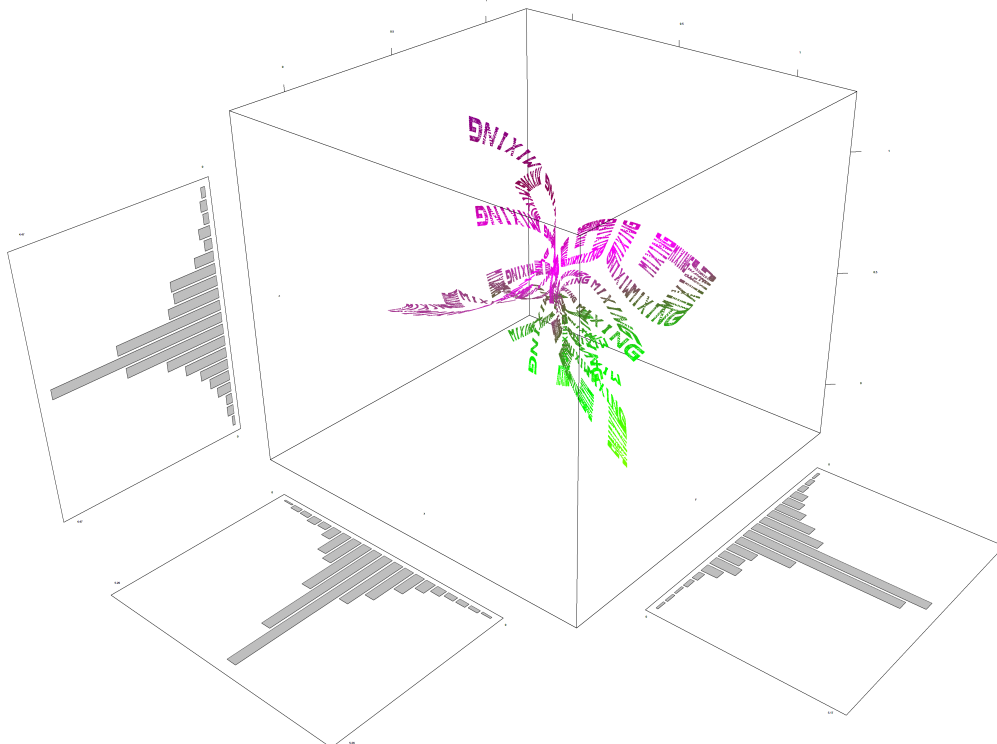


Draft Skriptum (Vers. 2021-03) zur Vorlesung

Einführung in die Ergodentheorie

LV-Nr. 405.514

Dr. Sebastian Fuchs
Univ.-Prof. Dr. Wolfgang Trutschnig
www.trutschnig.net
Fachbereich Mathematik
Paris Lodron Universität Salzburg
Hellbrunner Strasse 34
A-5020 Salzburg



Inhaltsverzeichnis

1	Motivation	3
2	Invariante Maße und maßtreue Transformationen	8
3	Ergodentheorie - Grundlagen	15
3.1	Rekurrenz	16
3.2	Ergodizität	19
3.3	Ergodensatz	25
3.4	Mischungseigenschaften	33
4	Entropie	45
4.1	Zufall und Information	45
4.2	Definition und Eigenschaften	46
4.3	Berechnung der Entropie	51
	Literatur	??

Kapitel 1

Motivation

Unter einem dynamischen System in diskreter Zeit versteht man meist eine Grundmenge Ω zusammen mit einer Transformation $T : \Omega \rightarrow \Omega$. Üblicherweise ist Ω versehen mit einer topologischen/metrischen Struktur - wir schreiben (Ω, d, T) - oder mit einer probabilistischen Struktur - wir schreiben $(\Omega, \mathcal{A}, \mu, T)$. In der Vorlesung werden wir meist Räume Ω mit einer topologischen und einer probabilistischen Struktur betrachten da das Zusammenspiel zahlreiche interessante Zusammenhänge hervorbringt. Für $x \in \Omega$ heißt die Menge

$$O_T(x) := \{x, T(x), T(T(x)), \dots\} = \{T^n(x) : n \in \mathbb{N}_0\}$$

der Orbit von x unter T (oder auch kurz der T -Orbit von x).

