = \frac{1}{\beta}\min(y, T_\beta^{k+1}(1))$ und daher

$$\begin{aligned}
\mu_\beta(T_\beta^{-1}(J)) &= c \sum_{k=0}^{\infty} \beta^{-k} \left(e_k \frac{y}{\beta} + \frac{1}{\beta} \min(y, T_\beta^{k+1}(1))\right) \\
&= yc \sum_{k=0}^{\infty} \frac{e_k}{\beta^{k+1}} + c \sum_{k=0}^{\infty} \beta^{-k-1} \min(y, T_\beta^{k+1}(1)) \\
&= yc + c \sum_{k=1}^{\infty} \beta^{-k} \min(y, T_\beta^k(1)) = c \sum_{k=0}^{\infty} \beta^{-k} \min(y, T_\beta^k(1)) = \mu_\beta(J).
\end{aligned}$$

Die Menge $\mathcal{E} = \{[0, y) : 0 < y \leq 1\} \cup I$ erzeugt die σ-Algebra $\mathcal{B}_I$, und die Menge der Differenzen $\mathcal{E}' = \{A \setminus B : A, B \in \mathcal{E}\}$ ist ein Semiring, der natürlich ebenfalls $\mathcal{B}_I$ erzeugt.

Da die Maße μ_β und $(T_\beta)_*\mu_\beta$ auf $\mathcal{E}$, und daher auch auf $\mathcal{E}'$, übereinstimmen, folgt aus Lemma 1.58, dass μ_β invariant unter T_β ist. □

Wir zeigen, dass T_β eine ergodische Transformation auf $(I, \mathcal{B}_I, \mu_\beta)$ ist. Dazu definieren wir die Zylindermengen $Z_{i_0 i_1 \dots i_{n-1}} = \bigcap_{k=0}^{n-1} T_\beta^{-k}(Z_{i_k})$ für $n \geq 1$ und $i_0, i_1, \dots, i_{n-1} \in \{0, 1, \dots, b\}$. Wir nennen n den Rang der Zylindermenge. Eine Zylindermenge vom Rang n heißt *voll*, wenn ihr Bild unter T_β^n die Menge $[0,1)$ enthält.

Lemma 2.41 Eine nichtleere Zylindermenge vom Rang n ist ein Intervall $[u, v)$ mit $0 \leq u < v \leq 1$, sodass $T_\beta^n(x) = \beta^n(x-u)$ für $x \in [u, v)$ gilt. Für jedes $x \in [0,1)$ und jedes $\varepsilon > 0$ gibt es eine volle Zylindermenge C mit $x \in C$ und $\lambda(C) < \varepsilon$.

Beweis Die erste Aussage zeigen wir durch Induktion. Für $n = 1$ folgt sie direkt aus der Definition der β-Transformation. Wir nehmen an, dass sie für Zylindermengen vom Rang n bereits bewiesen ist. Ist C eine Zylindermenge vom Rang $n+1$, dann gilt $C = D \cap T_\beta^{-n}(Z_m)$, wobei D eine Zylindermenge vom Rang n ist und $0 \leq m \leq b$ gilt. Nach Annahme ist D ein Intervall $[u, v)$ mit $T_\beta^n(x) = \beta^n(x-u)$ für $x \in [u, v)$. Für $j \geq 0$ sei $u_j = u + j\beta^{-(n+1)}$. Es folgt $C = [u, v) \cap [u_m, u_{m+1})$. Das ist ein halboffenes Intervall mit linkem Endpunkt u_m. Auf $[u, v)$ ist T_β^n eine lineare Abbildung mit Anstieg β^n. Auf $T_\beta^n(C) \subset Z_m$ ist T_β eine lineare Abbildung mit Anstieg β. Für alle $x \in C$ gilt somit $T_\beta^{n+1}(x) = \beta^{n+1}(x - u_m)$, da ja auch $T_\beta^{n+1}(u_m) = T_\beta(T_\beta^n(u_m)) = T_\beta(\beta^n m \beta^{-(n+1)}) = T_\beta(\frac{m}{\beta}) = 0$ erfüllt ist. Damit ist die erste Aussage durch Induktion bewiesen.

Das zeigt auch, dass eine Zylindermenge $D = [u, v)$ vom Rang n eine disjunkte Vereinigung von k Zylindermengen $[u_0, u_1)$, $[u_1, u_2)$,…,$[u_{k-1}, v)$ vom Rang $n+1$ ist, wobei $1 \leq k \leq b$ gilt. Außerdem sind die Zylindermengen $[u_m, u_{m+1})$ für $0 \leq m \leq k-2$ voll.

Sei $x \in [0,1)$ und $\varepsilon > 0$. Wir wählen r so, dass $\beta^{-r} < \varepsilon$ gilt. Indem wir mit der Zylindermenge vom Rang 1, die x enthält, beginnen und diese immer weiter zerlegen, finden wir eine Zylindermenge $E = [a, b)$ vom Rang r, die x enthält. Ist sie voll, dann sind wir fertig, da $b - a = \beta^{-r} < \varepsilon$ gilt. Ist sie nicht voll, dann zerlegen wir sie in Zylindermengen vom Rang $r+1$. Entweder liegt x in einer vollen Zylindermenge vom Rang $r+1$, dann sind wir fertig, da diese Durchmesser $\beta^{-(r+1)} < \varepsilon$ hat, oder x liegt in der Zylindermenge $[a_1, b)$, für die $b - a_1 < \beta^{-(r+1)}$ gilt. In diesem Fall wenden wir dieselbe Vorgangsweise auf $[a_1, b)$ an. Auf diese Weise finden wir $a_1 \leq a_2 \leq a_3 \leq \dots < b$ mit $b - a_k < \beta^{-(r+k)}$. Wegen $x < b$ existiert ein j mit $a_{j-1} \leq x < a_j$. Dann liegt x in einer vollen Zylindermenge C vom Rang $r + j$ und es gilt $\lambda(C) = \beta^{-(r+j)} < \varepsilon$. □

Das folgende Lemma von Knopp[10] spielt eine wesentliche Rolle im Beweis der Ergodizität.

[10] Konrad Hermann Theodor Knopp (1882–1957) war ein deutscher Mathematiker, der auf dem Gebiet der Analysis, v. a. über Summierungs- und Limitierungsverfahren, arbeitete.

Lemma 2.42 Sei μ ein Wahrscheinlichkeitsmaß auf der Borel-σ-Algebra $\mathcal{B}_I$ des Intervalls I. Es seien weiterhin $B \in \mathcal{B}_I$ und $\mathcal{H}$ eine abzählbare Menge von Intervallen mit folgenden Eigenschaften:

(1) Jedes offene Intervall $U \subset I$ hat eine Teilmenge V, die eine disjunkte Vereinigung von Mengen aus $\mathcal{H}$ ist und $\mu(U \setminus V) = 0$ erfüllt.
(2) Es existiert ein $c > 0$ mit $\mu(B \cap H) \geq c\mu(H)$ für alle $H \in \mathcal{H}$.

Dann gilt $\mu(B) = 1$.

Beweis Wir nehmen an, dass $\mu(B^c) > 0$ gilt, und wählen ein $\varepsilon > 0$, das kleiner als $c\mu(B^c)$ ist. Es gibt eine offene Menge E mit $B^c \subset E$ und $\mu(E \setminus B^c) < \varepsilon$. Nun ist E eine abzählbare Vereinigung von offenen paarweise disjunkten Intervallen, auf die wir die Bedingung (1) anwenden: Wir finden eine Teilmenge F von E, die disjunkte Vereinigung von Mengen aus $\mathcal{H}$ und $\mu(E \setminus F) = 0$ erfüllt. Es folgt $\mu(F \setminus B^c) < \varepsilon$ und $\mu(B^c \setminus F) = 0$. Da F eine disjunkte Vereinigung von Mengen aus $\mathcal{H}$ und $\mathcal{H}$ abzählbar ist, folgt $\mu(B \cap F) \geq c\mu(F)$ aus (2). Nun gilt $\varepsilon > \mu(F \setminus B^c) = \mu(F \cap B) \geq c\mu(F)$ und wegen $\mu(B^c \setminus F) = 0$ auch $\mu(B^c) \leq \mu(F)$. Damit ist $\varepsilon > c\mu(B^c)$ gezeigt, ein Widerspruch zur Wahl von ε. □

Satz 2.43 (Ergodizität der β-Transformation) *Sei $\beta > 1$. Dann ist die β-Transformation T_β auf dem Wahrscheinlichkeitsraum $(I, \mathcal{B}_I, \mu_\beta)$ ergodisch.*

Beweis Sei $B \in \mathcal{B}_I$ mit $T_\beta^{-1}(B) = B$ und $\mu_\beta(B) > 0$. Wir müssen $\mu_\beta(B) = 1$ zeigen. Dazu verwenden wir Lemma 2.42. Sei $\mathcal{H}$ die Menge aller vollen Zylindermengen. Wir zeigen, dass (1) und (2) aus Lemma 2.42 gelten.

Sei $U \subset I$ ein offenes Intervall. Nach Lemma 2.41 existiert für jedes $x \in U$ eine volle Zylindermenge C_x mit $x \in C_x \subset U$. Sei $\tilde{\mathcal{U}} = \{C_x : x \in U\}$. Dann ist $\tilde{\mathcal{U}}$ eine Überdeckung von U. Da zwei Zylindermengen entweder disjunkt sind oder die eine in der anderen enthalten ist, können wir für jedes $D \subset \tilde{\mathcal{U}}$ alle $C \in \tilde{\mathcal{U}}$ mit $C \subsetneq D$ herausnehmen und erhalten eine Teilmenge $\mathcal{U}$ von $\tilde{\mathcal{U}}$, die aus paarweise disjunkten in U enthaltenen Zylindermengen besteht und U immer noch überdeckt. Somit ist (1) mit $V = U$ erfüllt.

Sei $C \in \mathcal{H}$ eine volle Zylindermenge vom Rang n und $\mathcal{E}$ die Menge aller Intervalle in I. Aus Lemma 2.41 folgt $\lambda(T_\beta^{-n}(I) \cap C) = \beta^{-n}\lambda(I) = \lambda(C)\lambda(I)$ für alle $I \in \mathcal{E}$. Sei $\nu(A) = \frac{\lambda(T_\beta^{-n}(A) \cap C)}{\lambda(C)}$ für $A \in \mathcal{B}_I$. Dann ist ν ein Maß auf $\mathcal{B}_I$. Da $\mathcal{E}$ ein Semiring ist, der $\mathcal{B}_I$ erzeugt, und $\nu(I) = \lambda(I)$ für alle $I \in \mathcal{E}$ gilt, erhalten wir $\nu = \lambda$ aus dem Eindeutigkeitssatz für Maße in [3, Kapitel VII]. Damit ist $\lambda(T_\beta^{-n}(A) \cap C) = \lambda(A)\lambda(C)$ für alle $A \in \mathcal{B}_I$ gezeigt. Es folgt $\mu_\beta(T_\beta^{-n}(A) \cap C) \geq c\lambda(T_\beta^{-n}(A) \cap C) = c\lambda(A)\lambda(C) \geq \frac{(\beta-1)^2}{c\beta^2}\mu_\beta(A)\mu_\beta(C)$ für alle $A \in \mathcal{B}_I$, da $c \leq g \leq \frac{c\beta}{\beta-1}$ für die Dichte g von μ_β gilt. Setzt man $A = B$ und beachtet, dass $T_\beta^{-n}(B) = B$

gilt, dann erhält man $\mu_\beta(B \cap C) \geq d\mu_\beta(C)$ mit $d = \frac{(\beta-1)^2}{c\beta^2}\mu_\beta(B)$. Das gilt für alle $C \in \mathcal{H}$. Damit ist (2) gezeigt.

Aus Lemma 2.42 folgt jetzt, dass $\mu_\beta(B) = 1$ gilt. Damit ist gezeigt, dass T_β eine ergodische Transformation ist. □

Wir können den Ergodensatz auf die β-Transformation anwenden. Wählt man $f = 1_{Z_m}$, dann ist $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T_\beta^j(x))$ die Frequenz, mit der die Ziffer m in der β-Entwicklung der Zahl x auftritt. Da die β-Transformation ergodisch ist, folgt aus dem Ergodensatz, dass für μ_β-fast alle x die Ziffer m mit Frequenz $\mu_\beta(Z_m)$ in der β-Entwicklung der Zahl x auftritt. Da μ_β eine Dichte hat, die auf ganz I größer als 0 ist, gilt das auch für λ-fast alle x.

Ist β eine ganze Zahl $N \geq 2$, dann ist μ_β das Lebesguemaß λ. In diesem Fall gilt $\lambda(Z_m) = \frac{1}{N}$ für $0 \leq m \leq N-1$, sodass für λ-fast alle $x \in I$ in der Ziffernentwicklung von x zur Basis N jede Ziffer mit Frequenz $\frac{1}{N}$ vorkommt.

Kettenbrüche

Auch die Kettenbruchentwicklung einer Zahl $x \in (0,1]$ kann man durch eine Transformation beschreiben. Wir setzen wiederum $I = [0,1]$ und definieren die Kettenbruchtransformation $K: I \longrightarrow I$ durch

$$K(x) = \begin{cases} \frac{1}{x} \pmod 1 & \text{für } x \in (0,1], \\ 0 & \text{für } x = 0. \end{cases} \tag{2.9}$$

Diese Transformation ist nicht stetig. Für $m \geq 1$ sei $Z_m = (\frac{1}{m+1}, \frac{1}{m}]$. Dann gilt $K(x) = \frac{1}{x} - m$ für $x \in Z_m$. Weiterhin sei $g_m: I \longrightarrow [\frac{1}{m+1}, \frac{1}{m}]$ durch $g_m(y) = \frac{1}{m+y}$ definiert. Es gilt dann $g_m(K(x)) = x$ für $x \in Z_m$, und g_m ist die Umkehrfunktion von $K|_{Z_m}$.

Sei $x \in I$. Für $k \geq 0$ sei $i_k \in \mathbb{N}$ so gewählt, dass $K^k(x) \in Z_{i_k}$ gilt. Ist $K^k(x) = 0$, dann ist i_k nicht definiert. Es gilt dann $g_{i_k}(K^{k+1}(x)) = K^k(x)$. Für $r \geq 1$ erhalten wir damit $x = g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(K^r(x))$, und $g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(0)$ ist der aus den Ziffern $i_0, i_1, \ldots, i_{r-1}$ gebildete Kettenbruch. Man nennt die Folge $i_0 i_1 i_2 \ldots$ die *Kettenbruchentwicklung* von x.

Lemma 2.44 Ist $x \in I$ und $K^k(x) = 0$ für ein $k \geq 0$, dann gilt $x \in \mathbb{Q}$.

Beweis Sei r minimal mit $K^r(x) = 0$ und $i_0, i_1, \ldots, i_{r-1}$ so, dass $K^k(x) \in Z_{i_k}$ für $0 \leq k \leq r-1$ gilt. Wir haben dann $x = g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(0)$. Mit Hilfe der Definition von g_m ist leicht zu erkennen, dass x eine rationale Zahl ist. □

Dieses Lemma zeigt, dass die Kettenbruchentwicklung einer irrationalen Zahl $x \in I$ nicht abbricht.

Definition 2.45
Sind $i_0, i_1, i_2, \ldots$ Zahlen in $\mathbb{N}$, dann definieren wir p_k und q_k für $k \geq 0$ durch $p_0 = 0$, $q_0 = 1$, $p_1 = 1$, $q_1 = i_0$ und $p_{k+1} = i_k p_k + p_{k-1}$, $q_{k+1} = i_k q_k + q_{k-1}$ für $k \geq 1$. Ist $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung einer Zahl x, dann schreiben wir $p_k(x)$ und $q_k(x)$ für die so entstehenden natürlichen Zahlen.

Lemma 2.46 Seien $i_0, i_1, i_2, \ldots$ in $\mathbb{N}$. Für $r \geq 1$ gilt dann $g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(y) = \frac{p_{r-1}y+p_r}{q_{r-1}y+q_r}$ und $p_{r-1}q_r - q_{r-1}p_r = (-1)^r$. Außerdem gilt $p_r \geq \frac{1}{2}(\sqrt{2})^r$ und $q_r \geq \frac{1}{2}(\sqrt{2})^r$.

Beweis Alle diese Aussagen ergeben sich durch Induktion aus Definition 2.45. □

Ist $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x, dann zeigt Lemma 2.46, dass der aus den ersten r Ziffern gebildete Kettenbruch $g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(0)$ gleich $\frac{p_r(x)}{q_r(x)}$ ist, wobei

$$\frac{p_0(x)}{q_0(x)} = \frac{1}{i_0}, \qquad \frac{p_1(x)}{q_1(x)} = \frac{1}{i_0 + \frac{1}{i_1}}, \qquad \frac{p_2(x)}{q_2(x)} = \frac{1}{i_0 + \frac{1}{i_1 + \frac{1}{i_2}}} \qquad \text{usw.} \tag{2.10}$$

Das folgende Lemma zeigt u. a., dass die Zahl x durch die Partialbrüche $\frac{p_r(x)}{q_r(x)}$ ihrer Kettenbruchentwicklung sehr gut approximiert wird.

Lemma 2.47 Sei $x \in I \setminus \mathbb{Q}$. Für $r \geq 1$ gilt dann $\frac{1}{2q_r(x)q_{r+1}(x)} \leq \left|x - \frac{p_r(x)}{q_r(x)}\right| \leq \frac{1}{q_r(x)^2}$ und $p_r(x) = q_{r-1}(K(x))$.

Beweis Wir erhalten $\left|x - \frac{p_r(x)}{q_r(x)}\right| = \frac{K^r(x)}{q_r(x)^2 + q_r(x)q_{r-1}(x)K^r(x)}$ mit Hilfe von Lemma 2.46, da ja $x = g_{i_0} \circ g_{i_1} \circ \cdots \circ g_{i_{r-1}}(K^r(x))$ gilt. Wegen $0 \leq K^r(x) \leq 1$ folgt daraus auch $\left|x - \frac{p_r(x)}{q_r(x)}\right| \leq \frac{1}{q_r(x)^2}$.

Sei $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x. Wegen $K^r(x) \in Z_{i_r}$ haben wir $\frac{1}{K^r(x)} \leq i_r + 1 \leq i_r + \frac{q_{r+1}(x)}{q_r(x)}$. Es folgt $q_r(x) \leq i_r q_r(x) K^r(x) + q_{r+1}(x) K^r(x)$ und daraus dann $q_r(x) + q_{r-1}(x) K^r(x) \leq 2q_{r+1}(x) K^r(x)$ mit Hilfe von Definition 2.45. Daraus ergibt sich $\frac{1}{2q_r(x)q_{r+1}(x)} \leq \left|x - \frac{p_r(x)}{q_r(x)}\right|$.

Die letzte Aussage zeigen wir mit Induktion. Ist $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x, dann ist $i_1 i_2 i_3 \ldots$ die Kettenbruchentwicklung von $K(x)$. Es gilt $p_1(x) = 1 = q_0(K(x))$ und $p_2(x) = i_1 p_1(x) + p_0(x) = i_1 = q_1(K(x))$ wegen Definition 2.45. Ist $k \geq 2$ und $p_r(K(x)) = q_{r-1}(x)$ für $r \leq k$ gezeigt, dann folgt $p_{k+1}(x) = i_k p_k(x) + p_{k-1}(x) = i_k q_{k-1}(K(x)) + q_{k-2}(K(x)) = q_{k-1}(K(x))$ wieder aus Definition 2.45. Damit ist auch die letzte Aussage bewiesen. □

Wenn $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von $x \in I \setminus \mathbb{Q}$ ist, so besagt Lemma 2.47, dass

$$x = \lim_{r\to\infty} \frac{p_r(x)}{q_r(x)} = \cfrac{1}{i_0 + \cfrac{1}{i_1 + \cfrac{1}{i_2 + \cfrac{1}{\ddots}}}}.$$

Diese Formel erklärt die Bezeichnung „Kettenbruchentwicklung" für diese Darstellung von x.

Sei $h(x) = \frac{1}{(1+x)\log 2}$, $x \in I$. Wir bezeichnen mit γ das Maß auf I, das Dichte h bezüglich des Lebesguemaßes λ hat. Es wird nach seinem Entdecker *Gaußmaß*[11] genannt. Für ein Intervall $J \subset I$ mit den Endpunkten u und v gilt $\gamma(J) = \frac{\log(1+v)}{\log 2} - \frac{\log(1+u)}{log 2}$. Insbesondere haben wir $\gamma(I) = 1$, sodass γ ein Wahrscheinlichkeitsmaß ist.

Lemma 2.48 Das Gaußmaß γ ist invariant unter der Kettenbruchtransformation K.

Beweis Für $J = [0, y)$ gilt $K^{-1}(J) = \{0\} \cup \bigcup_{m=1}^{\infty} (\frac{1}{y+m}, \frac{1}{m}]$. Es folgt

$$\begin{aligned}
\gamma(K^{-1}(J)) &= \frac{1}{\log 2} \sum_{m=1}^{\infty} \left(\log\left(1 + \frac{1}{m}\right) - \log\left(1 + \frac{1}{y+m}\right)\right) \\
&= \frac{1}{\log 2} \sum_{m=1}^{\infty} \big(-\log m + \log(m+1) + \log(y+m) - \log(y+m+1)\big) \\
&= \frac{1}{\log 2} \lim_{k\to\infty} \big(\log(k+1) + \log(y+1) - \log(y+k+1)\big) = \frac{\log(y+1)}{\log 2} = \gamma(J).
\end{aligned}$$

Wie im letzten Schritt des Beweises von Lemma 2.40 folgt aus dieser Rechnung, dass γ invariant unter K ist. □

Im folgenden Lemma untersuchen wir Zylindermengen, die wir jetzt als offene Intervalle auffassen. Für $m \geq 1$ sei $Z_m^0 = (\frac{1}{m+1}, \frac{1}{m})$ das Innere des Intervalls Z_m.