Die Ergodentheorie studiert das Langzeitverhalten des dynamischen Systems und versucht dabei insbesondere, folgende Fragen zu beantworten:

- Lässt sich das asymptotische Verhalten spezifischer Punkte $x \in \Omega$ 'prognostizieren'?
Bsp: Konvergiert die Folge $(T^n(x))_{n \in \mathbb{N}}$? Wie ist der Orbit $O_T(x)$ über Ω verteilt?
- Wie verhalten sich 'typische Punkte'?
- Wenn die Punkte $x, y \in \Omega$ nahe beieinander liegen, haben sie dann auch ähnliche Orbits (und damit auch eine ähnliche Verteilung über Ω)?

Wir starten mit einigen einfachen, aber spannenden und teilweise überraschenden Beispielen, die uns durch die Vorlesung begleiten werden.

Beispiel 1.1 (Logistic Map - the route to chaos).

Wir betrachten $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, $\mu = \lambda$, $d = |\cdot|$ (üblicher Abstand in \mathbb{R}) und die sog. logistische Transformation $L_a : [0, 1] \rightarrow [0, 1]$ mit $a \in [0, 4]$, definiert durch

$$L_a(x) = ax(1 - x).$$

- (i) Für $a < 1$ ist das Langzeitverhalten von $L_a^n(x)$ für alle $x \in [0, 1]$ sehr einfach zu bestimmen: In diesem Fall gilt $\lim_{n \rightarrow \infty} L_a^n(x) = 0$ für jedes $x \in [0, 1]$. Für großes n kann der Wert von $L_a^n(x)$ also sehr gut prognostiziert werden.
- (ii) Für den Fall $a > 1$ machen wir uns ein erstes Bild mit Hilfe von R und gehen wie folgt vor: Für jedes $a \in \{1, \frac{1001}{1000}, \frac{1002}{1000}, \dots, 4\}$ ziehen wir eine Zufallsstichprobe x_1, \dots, x_R der

Größe $R = 20$ von $X \sim \mathcal{U}(0,1)$ und berechnen dann für jeden Wert von a die Werte $\{L_a^n(x_1), \dots, L_a^n(x_R)\}$ für $n = 1.000$. Für $a \in [1, 3]$ sieht es so aus als würde $(L_a^n(x))_{n \in \mathbb{N}}$ gegen einen einzigen, von a abhängigen Wert konvergieren, danach verzweigt sich die Kurve in zwei Teile (wir sprechen von einer Bifurkation), dann in vier, usw. Ab $a = 3.6$ ist keinerlei Struktur mehr erkennbar, erst bei ca. 3.83 tritt plötzlich wieder ein 3-punktige Struktur auf, die für größere Werte von a wieder verloren geht.

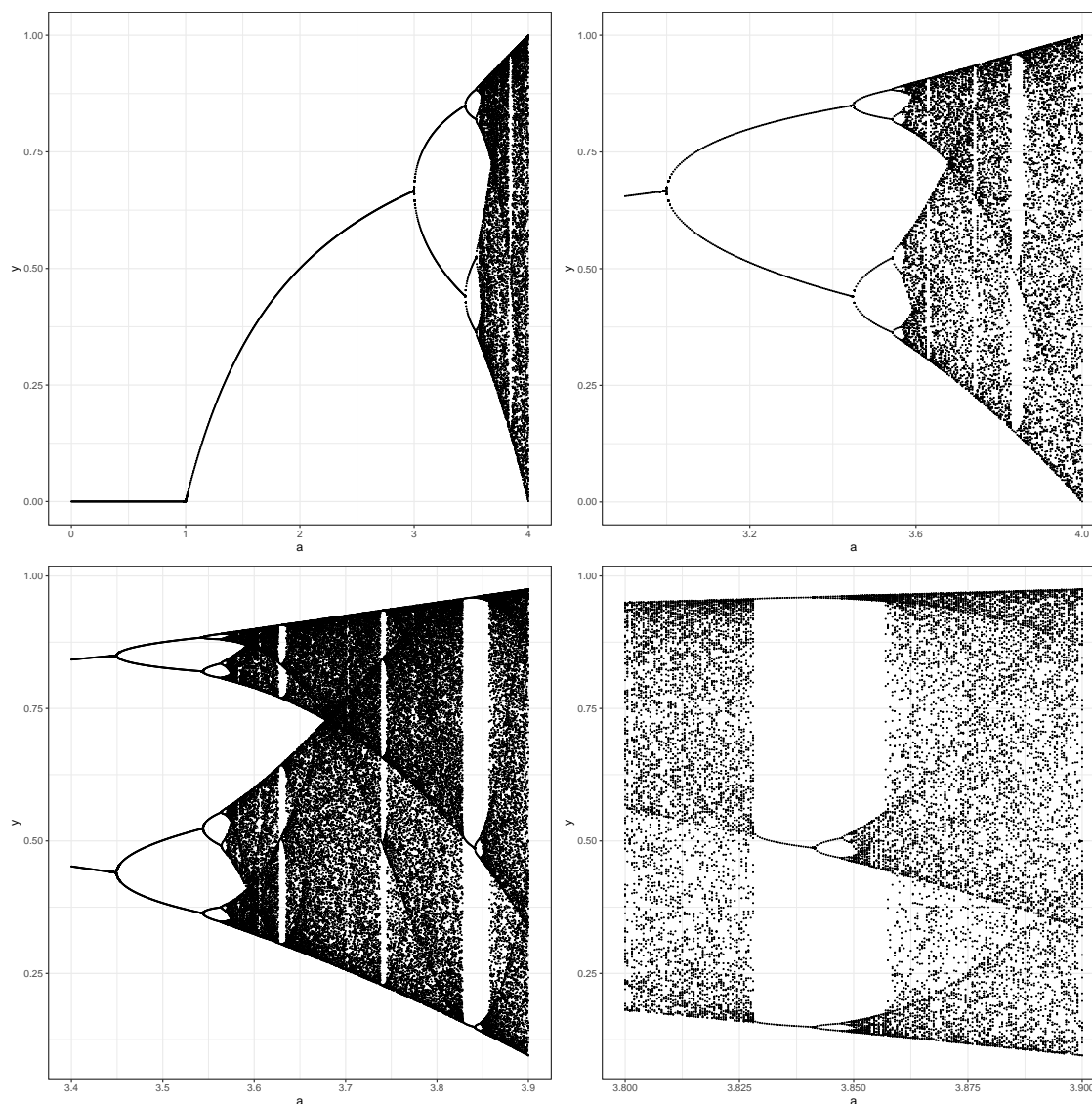


Abbildung 1.1: Die mittels Simulation in Beispiel 1.1 erhaltenen Punkte global und in Teilbereichen; die Bifurkationen und das Fenster für $a = 3.83$ sind klar erkennbar.

- (iii) Der Spezialfall $a = 4$ wird im Laufe der Vorlesung noch eine wichtige Rolle spielen. Für diesen Fall ist auf den ersten Blick keine Struktur erkennbar, es scheint so, also

würden sich die Punkte ‘zufällig’ über $[0, 1]$ verteilen. Wenn schon der Wert nicht prognostizierbar ist - kann zumindest über die Verteilung des Orbits $O_{L_a}(x)$ für ‘typische’ $x \in [0, 1]$ etwas ausgesagt werden?