Lemma 2.49 Sei $C = \bigcap_{k=0}^{n-1} K^{-k}(Z_{i_k}^0)$ mit $n \geq 1$ und $i_0, i_1, \ldots, i_{n-1} \in \mathbb{N}$. Dann ist C ein offenes Intervall und die Abbildung $K^n\colon C \longrightarrow (0,1)$ ist bijektiv und streng monoton. Ist $J = (a,b)$ ein Intervall mit $0 \leq a < b \leq 1$, dann ist $K^{-n}(J) \cap C$ ein offenes Intervall mit Länge $\frac{b-a}{(q_{n-1}a+q_n)(q_{n-1}b+q_n)}$. Insbesondere hat C die Länge $\frac{1}{q_n(q_{n-1}+q_n)}$.

[11] Carl Friedrich Gauß (1777–1855) war ein deutscher Mathematiker, der fundamentale Beiträge zur Zahlentheorie, Statistik, Analysis, Differentialgeometrie, Geophysik, Astronomie, Optik und vielen anderen Gebieten leistete.

Beweis Seien $D_j = \bigcap_{k=0}^{j-1} K^{-k}(Z^0_{i_{n-j+k}})$ und $f_j = g_{i_{n-j}} \circ g_{i_{n-j+1}} \circ \dots \circ g_{i_{n-1}}$ für $1 \le j \le n$. Zuerst beweisen wir mit Induktion, dass D_j ein offenes Intervall und $K^j\colon D_j \longrightarrow (0,1)$ bijektiv und streng monoton mit Umkehrfunktion f_j ist.

Es gilt $D_1 = Z^0_{i_{n-1}}$. Das ist ein offenes Intervall, das durch K streng monoton auf $(0,1)$ abgebildet wird und $f_1 = g_{i_{n-1}}$ als Umkehrfunktion hat. Somit gilt obige Aussage für $j = 1$.

Für den Induktionsschritt nehmen wir an, dass die obige Aussage für $j = r-1$ bewiesen ist. Nun ist $K\colon Z^0_{i_{n-r}} \longrightarrow (0,1)$ bijektiv und streng monoton mit Umkehrfunktion $g_{i_{n-r}}$. Es folgt, dass $D_r = Z^0_{i_{n-r}} \cap K^{-1}(D_{r-1})$ ein offenes Intervall ist und $K^r\colon D_r \longrightarrow (0,1)$ als Hintereinanderausführung von $K\colon D_r \longrightarrow D_{r-1}$ und $K^{r-1}\colon D_{r-1} \longrightarrow (0,1)$ bijektiv und streng monoton ist. Die Umkehrfunktion dieser Abbildung erhält man als Hintereinanderausführung der entsprechenden Umkehrfunktionen. Sie ist daher die Funktion $g_{i_{n-r}} \circ f_{r-1} = f_r$.

Somit ist obige Aussage durch Induktion bewiesen. Für $j = n$ besagt sie, dass C ein offenes Intervall und $K^n\colon C \longrightarrow (0,1)$ bijektiv und streng monoton mit Umkehrfunktion f_n ist. Ist $J = (a,b)$, dann ist $K^{-n}(J) \cap C$ ein offenes Intervall mit Endpunkten $f_n(a) = \frac{p_{n-1}a+p_n}{q_{n-1}a+q_n}$ und $f_n(b) = \frac{p_{n-1}b+p_n}{q_{n-1}b+q_n}$, dessen Länge $\frac{b-a}{(q_{n-1}a+q_n)(q_{n-1}b+q_n)}$ ist, wie man mit Hilfe von Lemma 2.46 berechnet. Die Länge des Intervalls C erhält man, indem man $a = 0$ und $b = 1$ setzt. □

Nachdem wir diese Resultate als Vorbereitung bewiesen haben, kehren wir jetzt zurück zur Ergodentheorie.

Satz 2.50 *Die Kettenbruchtransformation K ist eine ergodische Transformation auf dem Wahrscheinlichkeitsraum $(I, \mathcal{B}_I, \gamma)$.*

Beweis Sei $B \in \mathcal{B}_I$ mit $K^{-1}(B) = B$ und $\gamma(B) > 0$. Wir müssen $\gamma(B) = 1$ zeigen. Dazu verwenden wir Lemma 2.42. Seien $\mathcal{H}$ die Menge aller Zylindermengen $Z_{i_0 i_1 \dots i_{n-1}} := \bigcap_{k=0}^{n-1} K^{-k}(Z^0_{i_k})$ mit $n \ge 1$ und $i_0, i_1, \dots, i_{n-1} \in \mathbb{N}$. Wir zeigen, dass (1) und (2) aus Lemma 2.42 erfüllt sind.

Sei $U \subset I$ ein offenes Intervall. Für $x \in U \setminus \mathbb{Q}$ sei $i_0 i_1 i_2 \dots$ die Kettenbruchentwicklung von x. Wegen Lemma 2.49 hat das Intervall $Z_{i_0 i_1 \dots i_{n-1}}$ die Länge $\frac{1}{q_n(q_{n-1}+q_n)}$, die wegen Lemma 2.46 mit $n \to \infty$ gegen 0 geht. Wir finden daher ein n, sodass $x \in C_x \subset U$ für $C_x = Z_{i_0 i_1 \dots i_{n-1}}$ gilt. Sei $\tilde{\mathcal{U}} = \{C_x : x \in U \setminus \mathbb{Q}\}$. Dann ist $\tilde{\mathcal{U}}$ eine Überdeckung von $U \setminus \mathbb{Q}$. Da zwei Zylindermengen entweder disjunkt sind oder die eine in der anderen enthalten ist, können wir alle $C \in \tilde{\mathcal{U}}$ mit $C \subsetneq D$ für ein $D \in \tilde{\mathcal{U}}$ aus $\tilde{\mathcal{U}}$ herausnehmen und erhalten eine Teilmenge $\mathcal{U}$ von $\tilde{\mathcal{U}}$, die aus paarweise disjunkten, in U enthaltenen Zylindermengen besteht und immer noch $U \setminus \mathbb{Q}$ überdeckt. Setzt man $V = \bigcup_{C \in \mathcal{U}} C$, dann gilt $U \setminus \mathbb{Q} \subset V \subset U$. Es folgt $\gamma(U \setminus V) = 0$, und (1) ist gezeigt.

Um (2) zu zeigen, sei $C = Z_{i_0 i_1 \dots i_{n-1}} \in \mathcal{H}$ und $J = (a,b) \subset I$. Es gilt dann $(q_{n-1}a + q_n)(q_{n-1}b + q_n) \le 2q_n(q_{n-1} + q_n)$ wegen $q_{n-1} \le q_n$. Mit Hilfe von Lemma 2.49 folgt $\lambda(K^{-n}(J) \cap C) \ge \frac{1}{2}\lambda(J)\lambda(C)$. Sei $\nu(A) = \frac{2\lambda(K^{-n}(A)\cap C)}{\lambda(C)}$ für $A \in \mathcal{B}_I$. Dann ist ν ein Maß auf

$\mathcal{B}_I$ und wir haben $\nu(J) \geq \lambda(J)$ für alle offenen Intervalle $J \subseteq I$ gezeigt. Ist $G \subset I$ eine offene Menge, dann lässt sich G schreiben als $E \cup \bigcup_{k\geq 1} J_k$ mit $E \subset \{0,1\}$ und offenen Intervallen $J_1, J_2, \ldots$, die paarweise disjunkt sind. Wegen $\nu(G) = \sum_{k\geq 1} \nu(J_k)$ und $\lambda(G) = \sum_{k\geq 1} \lambda(J_k)$ erhalten wir $\nu(G) \geq \lambda(G)$. Da endliche Maße auf der Borel-σ-Algebra regulär sind [3, Kapitel VII], existiert für alle $A \in \mathcal{B}_I$ und alle $\varepsilon > 0$ eine offene Teilmenge G von I mit $A \subset G$ und $\nu(G \setminus A) < \varepsilon$. Es folgt $\nu(A) + \varepsilon > \nu(G) \geq \lambda(G) \geq \lambda(A)$. Da $\varepsilon > 0$ beliebig ist, ist damit $\nu(A) \geq \lambda(A)$, d. h. $\lambda(K^{-n}(A) \cap C) \geq \frac{1}{2}\lambda(A)\lambda(C)$, für alle $A \in \mathcal{B}_I$ gezeigt.

Außerdem gilt noch $\frac{1}{2\log 2} \leq h \leq \frac{1}{\log 2}$ für die Dichte h von γ. Wir erhalten $\gamma(K^{-n}(A) \cap C) \geq \frac{1}{2\log 2}\lambda(K^{-n}(A) \cap C) \geq \frac{1}{4\log 2}\lambda(A)\lambda(C) \geq \frac{\log 2}{4}\gamma(A)\gamma(C)$ für alle $A \in \mathcal{B}_I$. Setzt man $A = B$ und $d = \frac{\log 2}{4}\gamma(B) > 0$, dann hat man $\gamma(B \cap C) \geq d\gamma(C)$ wegen $K^{-n}(B) = B$. Das gilt für alle $C \in \mathcal{H}$. Damit ist (2) gezeigt.

Aus Lemma 2.42 folgt jetzt, dass $\gamma(B) = 1$ gilt. Damit haben wir bewiesen, dass K eine ergodische Transformation ist. □

Satz 2.51 *Sei $\varphi\colon \mathbb{N} \longrightarrow [0,\infty)$ eine Abbildung und $S(\varphi) = \frac{1}{\log 2}\sum_{j=1}^{\infty} \varphi(j)\log(1+\frac{1}{j(j+2)})$. Wir lassen $S(\varphi) = \infty$ zu. Dann gilt $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \varphi(i_k) = S(\varphi)$ für λ-fast alle $x \in I$, wobei $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x ist.*

Beweis Für $m \geq 1$ sei $Z_m = (\frac{1}{m+1}, \frac{1}{m}]$. Sei $f(x) = \varphi(m)$, wenn $x \in Z_m$. Für $x \in I \setminus \mathbb{Q}$ und $n \geq 1$ gilt dann $\frac{1}{n}\sum_{k=0}^{n-1} \varphi(i_k) = \frac{1}{n}\sum_{k=0}^{n-1} f(K^k(x))$, wobei $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x ist.

Wir nehmen zunächst an, dass $\int f\,d\gamma < \infty$. Da die Kettenbruchtransformation nach Satz 2.50 ergodisch ist, folgt $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_j(T^k(x)) = \int f\,d\gamma$ γ-fast sicher aus dem Ergodensatz. Wegen $\gamma(Z_j) = \frac{1}{\log 2}\log(1 + \frac{1}{j(j+2)})$ ist $\int f\,d\gamma = S(\phi)$. Da die Maße λ und γ dieselben Nullmengen haben, ist γ-fast alle gleichbedeutend mit λ-fast alle.