Für den Moment holen wir uns eine erste Vermutung mittels Simulationen, und gehen wie folgt vor: Wir ziehen eine Zufallsstichprobe x_1, \dots, x_R der Größe $R = 20$ von $X \sim \mathcal{U}(0, 1)$ und berechnen dann für $a = 4$ die Mengen $\{L_a(x_i), L_a^2(x_i), \dots, L_a^n(x_i)\}$ für $n = 1.000.000$. Abbildung 1 zeigt (probability) Histogramme von $\{L_a(x_i), L_a^2(x_i), \dots, L_a^n(x_i)\}$ für $i \in \{1, 2, 3, 4\}$. Folgende zwei Beobachtungen fallen sofort auf:

- Die erhaltenen Histogramme sind nahezu gleich.
- Jeder Punkte scheinen über die Zeit hinweg ‘überall’[†] hinzukommen und besucht am Rand liegende Intervalle öfter als Intervalle in der Mitte.

Das ‘Schicksal’ eines zufällig gezogenen Punkts unter der Abbildung L_4 scheint nicht prognostizierbar zu sein (keine Stabilität oder Konvergenz erkennbar) - die Verteilung über Ω aber anscheinend schon.

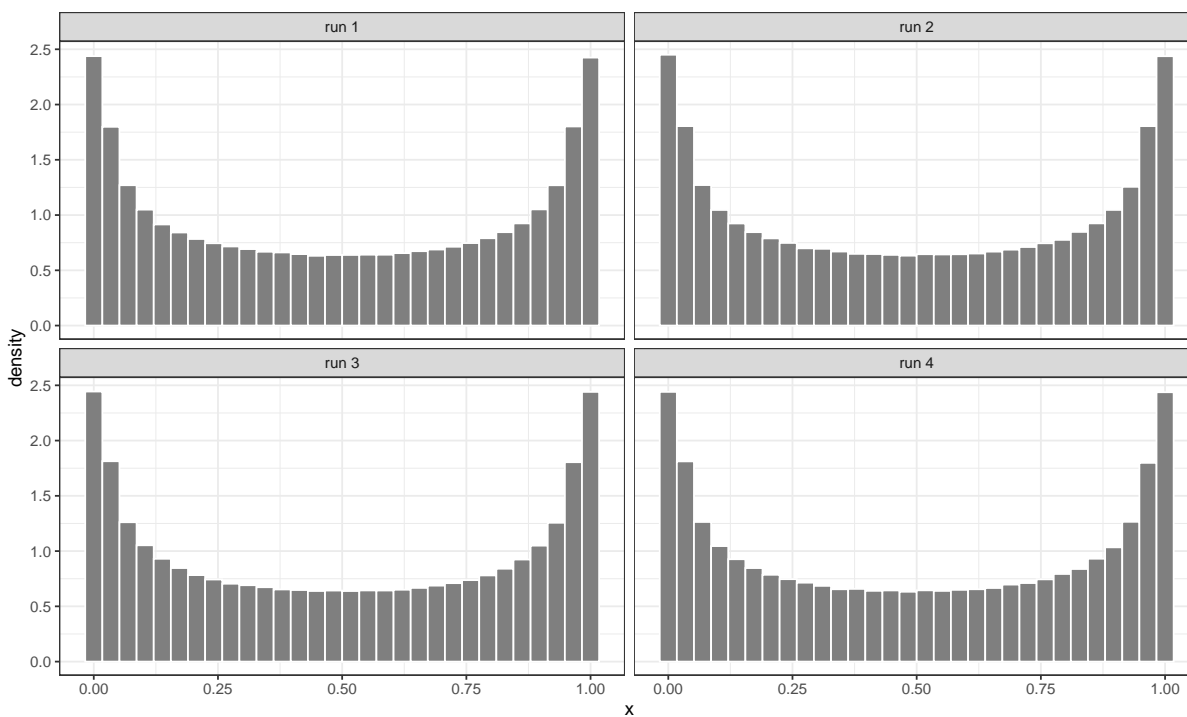


Abbildung 1.2: Iterationen von L_4 : Die in 4 runs erhaltenen Punkte $\{L_a(x_i), L_a^2(x_i), \dots, L_a^n(x_i)\}$ gemäß (iii) in Beispiel 1.1.

[†]wird noch präzisiert

Beispiel 1.2 (Tent Map). Die Zelt-Abbildung (a.k.a. Tent Map) $t : [0, 1] \rightarrow [0, 1]$ ist definiert durch $t(x) = 2x\mathbf{1}_{[0, \frac{1}{2}]}(x) + (2 - 2x)\mathbf{1}_{(\frac{1}{2}, 1]}(x)$. Die Abbildung t ist stückweise expandierend in dem Sinne, dass sie die Intervalle $[0, \frac{1}{2}]$ und $[\frac{1}{2}, 1]$ affin auf ganz $[0, 1]$ abbildet und dabei (innerhalb der Intervalle) den Abstand zweier Punkt verdoppelt. Ob der einfachen Form von

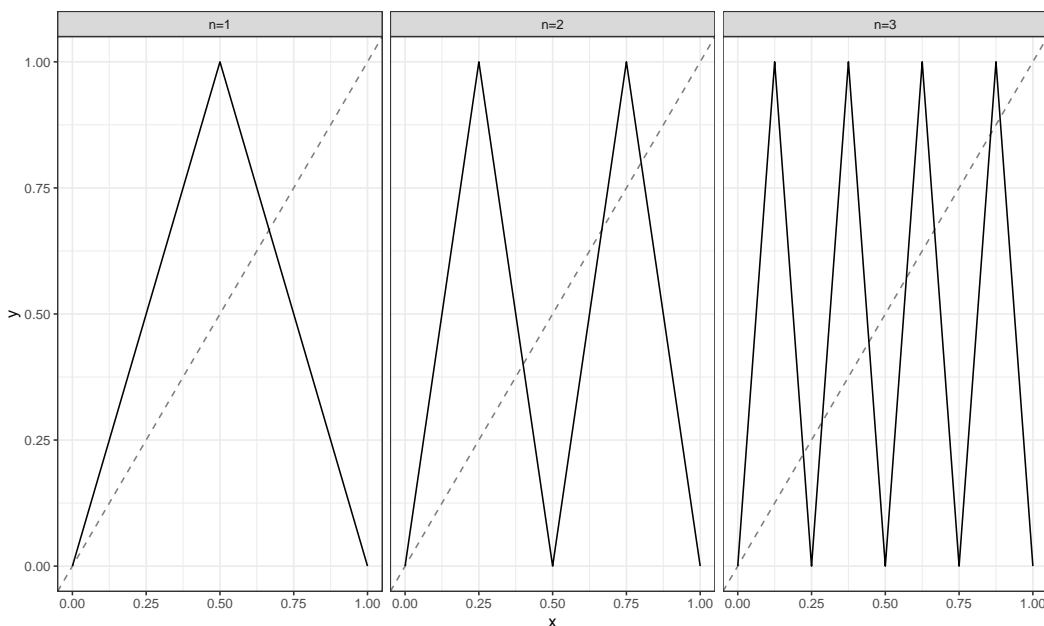


Abbildung 1.3: Die ersten drei Iterationen der Zelt-Abbildung gemäß Beispiel 1.2.

t^n ist leicht zu sehen, dass t genau zwei, t^2 genau 4, und allgemein t^n genau 2^n Fixpunkte hat. Ein Punkt $x \in [0, 1]$ heißt periodischer Punkt von t , genau dann wenn ein $n \in \mathbb{N}$ existiert, sodass $t^n(x) = x$ gilt. Das kleinste $n \in \mathbb{N}$ mit dieser Eigenschaft heißt die Periode von x . $\text{Per}(t)$ wird in der Folge die Menge aller periodischen Punkte von t bezeichnen. Für t lässt sich leicht überprüfen, dass $\text{Per}(t)$ eine abzählbare dichte Teilmenge von $[0, 1]$ ist, i.e., die periodischen Punkte liegen dicht.