Die Aussage des Satzes gilt aber auch, wenn $\int f\,d\gamma = \infty$ ist. Um das zu zeigen, sei $f_j = \min(f, j)$ für $j \geq 1$. Wegen $f \geq 0$ folgt $\lim_{j\to\infty} \int f_j d\gamma = \infty$ aus dem Satz über monotone Konvergenz. Nach dem Ergodensatz existiert eine Menge M_j mit $\gamma(M_j) = 0$, sodass $\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) \geq \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_j(T^k(x)) = \int f_j d\gamma$ für alle $x \notin M_j$ gilt. Für $M = \bigcup_{j=1}^{\infty} M_j$ gilt $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) = \infty$ für alle $x \notin M$ und $\gamma(M) = 0$. Dann ist aber auch $\lambda(M) = 0$, womit der Satz auch in diesem Fall bewiesen ist. □

Setzt man in Satz 2.51 für φ die Funktion ψ_m ein, die durch $\psi_m(j) = 0$ für $j \neq m$ und $\psi_m(m) = 1$ definiert ist, dann erhält man, dass für λ-fast alle x die Ziffer m in der Kettenbruchentwicklung von x mit Frequenz $\frac{1}{\log 2}\log(1 + \frac{1}{m(m+2)})$ vorkommt. Setzt man $\varphi(m) = m$ und $\varphi(m) = \log m$, so erhält man Aussagen über Mittelwerte. Für λ-fast alle $x \in I$ gilt $\lim_{n\to\infty} \frac{1}{n}(i_0 + i_1 + \cdots + i_{n-1}) = \infty$ und $\lim_{n\to\infty}(i_0 i_1 \ldots i_{n-1})^{1/n} = \prod_{j=0}^{\infty}(\frac{(j+1)^2}{j(j+2)})^{\log j/\log 2}$, wobei $i_0 i_1 i_2 i_3 \ldots$ die Kettenbruchentwicklung von x ist.

Aufgabe 2.52
Man zeige $\int \log x \, d\gamma(x) = -\frac{\pi^2}{12\log 2}$.

Satz 2.53 *Für $x \in I \setminus \mathbb{Q}$ und $n \geq 1$ sei $Y_n(x) = \bigcap_{k=0}^{n-1} K^{-k}(Z_{i_k}^0)$, wobei $i_0 i_1 i_2 \ldots$ die Kettenbruchentwicklung von x ist. Für λ-fast alle $x \in I$ gilt dann*

(1) $\lim_{n\to\infty} \frac{1}{n} \log q_n(x) = \frac{\pi^2}{12\log 2}$.
(2) $\lim_{n\to\infty} \frac{1}{n} \log |x - \frac{p_n(x)}{q_n(x)}| = -\frac{\pi^2}{6\log 2}$.
(3) $\lim_{n\to\infty} \frac{1}{n} \log \lambda(Y_n(x)) = \lim_{n\to\infty} \frac{1}{n} \log \gamma(Y_n(x)) = -\frac{\pi^2}{6\log 2}$.

Beweis Sei $x \in I \setminus \mathbb{Q}$. Dann gilt $p_r(x) = q_{r-1}(K(x))$ für $r \geq 1$ nach Lemma 2.47 und daher auch $-\log q_n(x) = \sum_{j=0}^{n-1} \log \frac{p_{n-j}(K^j(x))}{q_{n-j}(K^j(x))}$ für $n \geq 1$.

Für $y \in I \setminus \mathbb{Q}$ sei $\rho_k(y) = \log \frac{p_k(y)}{q_k(y)} - \log y$ und $\vartheta_k(y) = |\frac{p_k(y)}{q_k(y)} - y|$. Aus dem Mittelwertsatz folgt $|\rho_k(y)| = \frac{1}{\xi}\vartheta_k(y)$ für ein ξ zwischen y und $\frac{p_k(y)}{q_k(y)}$. Wegen Lemma 2.47 gilt $\vartheta_k(y) \leq \frac{1}{q_k(y)^2}$, woraus wir $\xi \geq \frac{p_k(y)}{q_k(y)} - \frac{1}{q_k(y)^2} = \frac{p_k(y)q_k(y)-1}{q_k(y)^2}$ und $|\rho_k(y)| \leq \frac{q_k(y)^2}{p_k(y)q_k(y)-1}\frac{1}{q_k(y)^2} = \frac{1}{p_k(y)q_k(y)-1}$ erhalten. Wegen Lemma 2.46 gilt aber $p_k(y) \geq \frac{1}{2}\sqrt{2}^k$ und $q_k(y) \geq \frac{1}{2}\sqrt{2}^k$, womit $|\rho_k(y)| \leq \frac{4}{2^k - 4}$ für $k \geq 3$ folgt. Da das für alle $y \in I \setminus \mathbb{Q}$ gilt, erhalten wir $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \rho_{n-j}(K^j(x)) = 0$.

Da die Kettenbruchtransformation ergodisch ist und $\int \log x \, d\gamma(x) = -\frac{\pi^2}{12\log 2}$ gilt, erhalten wir $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log(K^j(x)) = -\frac{\pi^2}{12\log 2}$ für γ-fast alle und damit auch für λ-fast alle x aus dem Ergodensatz. Oben haben wir gezeigt, dass $-\frac{1}{n}\log q_n(x) = \frac{1}{n}\sum_{j=0}^{n-1}\log(K^j(x)) + \frac{1}{n}\sum_{j=0}^{n-1}\rho_{n-j}(K^j(x))$ für alle $x \in I \setminus \mathbb{Q}$ und $n \geq 1$ gilt. Es folgt $\lim_{n\to\infty} \frac{1}{n}\log q_n(x) = \frac{\pi^2}{12\log 2}$ für λ-fast alle x.

Für $x \in I \setminus \mathbb{Q}$ und $n \geq 1$ gilt $\frac{1}{2q_n(x)q_{n+1}(x)} \leq |x - \frac{p_n(x)}{q_n(x)}| \leq \frac{1}{q_n(x)^2}$ nach Lemma 2.47. Mit Hilfe von (1) folgt $\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{2q_n(x)q_{n+1}(x)} = -\frac{\pi^2}{6\log 2}$ und $\lim_{n\to\infty} \frac{1}{n}\log\frac{1}{q_n(x)^2} = -\frac{\pi^2}{6\log 2}$ für λ-fast alle x. Daraus wieder ergibt sich, dass $\lim_{n\to\infty} \frac{1}{n}\log|x - \frac{p_n(x)}{q_n(x)}| = -\frac{\pi^2}{6\log 2}$ für λ-fast alle x gilt.

Für $x \in I \setminus \mathbb{Q}$ und $n \geq 1$ gilt $\lambda(Y_n(x)) = \frac{1}{q_n(x)(q_{n-1}(x)+q_n(x))}$ nach Lemma 2.49, woraus wir

$$\frac{1}{2q_n(x)^2} \leq \lambda(Y_n(x)) \leq \frac{1}{q_n(x)^2}$$

erhalten. Für λ-fast alle x gilt $\lim_{n\to\infty} \frac{1}{n}\log\frac{1}{2q_n(x)^2} = -\frac{\pi^2}{6\log 2}$ und $\lim_{n\to\infty}\frac{1}{n}\log\frac{1}{q_n(x)^2} = -\frac{\pi^2}{6\log 2}$ wegen (1) und daher auch $\lim_{n\to\infty}\frac{1}{n}\log\lambda(Y_n(x)) = -\frac{\pi^2}{6\log 2}$. Da $\frac{1}{2\log 2} \leq h \leq \frac{1}{\log 2}$

für die Dichte h von γ gilt, haben wir

$$\frac{1}{2\log 2}\lambda(Y_n(x)) \le \gamma(Y_n(x)) \le \frac{1}{\log 2}\lambda(Y_n(x)),$$

womit dann

$$\lim_{n\to\infty}\frac{1}{n}\log\gamma(Y_n(x)) = -\frac{\pi^2}{6\log 2}$$

für λ-fast alle x folgt. □

Zum Abschluss stellen wir noch die Frage, wie viele Ziffern der Kettenbruchentwicklung man typischerweise aus den ersten n Dezimalstellen einer Zahl bestimmen kann. Für $x \in I \setminus \mathbb{Q}$ und $n \ge 1$ sei $D_n(x) = (\frac{j}{10^n}, \frac{j+1}{10^n})$, wobei $j \in \{0, 1, \dots, 10^n - 1\}$ so gewählt ist, dass $x \in D_n(x)$ gilt. Wegen $x \notin \mathbb{Q}$ ist das möglich. Weiterhin sei $m_n(x)$ die größte natürliche Zahl m, sodass alle Zahlen in $D_n(x)$ dieselben ersten m Ziffern in der Kettenbruchentwicklung haben. Dann ist $m_n(x)$ die Länge des Anfangsstücks in der Kettenbruchentwicklung der Zahl x, das durch die ersten n Dezimalstellen bestimmt ist.

Wir versuchen $\lim_{n\to\infty}\frac{m_n(x)}{n}$ zu berechnen. Existiert dieser Grenzwert, dann ist $m_n(x)$ für große n näherungsweise bestimmt.

Sei $Z_m^0 = (\frac{1}{m+1}, \frac{1}{m})$ das Innere von Z_m. Für $n \ge 1$ und $i_0, i_1, \dots, i_{n-1} \in \mathbb{N}$ nennen wir $C = \bigcap_{k=0}^{n-1} K^{-k}(Z_{i_k}^0)$ eine Zylindermenge vom Rang n. Wir definieren dann $C^+ = \bigcap_{k=0}^{n-2} K^{-k}(Z_{i_k}^0) \cap K^{-(n-1)}(Z_{i_{n-1}+1}^0)$.

Lemma 2.54 Sei C eine Zylindermenge. Dann gilt $\lambda(C) \le 3\lambda(C^+)$.

Beweis Sei $C = \bigcap_{k=0}^{n-1} K^{-k}(Z_{i_k}^0)$. Wegen Lemma 2.49 und Definition 2.45 erhalten wir $\lambda(C) = \frac{1}{q_n(q_{n-1}+q_n)} = \frac{1}{(i_{n-1}q_{n-1}+q_{n-2})((i_{n-1}+1)q_{n-1}+q_{n-2})}$. Wendet man diese Formel auf C^+ an, so ergibt sich $\lambda(C^+) = \frac{1}{((i_{n-1}+1)q_{n-1}+q_{n-2})((i_{n-1}+2)q_{n-1}+q_{n-2})}$. Wegen $i_{n-1} \ge 1$ folgt daraus $\lambda(C) \le 3\lambda(C^+)$. □

Für $x \in I \setminus \mathbb{Q}$ und $n \ge 1$ sei $Y_n(x)$ wie in Satz 2.53 definiert.

Lemma 2.55 Seien $x \in I \setminus \mathbb{Q}$ und $n \ge 1$. Sei $m = m_n(x)$. Dann gilt $Y_{m+2}^+(x) \subset D_n(x)$ oder $Y_{m+3}^+(x) \subset D_n(x)$.