Was passiert mit Punkten $x \notin \text{Per}(t)$? Wir werden im Laufe der Vorlesung beweisen, dass ein ‘typischer’ Punkt sich ähnlich verhält wie bei L_4 : er kommt ‘überall’ hin; im Gegensatz zu L_4 besucht er aber (asymptotisch) gleich große Intervalle gleich oft. Mathematisch sauber formuliert: Für jedes Intervall $I \subseteq [0, 1]$ und λ -fast jeden Punkt $x \in [0, 1]$ gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_I(t^i(x)) = \lambda(I). \quad (1.1)$$

Anhand der Zeltabbildung lässt sich weiters illustrieren, warum sog. symbolische dynamische System extrem wichtig in der Ergodentheorie sind: Zerlegen wir das Intervall in $[0, 1] = J_0 \cup J_1$ mit $J_0 = [0, \frac{1}{2}]$, $J_1 = (\frac{1}{2}, 1]$, dann können wir jeden Orbit $O_t(x)$ wiederholt ‘binär’ codieren. Wir schreiben $C_t(x) = \mathbf{a} = (a_0, a_1, a_2, \dots) \in \Sigma_2 = \{0, 1\}^{\mathbb{N}_0}$ genau dann, wenn $t^n(x) \in J_{a_n}$ für jedes $n \in \mathbb{N}_0$ gilt[†]. $C_t(x)$ wird oft als t -Code von x bezeichnet, Σ_2 heißt Codespace mit 2 Symbolen.

[†] $t^0 := id_{[0,1]}$

Die damit induzierte Abbildung $C_t : [0, 1] \rightarrow \Sigma_2$ hat sehr schöne Eigenschaften - die wohl wichtigste ist die folgende: Bezeichnen wir mit $\sigma : \Sigma_2 \rightarrow \Sigma_2$ den Linksshift Operator, definiert durch

$$\sigma((a_0, a_1, a_2, \dots)) = (a_1, a_2, a_3, \dots), \quad (1.2)$$

dann gilt nämlich für $C_t(x) = \mathbf{a}$ offensichtlich

$$C_t(t(x)) = \sigma(\mathbf{a}), \quad (1.3)$$

Anwenden von t entspricht also Anwendung von σ im Codespace. Ist C_t bijektiv oder zumindest injektiv? Ist C_t stetig bezüglich einer ‘passenden’ Metrik ρ_2 auf Σ_2 , ist C_t messbar? Welche Eigenschaften haben die dynamischen Systeme $([0, 1], |\cdot|, t)$ und $(\Sigma_2, \rho_2, \sigma)$ gemeinsam?

Beispiel 1.3 (Nested square roots of 2). Beginn der 2010er Jahren wurde elementar (und relativ mühsam, siehe [1]) bewiesen, dass für jedes $\mathbf{a} = (a_0, a_1, a_2, \dots) \in \{-1, 1\}^{\mathbb{N}_0}$ der Grenzwert von geschachtelten Quadratwurzeln (nested square roots) existiert:

$$\lim_{n \rightarrow \infty} a_0 \sqrt{2 - a_1 \sqrt{2 - a_2 \sqrt{2 - a_3 \sqrt{\dots - a_n \sqrt{2}}}}} \in [-2, 2] \quad (1.4)$$

Wir nennen $\mathbf{a} = (a_0, a_1, a_2, \dots) \in \{-1, 1\}^{\mathbb{N}_0}$ normal, genau dann wenn asymptotische Häufigkeit von 1 und -1 gleich $\frac{1}{2}$ ist, i.e., wenn

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{1\}}(a_i) = \frac{1}{2}$$

gilt. Auch wenn in diesem Fall auf den ersten Blick nicht klar ist, wo hier überhaupt das dynamische System ist, kann mit Hilfe der logistischen Abbildung L_4 , der Zeltabbildung t , und dem Code space Σ_2^\dagger in Kombination mit standard Resultaten der Ergodentheorie die Existenz von eq. (1.4) sehr sehr einfach bewiesen werden. Wie wir in der Vorlesung zeigen werden, ist aber noch viel mehr machbar - es gilt das folgende überraschende Resultat:

Satz 1.4 ([2]). *Für λ -fast jedes $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ gilt: Es existiert ein normales $\mathbf{a} \in \{-1, 1\}^{\mathbb{N}_0}$ mit*

$$2 \sin(x) = \lim_{n \rightarrow \infty} a_0 \sqrt{2 - a_1 \sqrt{2 - a_2 \sqrt{2 - a_3 \sqrt{\dots - a_n \sqrt{2}}}}} \quad (1.5)$$

Im Zuge des Beweises von Satz 1.4 werden wir auch sehen, wie einfach die berühmte Formel

$$\pi = \lim_{n \rightarrow \infty} 2^n \cdot \sqrt{2 - \sqrt{2 + \sqrt{2 + \sqrt{\dots + \sqrt{2}}}}}$$

mit diesen Werkzeugen hergeleitet werden kann.

[†]diese drei dynamischen System haben sehr viel gemeinsam und sind, grob gesagt, dynamisch äquivalent

Kapitel 2

Invariante Maße und maßtreue Transformationen

Die im vorigen Abschnitt kennengelernten Transformationen L_4, t und σ haben viele Gemeinsamkeiten, insbesondere sind Sie *maßtreu* (bezüglich eines oder mehrerer Wahrscheinlichkeitsmaße).

Für eine messbare Transformation $T : (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega, \mathcal{A}, \mu)$ bezeichnen wir das Bildmaß von μ unter T mit μ^T , i.e.

$$\mu^T(A) := \mu(T^{-1}(A)), \quad A \in \mathcal{A}.$$

Ist μ ein Wahrscheinlichkeitsmaß/endlisches Maß, dann offensichtlich auch μ^T .

Definition 2.1 (Masstreue). *Eine messbare Abbildung T auf $(\Omega, \mathcal{A}, \mu)$ heißt maßtreu (präziser: μ -treu) genau dann, wenn $\mu = \mu^T$ gilt. Analog dazu nennen wir ein Maß μ mit $\mu = \mu^T$ invariant bezügl. T (bzw. kurz T -invariant).*

Um zu überprüfen, dass μ invariant bezüglich T ist, reicht es für endliche Maßräume, die Eigenschaft $\mu^T(E) = \mu(E)$ für Mengen aus einem durchschnittsstabilen Erzeuger zu überprüfen - es gilt das folgende Lemma:

Lemma 2.2. *Sei $(\Omega, \mathcal{A}, \mu)$ ein endlicher Maßraum, $T : \Omega \rightarrow \Omega$ messbar und \mathcal{E} ein durchschnittsstabiler Erzeuger von \mathcal{A} . Dann ist T maßtreu genau dann, wenn $\mu^T(E) = \mu(E)$ für jedes $E \in \mathcal{E}$.*

Beweis: Direkte Folgerung aus dem Eindeigkeitssatz aus der Maßtheorie.

Wir werden im Laufe dieses Abschnitt zeigen, dass maßtreue Transformationen auf endlichen Maßräumen (mit denen sich die VO hauptsächlich beschäftigt) automatisch eine relativ starke Rekurrenzeigenschaft haben. Vorher betrachten wir aber einige Beispiele für maßtreue Transformationen.

Bemerkung 2.3. In der Folge werden wir (wie in der Ergodentheorie üblich) statt $T(x)$ auch oft einfach Tx schreiben.

Beispiel 2.4. Sei (Ω, \mathcal{A}) ein beliebiger Messraum und $T : (\Omega, \mathcal{A}) \rightarrow (\Omega, \mathcal{A})$ messbar. Das Dirac Maß δ_z im Punkt $z \in \Omega$, definiert durch

$$\delta_z(A) = \mathbf{1}_A(z)$$

erfüllt offensichtlich

$$\delta_z^T(A) = \delta_z(T^{-1}(A)) = \mathbf{1}_{T^{-1}(A)}(z) = \mathbf{1}_A(Tz) = \delta_{Tz}(A).$$

Für jeden Fixpunkt z von T ist daher δ_z invariant bezüglich T und T ist δ_z -invariant.