Beweis Nach Definition von $m = m_n(x)$ liegt ein Endpunkt a des Intervalls $Y_{m+1}(x)$ in $D_n(x)$, da ja nicht alle Zahlen in $D_n(x)$ dieselben ersten $m+1$ Ziffern in der Kettenbruchentwicklung haben. Nun zerfällt $Y_{m+1}(x)$ in abzählbar viele Zylindermengen vom Rang

$m+2$, die sich an einem Endpunkt von $Y_{m+1}(x)$ häufen. Ist a dieser Endpunkt, dann liegt $Y^+_{m+2}(x)$ zwischen a und x, woraus $Y^+_{m+2}(x) \subset D_n(x)$ folgt. Ansonsten betrachten wir $Y_{m+2}(x)$. Dieses Intervall hat einen Endpunkt b, der zwischen a und x liegt oder gleich a ist. Da K stückweise monoton fallend ist, häufen sich die abzählbar vielen Zylindermengen vom Rang $m+3$, in die $Y_{m+2}(x)$ zerfällt, jetzt beim Endpunkt b. Es folgt, dass $Y^+_{m+3}(x)$ zwischen b und x liegt, woraus $Y^+_{m+3}(x) \subset D_n(x)$ folgt. □

Jetzt können wir das gewünschte Resultat beweisen.

Satz 2.56 *Für λ-fast alle $x \in I$ gilt* $\lim_{n\to\infty} \frac{m_n(x)}{n} = \frac{6 \log 2 \log 10}{\pi^2}$.

Beweis Wegen $D_n(x) \subset Y_{m_n(x)}(x)$ gilt $\log \lambda(D_n(x)) \le \log \lambda(Y_{m_n(x)}(x))$, woraus $\frac{m_n(x)}{n} \le \frac{1}{n} \log \lambda(D_n(x)) / \frac{1}{m_n(x)} \log \lambda(Y_{m_n(x)}(x))$ folgt. Wegen $\lambda(D_n(x)) = 10^{-n}$ und Satz 2.53 erhalten wir $\limsup_{n\to\infty} \frac{m_n(x)}{n} \le \frac{6 \log 2 \log 10}{\pi^2}$ für λ-fast alle x, da mit n ja auch $m_n(x)$ gegen ∞ geht.

Lemma 2.55 zeigt, dass $D_n(x) \supset Y^+_{\tilde{m}_n(x)}(x)$ ist, wobei $\tilde{m}_n(x)$ entweder $m_n(x)+2$ oder $m_n(x)+3$ ist. Es folgt $\log \lambda(D_n(x)) \ge \log \lambda(Y^+_{\tilde{m}_n(x)}(x))$ und daraus dann $\frac{m_n(x)}{n} \ge \frac{m_n(x)}{\tilde{m}_n(x)} \frac{1}{n} \log \lambda(D_n(x)) / \frac{1}{\tilde{m}_n(x)} \log \lambda(Y^+_{\tilde{m}_n(x)}(x))$. Wegen Lemma 2.54 haben wir $\lambda(Y^+_{\tilde{m}_n(x)}(x)) \ge \frac{1}{3} \lambda(Y_{\tilde{m}_n(x)}(x))$, und es gilt $\lambda(D_n(x)) = 10^{-n}$, womit $\liminf_{n\to\infty} \frac{m_n(x)}{n} \ge \frac{6 \log 2 \log 10}{\pi^2}$ für λ-fast alle x aus Satz 2.53 folgt. □

2.3.3 Stochastische Prozesse

Es sei $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum. In der Wahrscheinlichkeitstheorie nennt man eine Funktion $Y \in \mathcal{L}_1(\mu)$ Zufallsvariable und das Integral $\int Y \, d\mu$ ihren Erwartungswert. Seien T eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ und $f \in \mathcal{L}_1(\mu)$. Durch $Y_k = f \circ T^k$ für $k \ge 0$ ist eine Folge von Zufallsvariablen definiert. Ist $m \ge 1$ und sind $B_1, B_2, \ldots, B_m$ Mengen in der Borel-σ-Algebra $\mathcal{B}_{\mathbb{R}}$ von $\mathbb{R}$, dann gilt $Y_k^{-1}(B_k) = T^{-k}(f^{-1}(B_k)) = T^{-1}(Y_{k-1}^{-1}(B_k))$ für $1 \le k \le m$ und daher auch $\mu(\bigcap_{k=1}^m Y_k^{-1}(B_k)) = \mu(\bigcap_{k=1}^m Y_{k-1}^{-1}(B_k))$, da T maßtreu ist. Eine Folge $Y_0, Y_1, Y_2, \ldots$ von Zufallsvariablen mit dieser Eigenschaft nennt man einen stationären stochastischen Prozess. Der Ergodensatz macht eine Aussage über die fast sichere Konvergenz von $\frac{1}{n}(Y_0 + Y_1 + \cdots + Y_{n-1})$ für $n \to \infty$. Derartige Konvergenzaussagen nennt man ein starkes Gesetz der großen Zahlen. Der Ergodensatz ist somit ein starkes Gesetz der großen Zahlen für stationäre stochastische Prozesse. Wenn T ergodisch ist, dann erhält man den Erwartungswert $\int f \, d\mu$ als Grenzwert.

Unabhängige Zufallsvariablen

Der einfachste Fall eines stationären Prozesses ist eine Folge von unabhängigen Zufallsvariablen, die alle die gleiche Verteilung haben. Ist diese Verteilung diskret und auf einer endlichen Teilmenge A von $\mathbb{R}$ konzentriert, dann ist der Shiftraum $\mathsf{A}^{\mathbb{N}_0}$ mit einem Bernoullimaß ein geeigneter Wahrscheinlichkeitsraum. Wenn die Menge A N Elemente hat, so nehmen wir wie im Abschn. 1.2 über symbolische Dynamik der Einfachheit halber an, dass $\mathsf{A} = \{0, \ldots, N-1\}$ ist.

Satz 2.57 *Es sei Σ_N^+ der einseitige N-Shift. Seien $\pi = (\pi_i)_{i=0,\ldots,N-1}$ ein stochastischer Vektor und μ das π-Bernoullimaß auf Σ_N^+. Die Shifttransformation σ ist dann stark mischend und daher auch ergodisch.*

Beweis Siehe Aufgabe 14. □

Das starke Gesetz der großen Zahlen folgt jetzt aus Satz 2.57 und dem Ergodensatz. Ist μ das π-Bernoullimaß auf dem Shiftraum Σ_N^+ und $f \in \mathcal{L}_1(\mu)$, dann konvergiert $\frac{1}{n}(Y_0 + Y_1 + \cdots + Y_{n-1})$ für $n \to \infty$ fast sicher gegen $\int f\,d\mu$, wobei $Y_0, Y_1, Y_2, \ldots$ der durch $Y_k = f \circ \sigma^k$ für $k \geq 0$ definierte stochastische Prozess ist. Wählt man insbesondere das Alphabet A als Teilmenge von $\mathbb{R}$ und definiert f durch $f(i_0 i_1 i_2 \ldots) = i_0$, dann sind die Zufallsvariablen $Y_0, Y_1, Y_2, \ldots$ unabhängig und haben alle die durch den stochastischen Vektor π bestimmte Verteilung, da ja $\mu(\bigcap_{k=0}^m Y_k^{-1}(i_k)) = \mu({}_0[i_0 i_1 \ldots i_m]) = \prod_{k=0}^m \pi_{i_k} = \prod_{k=0}^m \mu(Y_k^{-1}(i_k))$ für alle $k \geq 0$ und alle $i_0, i_1, \ldots, i_m \in \mathsf{A}$ gilt. Wir haben somit ein starkes Gesetz der großen Zahlen für unabhängige Zufallsvariablen gefunden.

Dieses Beispiel lässt sich leicht verallgemeinern. Sei ν ein beliebiges Wahrscheinlichkeitsmaß auf der Borel-σ-Algebra $\mathcal{B}_\mathbb{R}$ in $\mathbb{R}$. Man kann dann den Produktraum $(X, \mathcal{S}, \mu) = (\mathbb{R}, \mathcal{B}_\mathbb{R}, \nu)^{\mathbb{N}_0}$ definieren. Für $\mathbf{x} = (x_0, x_1, x_2, \ldots) \in X = \mathbb{R}^{\mathbb{N}_0}$ sei $\sigma(\mathbf{x}) = (x_1, x_2, x_3, \ldots)$. Dann ist σ eine maßtreue Transformation auf $(X, \mathcal{S}, \mu)$. Die Menge $\mathcal{E} = \{A_0 \times A_1 \times \cdots \times A_n \times \mathbb{R} \times \mathbb{R} \times \cdots : n \geq 0,\ A_0, A_1, \ldots, A_n \in \mathcal{B}_\mathbb{R}\}$ ist ein Semiring, der die σ-Algebra $\mathcal{S}$ erzeugt. Wie im Beweis von Satz 2.57 zeigt man, dass $\mu(\sigma^{-n}(U) \cap V) = \mu(U)\mu(V)$ für beliebige Mengen U und V aus $\mathcal{E}$ gilt, wenn n nur groß genug ist. Es folgt, dass die Transformation σ auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$ stark mischend und daher auch ergodisch ist. Ist $f\colon X \longrightarrow \mathbb{R}$ definiert durch $f(x_0, x_1, \ldots) = x_0$ und $Y_k = f \circ \sigma^k$ für $k \geq 0$, dann sind die Zufallsvariablen $Y_0, Y_1, Y_2, \ldots$ unabhängig und haben alle die Verteilung ν. Aus dem Ergodensatz folgt jetzt, dass $\frac{1}{n}(Y_0 + Y_1 + \cdots + Y_{n-1})$ für $n \to \infty$ fast sicher gegen $\int f\,d\nu$ konvergiert. Das ist ein starkes Gesetz der großen Zahlen für unabhängige Zufallsvariablen, die alle Verteilung ν haben.

Markovketten

Sei A eine N-elementige Menge, P eine stochastische $\mathsf{A} \times \mathsf{A}$-Matrix und π ein stochastischer Vektor, sodass $\pi\mathsf{P} = \pi$ gilt. Wie wir in Abschn. 1.3 gesehen haben, ist dadurch ein Wahrscheinlichkeitsmaß μ auf der Borel-σ-Algebra $\mathcal{B}_{\mathsf{A}^{\mathbb{N}_0}}$ des einseitigen Shiftraums $\mathsf{A}^{\mathbb{N}_0}$

gegeben. Wir haben es (π, P)-Markovmaß genannt. Für $n \geq 0$ und $i_0, i_1, \dots, i_n \in \mathsf{A}$ gilt $\mu({}_0[i_0 i_1 \dots i_n]) = \pi_{i_0} \mathsf{P}_{i_0 i_1} \mathsf{P}_{i_1 i_2} \dots \mathsf{P}_{i_{n-1} i_n}$.

Ist $f\colon \Sigma_N^+ \longrightarrow \mathsf{A}$ durch $f(i_0 i_1 i_2 \dots) = f(i_0)$ definiert und $Y_n = f \circ T^n$ für $n \geq 0$, dann nennt man den stochastischen Prozess $Y_0, Y_1, Y_2, \dots$ eine Markovkette mit Zustandsraum A und Übergangsmatrix P.

Wir können annehmen, dass $\pi_i > 0$ für alle $i \in \mathsf{A}$ gilt. Ist das nicht erfüllt, dann können wir es erreichen, indem wir zum Zustandsraum $\mathsf{E} = \{i \in \mathsf{A} : \pi_i > 0\}$ übergehen. Da π ein stochastischer Vektor ist, kann E nicht leer sein. Ist $i \in \mathsf{E}$ und $j \in \mathsf{A} \setminus \mathsf{E}$, dann gilt $\pi_i > 0$ und $\pi_j = 0$, sodass $\mathsf{P}_{ij} = 0$ wegen $\sum_{k \in \mathsf{A}} \pi_k \mathsf{P}_{kj} = \pi_j$ gelten muss. Wir können daher den Vektor π und die Matrix P auf die Menge E einschränken und erhalten dadurch einen stochastischen Vektor $\tilde{\pi}$ und eine stochastische $\mathsf{E} \times \mathsf{E}$-Matrix $\tilde{\mathsf{P}}$, für die $\tilde{\pi}\tilde{\mathsf{P}} = \tilde{\pi}$ und $\tilde{\pi}_i > 0$ für alle $i \in \mathsf{E}$ gilt.

Im Folgenden nehmen wir wiederum ohne Beschränkung der Allgemeinheit an, dass $\mathsf{A} = \{0, \dots, N-1\}$ und $\mathsf{A}^{\mathbb{N}_0} = \Sigma_N^+$ ist. Wir bezeichnen die Borel-σ-Algebra $\mathcal{B}_{\Sigma_N^+}$ von Σ_N^+ mit $\mathcal{B}$.

Wir wollen die Ergodizität der Shifttransformation σ auf dem Wahrscheinlichkeitsraum $(\Sigma_N^+, \mathcal{B}, \mu)$ zu untersuchen. Dazu beginnen wir mit einem Satz über die Konvergenz der Mittelwerte der Potenzen von P, den wir anschließend beim Beweis der Ergodizität verwenden.

Satz 2.58 *Sei μ das durch die stochastische Matrix P und den stochastischen Vektor π definierte Markovmaß auf dem Shiftraum Σ_N^+. Wenn alle Koeffizienten von π positiv sind, dann existiert $\mathsf{Q} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathsf{P}^k$. Weiterhin ist Q eine stochastische Matrix, für die $\pi\mathsf{Q} = \pi$ und $\mathsf{Q} = \mathsf{QP} = \mathsf{Q}^2$ gilt. Ist die Shifttransformation σ ergodisch, dann ist jede Zeile von Q gleich π.*

Beweis Seien i und j in A. Wir setzen $A = {}_0[i]$ und $B = {}_0[j]$. Für $k \geq 1$ gilt dann $\mu(A \cap T^{-k}(B)) = \sum_{i_1 \in \mathsf{A}} \sum_{i_2 \in \mathsf{A}} \cdots \sum_{i_{k-1} \in \mathsf{A}} \mu({}_0[i i_1 i_2 \dots i_{k-1} j]) = \pi_i \mathsf{P}_{ij}^k$. Aus dem Ergodensatz folgt, dass $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_B \circ T^k = E(1_B \,|\, \mathcal{B}^\sigma)$ fast sicher gilt, wobei $\mathcal{B}^\sigma \subset \mathcal{B}$ die σ-Algebra der shiftinvarianten Borelmengen ist. Multipliziert man diese Gleichung mit 1_A und wendet den Satz über dominierte Konvergenz an, so erhält man $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}(B)) = \int 1_A E(1_B \,|\, \mathcal{B}^\sigma)\, d\mu$. Setzt man $\mathsf{Q}_{ij} = \frac{1}{\pi_i} \int 1_A E(1_B \,|\, \mathcal{I}^\sigma)\, d\mu$, dann ist $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathsf{P}_{ij}^k = \mathsf{Q}_{ij}$ gezeigt. Ist die Shifttransformation σ ergodisch, dann gilt $E(1_B \,|\, \mathcal{B}^\sigma) = \mu(B) = \pi_j$, woraus $\int 1_A E(1_B \,|\, \mathcal{B}^\sigma)\, d\mu = \pi_i \pi_j$ und $\mathsf{Q}_{ij} = \pi_j$ folgt. Das bedeutet, dass jede Zeile der Matrix Q gleich π ist.

Für $n \geq 1$ sei $\mathsf{R}_n = \frac{1}{n} \sum_{k=0}^{n-1} \mathsf{P}^k$. Für die Matrix Q gilt dann $\mathsf{Q} = \lim_{n\to\infty} \mathsf{R}_n$. Da die Koeffizienten von P nichtnegativ sind, gilt das auch für alle R_n und damit auch für Q. Des Weiteren gilt $\pi\mathsf{P} = \pi$. Da P eine stochastische Matrix ist, gilt auch $\mathsf{P}e = e$ für den Vektor e, dessen Koeffizienten alle 1 sind. Es folgt $\pi\mathsf{R}_n = \pi$ und $\mathsf{R}_n e = e$ für $n \geq 1$ und daraus dann $\pi\mathsf{Q} = \pi$ und $\mathsf{Q}e = e$. Insbesondere ist Q eine stochastische Matrix. Weiterhin gilt

$R_n P = R_n - \frac{1}{n}P + \frac{1}{n}P^n$ für $n \geq 1$. Da die Koeffizienten von P^n im Intervall I liegen, erhalten wir $QP = \lim_{n\to\infty} R_n P = \lim_{n\to\infty} R_n = Q$. Aus dieser Gleichung folgt jetzt $QP^k = Q$ für alle $k \geq 0$ und damit $QR_n = Q$ für alle $n \geq 1$. Daraus ergibt sich $Q^2 = \lim_{n\to\infty} QR_n = Q$. □

Satz 2.59 *Sei μ das durch die stochastische Matrix P und den stochastischen Vektor π definierte Markovmaß auf dem Shiftraum Σ_N^+. Der Vektor π habe positive Koeffizienten, sodass die Matrix $Q = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k$ existiert. Dann sind die folgenden Aussagen äquivalent.*

(1) Die Shifttransformation σ ist ergodisch.
(2) Alle Zeilen der Matrix Q sind gleich π.
(3) Die Matrix P ist irreduzibel (Definition 1.31).

Beweis Die Implikation (1) ⇒ (2) wurde bereits in Satz 2.58 gezeigt. Es gelte (2). Würden i und j in A existieren mit $P_{ij}^n = 0$ für alle $n \geq 1$, dann würde $Q_{ij} = 0$ folgen und daraus $\pi_j = 0$ wegen (2). Dieser Widerspruch zeigt, dass für jedes i und j in A ein n existiert mit $P_{ij}^n > 0$, womit (3) gezeigt ist.

Es gelte (3). Wir zeigen zuerst, dass Q positive Koeffizienten hat. Sei $i \in A$ beliebig. Da Q nach Satz 2.58 stochastisch ist, existiert ein $k \in A$ mit $Q_{ik} > 0$. Ist jetzt $j \in A$ beliebig, dann existiert ein $n \geq 1$ mit $P_{kj}^n > 0$. Da $Q = QP^n$ nach Satz 2.58 gilt, erhalten wir $Q_{ij} > 0$. Damit ist gezeigt, dass alle Koeffizienten von Q positiv sind.

Sei jetzt $j \in A$ beliebig und $m_j = \max_{i\in A} Q_{ij}$. Ist $u \in A$ so gewählt, dass $m_j = Q_{uj}$ gilt, dann haben wir $m_j = Q_{uj} = \sum_{i\in A} Q_{ui}Q_{ij} \leq \sum_{i\in A} Q_{ui} m_j = m_j$, da Q nach Satz 2.58 stochastisch ist und $Q = Q^2$ erfüllt. Es folgt $Q_{ij} = m_j$ für alle $i \in A$, sonst würde $\sum_{i\in A} Q_{ui}Q_{ij} < \sum_{i\in A} Q_{ui} m_j$ gelten, da ja alle Koeffizienten von Q positiv sind. Schließlich erhalten wir $m_j = \sum_{i\in A} \pi_i m_j = \sum_{i\in A} \pi_i Q_{ij} = \pi_j$, da $\pi = \pi Q$ nach Satz 2.58 gilt. Somit ist $Q_{ij} = \pi_j$ für alle i und j in A gezeigt. Das ist (2).

Um die Implikation (2) ⇒ (1) zu zeigen, gehen wir vor wie im Beweis von Satz 2.57. Die Menge $\mathcal{E} = \{{}_0[i_0 i_1 \ldots i_n] : n \geq 0,\ i_0, i_1, \ldots, i_n \in A\}$ ist ein Semiring, der die Borel-σ-Algebra $\mathcal{B}$ von Σ_N^+ erzeugt. Seien $U = {}_0[u_0 u_1 \ldots u_m]$ und $V = {}_0[v_0 v_1 \ldots v_k]$ beliebige Mengen aus $\mathcal{E}$. Es gilt $\mu(U) = \pi_{u_0} P_{u_0 u_1} \ldots P_{u_{m-1} u_m}$ und $\mu(V) = \pi_{v_0} P_{v_0 v_1} \ldots P_{v_{k-1} v_k}$. Für $n \geq k$ ergibt sich wie im Beweis von Satz 2.57, dass $\mu(\sigma^{-n}(U) \cap V) = \pi_{v_0} P_{v_0 v_1} \ldots P_{v_{k-1} v_k} \cdot P_{v_k u_0}^{n-k} P_{u_0 u_1} \ldots P_{u_{m-1} u_m}$ gilt. Es folgt $\lim_{r\to\infty} \frac{1}{r-k} \sum_{n=k}^{r-1} \mu(\sigma^{-n}(U) \cap V) = \pi_{v_0} P_{v_0 v_1} \ldots P_{v_{k-1} v_k} \cdot Q_{v_k u_0} P_{u_0 u_1} \ldots P_{u_{m-1} u_m}$ aus der Definition der Matrix Q. Da wir (2) voraussetzen, ist auch $Q_{v_k u_0} = \pi_{u_0}$ erfüllt. Wir erhalten somit, dass $\lim_{r\to\infty} \frac{1}{r} \sum_{n=0}^{r-1} \mu(\sigma^{-n}(U) \cap V) = \mu(U)\mu(V)$ gilt. Nach Satz 2.20 ist σ ergodisch und (1) ist gezeigt. □

Die stochastische $N \times N$-Matrix P sei irreduzibel. Ist π ein stochastischer Vektor mit $\pi P = \pi$, dann gilt $\pi_i > 0$ für alle $i \in A$, da für alle i und j in A ein $n \geq 1$ mit $P_{ji}^n > 0$ existiert und wegen $\sum_{u\in A} \pi_u P_{ui}^n = \pi_i$ auch $\pi_i > 0$ gelten muss, wenn $\pi_j > 0$ gilt. Ist μ das

(π, P)-Markovmaß, dann besagt Satz 2.59, dass die Shifttransformation σ auf dem Raum $(\Sigma_N^+, \mathcal{B}, \mu)$ ergodisch ist.

Die Matrix P heißt *aperiodisch*, wenn P^n für alle $n \geq 1$ irreduzibel ist. In diesem Fall kann man zeigen, dass $\lim_{n\to\infty} \mathsf{P}_{ij}^n = \pi_j$ für alle $i, j \in \mathsf{A}$ gilt. Für $U = {}_0[u_0 u_1 \dots u_m]$ und $V = {}_0[v_0 v_1 \dots v_k]$ beweist man dann wie im letzten Teil des Beweises von Satz 2.59, dass $\lim_{n\to\infty} \mu(\sigma^{-n}(U) \cap V) = \mu(U)\mu(V)$ gilt. Nach Satz 2.20 ist σ stark mischend.

Wenn σ auf dem Wahrscheinlichkeitsraum $(\Sigma_N^+, \mathcal{B}, \mu)$ ergodisch ist, so hat für jedes $f \in \mathcal{L}(\mu)$ der reellwertige stochastische Prozess $(Y_k = f \circ \sigma^k)_{k\geq 0}$ aufgrund des Ergodensatzes die Eigenschaft, dass $\frac{1}{n}(Y_0 + Y_1 + \cdots + Y_{n-1})$ für $n \to \infty$ fast sicher gegen $\int f \, d\mu$ konvergiert.

Wenn wir statt der Menge $\mathsf{A} = \{0, \dots, N-1\}$ wiederum eine beliebige endliche Teilmenge von $\mathbb{R}$ betrachten, so folgt aus Satz 2.59 und dem Ergodensatz, dass jeder Markovprozess $(Y_k)_{k\geq 0}$ mit Zustandsraum A und irreduzibler Übergangsmatrix das starke Gesetz der großen Zahlen erfüllt: $\lim_{n\to\infty} \frac{1}{n}(Y_0 + Y_1 + \cdots + Y_{n-1}) = E(Y_0)$ fast sicher, wobei $E(Y_0)$ der Erwartungswert der Zufallsvariablen Y_0 ist.

2.4 Aufgaben

Ergodensätze

1. Es seien (X, T) ein topologisches dynamisches System und μ ein T-invariantes Maß auf X. Zeigen Sie, dass $x \in \omega(x)$ für fast alle $x \in X$ gilt.
2. **Rekurrenz von Mengen** Es seien $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum und $T\colon X \longrightarrow X$ eine maßtreue Transformation. Zeigen Sie, dass es zu jedem $E \in \mathcal{S}$ mit $\mu(E) > 0$ ein n mit $1 \leq n \leq \frac{1}{\mu(E)}$ und $\mu(E \cap T^{-n}(E)) > 0$ gibt.
3. Es sei $(X, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum. Wenn $J \subset \mathbb{R}_+$ ein Intervall und $\psi\colon J \longrightarrow \mathbb{R}$ *konkav* ist (d. h., wenn $\phi = -\psi$ konvex ist), dann zeigen Sie die Ungleichung
$$E(\psi \circ f \,|\, \mathcal{T}) \leq \psi(E(f \,|\, \mathcal{T})) \tag{2.11}$$
für jedes $f\colon X \longrightarrow J$ mit $f \in \mathcal{L}_1(X, \mathcal{S}, \mu)$ und $\psi \circ f \in \mathcal{L}_1(X, \mathcal{S}, \mu)$.
4. Es sei $T\colon X \longrightarrow X$ eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Beweisen Sie Folgendes:
   (a) Eine Funktion $f\colon X \longrightarrow \mathbb{R}$ ist genau dann $\mathcal{S}_\mu^T$-messbar, wenn $f \circ T = f \pmod{\mu}$ gilt.
   (b) Wenn $f\colon X \longrightarrow \mathbb{R}$ $\mathcal{S}_\mu^T$-messbar ist, so existiert eine $\mathcal{S}^T$-messbare (und daher T-invariante) Funktion $g\colon X \longrightarrow \mathbb{R}$ mit $f = g \pmod{\mu}$.
5. Sei $T\colon X \longrightarrow X$ eine invertierbare maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Zeigen Sie, dass für jedes $f \in L_1(\mu)$ die Limiten der Mittel $\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j$ und $\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^{-j}$ für $n \to \infty$ μ-*f.ü.* übereinstimmen.
6. Es sei $(\mathbb{T}, T_3)$ das in Aufgabe 1.2 (5) definierte topologische dynamische System mit $T_3 x = 3x \pmod 1$ für jedes $x \in \mathbb{T}$. Des Weiteren sei $f = 1_{[\frac{1}{3}, \frac{2}{3}]}$ und $C \subset \mathbb{T}$ die klassische Cantormenge. Untersuchen Sie den Ausdruck $\frac{1}{n}\sum_{k=0}^{n-1} f(T_3^k x)$ für $x \in C$.
7. Es seien $\alpha_1 \in \mathbb{R}$ irrational, $\alpha_2 = \alpha_1 + \frac{1}{2} \pmod 1$ und $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. Wie in Aufgabe 1.35 definieren wir $R_{\boldsymbol{\alpha}}\colon \mathbb{T}^2 \longrightarrow \mathbb{T}^2$ durch $R_{\boldsymbol{\alpha}}(\mathbf{x}) = \mathbf{x} + \alpha \pmod 1$ für jedes $\mathbf{x} \in \mathbb{T}^2$.
   (1) Zeigen Sie, dass $R_{\boldsymbol{\alpha}}$ nicht eindeutig ergodisch ist (vgl. Aufgabe 1.35).

(2) Es seien $t \in [0,1)$ und $Q_t = [0,t) \times [0,t) \subset \mathbb{T}^2$. Zeigen Sie, dass die ergodischen Mittel $\frac{1}{n}\sum_{k=0}^{n-1} 1_{Q_t} \circ R_{\boldsymbol{\alpha}}^k$ punktweise gegen eine stetige Funktion $\phi_t \colon \mathbb{T}^2 \longrightarrow \mathbb{R}$ konvergieren, wobei natürlich $\phi_t = E_{\lambda^2}(1_{Q_t} \mid \mathcal{B}_{\mathbb{T}^2}^{R_{\alpha}}) \pmod{\lambda^2}$ ist.
(3) Bestimmen Sie die Funktion ϕ_t für verschiedene Werte von t (z. B. für $t = \frac{1}{4}$ und $t = \frac{1}{2}$).

8. Sei $T \colon X \longrightarrow X$ eine maßtreue Transformation auf dem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$, und $f \colon X \longrightarrow \mathbb{R}$ sei eine nichtnegative messbare Funktion mit $\int f\, d\mu = \infty$. Zeigen Sie, dass $\lim_{n\to\infty} \sum_{k=0}^{n-1} f \circ T^k(x) = \infty$ für fast alle $x \in X$ gilt.
9. (1) Zeigen Sie, dass $\lim_{n\to\infty}(\frac{1}{n}\sum_{k=0}^{n-1} U_T^k)[f] = P_T[f]$ für jedes $f \in \mathcal{L}_2(X, \mathcal{S}^T, \mu)$ gilt, wobei $P_T \colon L_2(X, \mathcal{S}, \mu) \longrightarrow L_2(X, \mathcal{S}, \mu)$ der Projektionsoperator auf den abgeschlossenen Teilraum der U_T-invarianten Elemente von $L_2(X, \mathcal{S}, \mu)$ ist.
(2) Es sei $U_T^* \colon L_2(X, \mathcal{S}, \mu) \longrightarrow L_2(X, \mathcal{S}, \mu)$ der durch die Formel

$$\langle\!\langle U_T v, w \rangle\!\rangle = \langle\!\langle v, U_T^* w \rangle\!\rangle, \quad v, w \in L_2(X, \mathcal{S}, \mu) \tag{2.12}$$

definierte zu U_T *adjungierte* Operator. Zeigen Sie, dass $U_T^* U_T$ die Identitätsabbildung auf $L_2(X, \mathcal{S}, \mu)$ ist, während $U_T U_T^*$ der Projektionsoperator auf den abgeschlossenen Teilraum

$$(\ker U_T^*)^{\perp} = \{v \in L_2(\mu) : U_T^* v = 0\}^{\perp} = U_T(L_2(\mu)) = \mathrm{Im}(U_T)$$

von $L_2(X, \mathcal{S}, \mu)$ ist.

Mischungseigenschaften

10. **Ergodizität expansiver Automorphismen von $\mathbb{T}^d$** Sei A eine $d \times d$-Matrix mit Koeffizienten in $\mathbb{Z}$ und Determinante 1 oder −1. Die Transformation $T_A \colon \mathbb{T}^d \longrightarrow \mathbb{T}^d$ sei wie in Beispiel 2.17 definiert durch $T_A(\mathbf{x}) = A\mathbf{x} \pmod 1$. Beweisen Sie in Verallgemeinerung der Aufgaben 9, dass T_A genau dann expansiv ist (Definition 1.10), wenn A keinen Eigenwert mit Absolutbetrag 1 besitzt. Wegen Beispiel 2.17 ist also jeder expansive Automorphismus T_A von $\mathbb{T}^d$ ergodisch bezüglich λ^d.
11. **Nichtexpansiver ergodischer Automorphismus von $\mathbb{T}^4$** Zeigen Sie, dass die Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}$$

einen nichtexpansiven ergodischen Automorphismus T_A von $\mathbb{T}^4$ definiert.
12. Seien (X, T) ein invertierbares topologisches dynamisches System und $\mu \in \mathcal{M}(X)^T$. Beweisen Sie, dass für μ-fast alle x die Aussage $\overline{\mathcal{O}_T^+(x)} = \overline{\mathcal{O}_{T^{-1}}^+(x)} = \overline{\mathcal{O}_T(x)} = \omega(x)$ gilt.
13. **Beweis des monotonen Klassensatzes** Es seien $\mathcal{A}$ eine Mengenalgebra und $\mathcal{M}$ die kleinste monotone Klasse, die $\mathcal{A}$ enthält. Zeigen Sie folgende Aussagen:
   (i) $\mathcal{C} = \{C \in \mathcal{M} : X \smallsetminus C \in \mathcal{M}\}$ ist eine monotone Klasse und daher $\mathcal{C} = \mathcal{M}$.
   (ii) Sei $A \in \mathcal{A}$ und $\mathcal{V}_A = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\}$. Dann ist $\mathcal{V}_A$ eine monotone Klasse, die $\mathcal{A}$ enthält, und daher $\mathcal{V}_A$. Dies gilt auch für $A \in \mathcal{M}$.
   (iii) $\mathcal{M}$ ist eine Algebra und daher die kleinste σ-Algebra, die $\mathcal{A}$ enthält.
14. **Bernoulli-Maße sind stark mischend** Es sei $N \geq 2$, und $T = \sigma$ sei der Shift (1.11) auf dem Shiftraum $X = \Sigma_N$ in (1.13). Weiterhin sei $\pi = (\pi_0, \dots, \pi_{N-1})$ ein stochastischer Vektor und $\mu =$

μ_π das in Beispiel 1.64 definierte Bernoullimaß, das für jede Zylindermenge ${}_m[i_0 i_1 \ldots i_n]$ durch $\mu({}_m[i_0 i_1 \ldots i_n]) = \prod_{j=0}^{n} \pi_{i_j}$ gegeben ist. Zeigen Sie, dass σ eine stark mischende Transformation auf $(X, \mathcal{B}_X, \mu_\pi)$ ist.
Hinweis: Wenden Sie Satz 2.20 auf den Semiring $\mathcal{E}$ der Zylindermengen an.

15. Sei T eine maßerhaltende Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Zeigen Sie, dass T genau dann schwach mischend ist, wenn es für jedes Paar $A, B \in \mathcal{S}$ eine Teilfolge n_k der natürlichen Zahlen gibt sodass $\lim_{n\to\infty} \frac{1}{n}|\{k : n_k < n\}| = 1$ gilt und die Konvergenz in (2.5) zumindest auf dieser Teilfolge hält.
Hinweis: Verwenden Sie den Beweis von (2) $\Rightarrow$ (1) in Satz 2.24.
16. Beweisen Sie, dass die Gleichung (2.6) für *jedes* $\alpha \in \mathbb{R} \setminus \mathbb{Z}$ gilt.
Hinweis: Es sei $\alpha \in \mathbb{R} \setminus \mathbb{Z}$ rational und von der Form $\alpha = \frac{p}{q}$ mit p und q teilerfremd. Beweisen Sie, dass in diesem Fall die durch $S(x, j) = (Tx, j + p \pmod q)$ gegebene Transformation S auf $X \times \mathbb{Z}/q\mathbb{Z}$ ergodisch bezüglich des Maßes $\nu = \mu \otimes \rho$ ist, wobei ρ das gleichverteilte Wahrscheinlichkeitsmaß auf $\mathbb{Z}/q\mathbb{Z}$ ist.
17. Es sei T eine maßtreue Transformation auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$. Zeigen Sie, dass T genau dann schwach mischend ist, wenn für jede ergodische Transformation S auf einem Wahrscheinlichkeitsraum $(Y, \mathcal{T}, \nu)$ auch $T \times S$ ergodisch bezüglich $\mu \otimes \nu$ ist.
18. Es seien $(X, \mathcal{S}, \mu, T)$ ein invertierbares maßerhaltendes System und $\mathcal{T} \subset \mathcal{S}$ eine strikt invariante Teil-σ-Algebra. Beweisen Sie, dass der in Beispiel 2.31 beschriebene Faktor $(X, \mathcal{T}, \mu, T)$ invertierbar ist.
Wenn also $\mathcal{P}$ eine Zerlegung von X und $\mathcal{T}_\mathcal{P}$ die von $\bigcup_{j=1}^{\infty}(\bigvee_{i=-j}^{j} T^{-i}\mathcal{P})$ erzeugte σ-Algebra ist, so ist der Faktor $(X, \mathcal{T}_\mathcal{P}, \mu, T)$ invertierbar.
19. Es sei $\phi\colon \Sigma_2^+ \longrightarrow \mathbb{T}$ die in (1.17) definierte topologische Faktorabbildung. Außerdem seien μ das durch den stochastischen Vektor $\pi = (\frac{1}{2}, \frac{1}{2})$ definierte Bernoullimaß auf Σ_2^+ (vgl. Beispiel 1.64) und λ das Lebesguemaß auf $\mathbb{T}$. Zeigen Sie, dass ϕ ein Isomorphismus der maßerhaltenden dynamischen Systeme $(\Sigma_2^+, \mathcal{B}_{\Sigma_2^+}, \mu, \sigma)$ und $(\mathbb{T}, \mathcal{B}_T, \lambda, T_2)$ ist.
20. Es seien $(X, \mathcal{S}, \mu, T)$ ein maßerhaltendes dynamisches System und $(X', \mathcal{S}', \mu', T')$ ein Faktor von $(X, \mathcal{S}, \mu, T)$. Zeigen Sie Folgendes:
    (a) Wenn $(X, \mathcal{S}, \mu, T)$ ergodisch, mischend oder schwach mischend ist, so gilt dies auch für $(X', \mathcal{S}', \mu', T')$.
    (b) Wenn $(X, \mathcal{S}, \mu, T)$ und $(X', \mathcal{S}', \mu', T')$ isomorph sind, so ist $(X', \mathcal{S}', \mu', T')$ genau dann ergodisch, mischend oder schwach mischend, wenn die analoge Aussage für $(X, \mathcal{S}, \mu, T)$ gilt.
21. **Totale Ergodizität** Ein maßerhaltendes dynamisches System $(X, \mathcal{S}, \mu, T)$ (bzw. eine maßtreue Transformation T auf einem Wahrscheinlichkeitsraum $(X, \mathcal{S}, \mu)$) heißt *totalergodisch*, wenn $(X, \mathcal{S}, \mu, T^k)$ für jedes $k \geq 1$ ergodisch ist. Beweisen Sie, dass jedes schwach mischende maßerhaltende dynamische System totalergodisch ist und dass ein ergodisches maßerhaltendes dynamisches System $(X, \mathcal{S}, \mu, T)$ genau dann totalergodisch ist, wenn T keinen Eigenwert $\gamma \neq 1$ besitzt, der eine Einheitswurzel ist.
Hinweis: Nehmen Sie an, dass T ergodisch ist, dass aber T^k für ein $k > 1$ nicht ergodisch ist, und betrachten Sie den Faktor $(X, \mathcal{S}^{T^k}, \mu, T)$.
22. Es sei (X, T) ein topologisches dynamisches System, und μ sei ein T-invariantes Wahrscheinlichkeitsmaß auf $\mathcal{B}_X$. Weiterhin sei $(\tilde{X}, \sigma)$ das in Aufgabe 1.47 definierte invertierbare topologische dynamische System und $\pi_0\colon \tilde{X} \longrightarrow X$ die dort definierte Faktorabbildung. Beweisen Sie, dass es auf $\mathcal{B}_{\tilde{X}}$ ein shiftinvariantes Wahrscheinlichkeitsmaß $\tilde{\mu}$ mit $(\pi_0)_*\tilde{\mu} = \mu$ gibt. Damit ist also das maßerhaltende dynamische System $(X, \mathcal{B}_X, \mu, T)$ ein Faktor des invertierbaren maßerhaltenden dynamischen Systems $(\tilde{X}, \mathcal{B}_{\tilde{X}}, \tilde{\mu}, \sigma)$ mit der Faktorabbildung $\pi_0\colon \tilde{X} \longrightarrow X$.

23. **Homomorphismen und Isomorphismen von Maßalgebren** Es seien $(X, \mathcal{S}, \mu)$ und $(Y, \mathcal{S}', \mu')$ Wahrscheinlichkeitsräume und $\phi\colon X \longrightarrow X'$ eine messbare und maßerhaltende Abbildung.
    (a) Für jedes $A \in \mathcal{S}$ bezeichnen wir mit $[A] = \{B \in \mathcal{S} : \mu(A \triangle B) = 0\}$ die Äquivalenzklasse der Menge A. Zeigen Sie, dass für alle $A, B \in \mathcal{S}$ die Operationen $[A] \cup [B] := [A \cup B]$, $[A] \cap [B] := [A \cap B]$ und $[A] \smallsetminus [B] := [A \smallsetminus B]$ wohldefiniert sind und dass die Menge $\mathcal{S}_\mu = \{[A] : A \in \mathcal{S}\}$ mit diesen Operationen die Axiome einer σ-Algebra erfüllt.
    (b) Zeigen Sie, dass die Abbildung $\phi\colon X \longrightarrow X'$ einen Homomorphismus $\Phi\colon \mathcal{S}'_{\mu'} \longrightarrow \mathcal{S}_\mu$ induziert.[12] Falls ϕ invertierbar ist, so ist dieser Homomorphismus ein Isomorphismus.

Anwendungen der Ergodensätze

24. Beweisen Sie, dass eine Folge $(x_n)_{n\geq 0}$ in $\mathbb{T} = [0,1)$ genau dann gleichverteilt ist, wenn $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{I(x_k)} = \lambda(I)$ für jedes Intervall $I \subset [0,1)$ gilt (vgl. Aufgabe 1.36).
25. Sei $\mathsf{P} = (\mathsf{P}_{ij})$ eine strikt positive stochastische $N \times N$-Matrix.
    (1) Zeigen Sie, dass es ein $\gamma < 1$ gibt, sodass $|\sum_{i=0}^{N-1} a_i \mathsf{P}_{ij}| \leq \gamma \sum_{i=1}^{N-1} |a_i| \mathsf{P}_{ij}$ für alle $j = 0, \ldots, N-1$ und alle Vektoren $(a_0, \ldots, a_{N-1}) \in \mathbb{R}^N$ mit $\sum_{i=0}^{N-1} a_i = 0$.
    (2) Sei $\Delta = \{x = (x_0, \ldots, x_{N-1}) \in \mathbb{R}^N : x_i \geq 0 \text{ und } \sum_{i=0}^{N-1} x_1 = 1\}$. Zeigen Sie, dass $\mathsf{P}\colon \Delta \longrightarrow \Delta$ eine Kontraktion ist mit $\|x\mathsf{P} - y\mathsf{P}\|_1 \leq \gamma \|x - y\|$ für alle $x, y \in \Delta$. Schließen Sie daraus, dass es einen eindeutigen stochastischen linken Eigenvektor π von P gibt, für den $\lim_{n\to\infty} \nu \mathsf{P}^n = \pi$ für alle $\nu \in \Delta$ gilt.
    (3) Verallgemeinern Sie die obigen Aufgaben (1) und (2) auf stochastische Matrizen P, für die es ein $n_0 \geq 1$ gibt, sodass P^{n_0} strikt positiv ist.
26. Es seien $\alpha, \beta \in \mathbb{R} \smallsetminus \mathbb{Q}$. Zeigen Sie, dass die Folge $((\alpha n, \beta n^2) \pmod 1)$, $n \geq 1$, gleichverteilt in $\mathbb{T}^2$ ist.
    Hinweis: Für $\alpha = \beta$ ist diese Aussage in Beispiel 2.37 enthalten. Wenn $1, \alpha, \beta$ linear unabhängig über $\mathbb{Q}$ sind, so muss man Beispiel 2.37 verallgemeinern.

---

[12] Das heißt, $\Phi([A] \cup [B]) = \Phi([A]) \cup \Phi([B])$, $\Phi([A] \cap [B]) = \Phi([A]) \cap \Phi([B])$ und $\Phi([A] \smallsetminus [B]) = \Phi([A]) \smallsetminus \Phi([B])$ für alle $A, B \in \mathcal{S}'$.

http://www.springer.com/978-3-0348-0633-6

Dynamische Systeme
Ergodentheorie und topologische Dynamik
Einsiedler, M.; Schmidt, K.
2014, VIII, 159 S. 6 Abb., Softcover
ISBN: 978-3-0348-0633-6
A product of Birkhäuser Basel