Beispiel 2.5. Wir betrachten die Zeltabbildung t auf $(\Omega, \mathcal{A}, \mu) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Für jedes $x \in [0, 1]$ gilt offensichtlich

$$t^{-1}([0, x]) = \{z \in [0, 1] : t(z) \in [0, x]\} = [0, \frac{x}{2}] \cup [1 - \frac{x}{2}, 1],$$

und daher auch

$$\lambda^t([0, x]) = \lambda([0, \frac{x}{2}] \cup [1 - \frac{x}{2}, 1]) = x.$$

Nachdem $\mathcal{E} = \{[0, x] : x \in [0, 1]\}$ ein durchschnittsstabiler Erzeuger von $\mathcal{B}([0, 1])$ ist, ist t also λ -treu.

Beachten Sie, dass es weitere t -invariante Wahrscheinlichkeitsmaße auf $([0, 1], \mathcal{B}([0, 1]))$ gibt: Nachdem 0 und $\frac{2}{3}$ Fixpunkte von t sind, sind sowohl δ_0 als auch $\delta_{\frac{2}{3}}$ invariant bezüglich t .

Beispiel 2.6. Wir betrachten den unendlichen (aber σ -endlichen) Maßraum $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ und die Transformation $Tx = x + 1$. Dann ist T offensichtlich λ -treu (das Lebesgue-Maß ist translationsinvariant). Gibt es weitere bezügl. T invariante Maße? Gibt es bezügl. T invariante Wahrscheinlichkeitsmaße?

Beispiel 2.7. Wir kehren zurück zur logistischen Abbildung L_4 . Abgesehen von den trivialen invarianten Wahrscheinlichkeitsmaßen δ_0 und $\delta_{\frac{3}{4}}$ legt Abbildung 1 nahe, dass es auch ein absolut stetiges L_4 -invariantes Wahrscheinlichkeitsmaß geben sollte (offensichtlich ist λ nicht L_4 -invariant).

Angenommen, das Wahrscheinlichkeitsmaß μ ist L_4 -invariant und absolut stetig (bezügl. λ) mit Dichte f . Wegen

$$L_4^{-1}([0, x]) = [0, \frac{1}{2} - \frac{1}{2}\sqrt{1-x}] \cup [\frac{1}{2} + \frac{1}{2}\sqrt{1-x}, 1]$$

muss μ für jedes $x \in [0, 1]$ die folgende Gleichheit erfüllen:

$$\int_{[0, x]} f d\lambda = \mu([0, x]) = \mu(L_4^{-1}([0, x])) = \int_{[0, \frac{1}{2} - \frac{1}{2}\sqrt{1-x}]} f d\lambda + \int_{[\frac{1}{2} + \frac{1}{2}\sqrt{1-x}, 1]} f d\lambda \quad (2.1)$$

Für λ -fast jedes $x \in [0, 1]$ muss also die folgende Funktionalgleichung für f gelten (Differentiation beider Seiten nach x unter Verwendung der Kettenregel):

$$f(x) = \frac{1}{4\sqrt{1-x}} \left(f\left(\frac{1}{2} - \frac{1}{2}\sqrt{1-x}\right) + f\left(\frac{1}{2} + \frac{1}{2}\sqrt{1-x}\right) \right) \quad (2.2)$$

Welche Dichte f erfüllt Gleichung (2.2)? Die Antwort auf diese Frage wurde 1947 von Ulam und von Neumann gegeben (siehe [5]), die zeigen konnten, dass die Funktion f_* , gegeben durch

$$f_*(x) = \frac{1}{\pi\sqrt{x(1-x)}} \quad (2.3)$$

Gleichung (2.2) erfüllt. Diese Tatsache ist analytisch leicht nachzurechnen, wesentlich spannender ist aber, wie man auf diese f_* kommt - wir wollen diese Frage nun beantworten (und

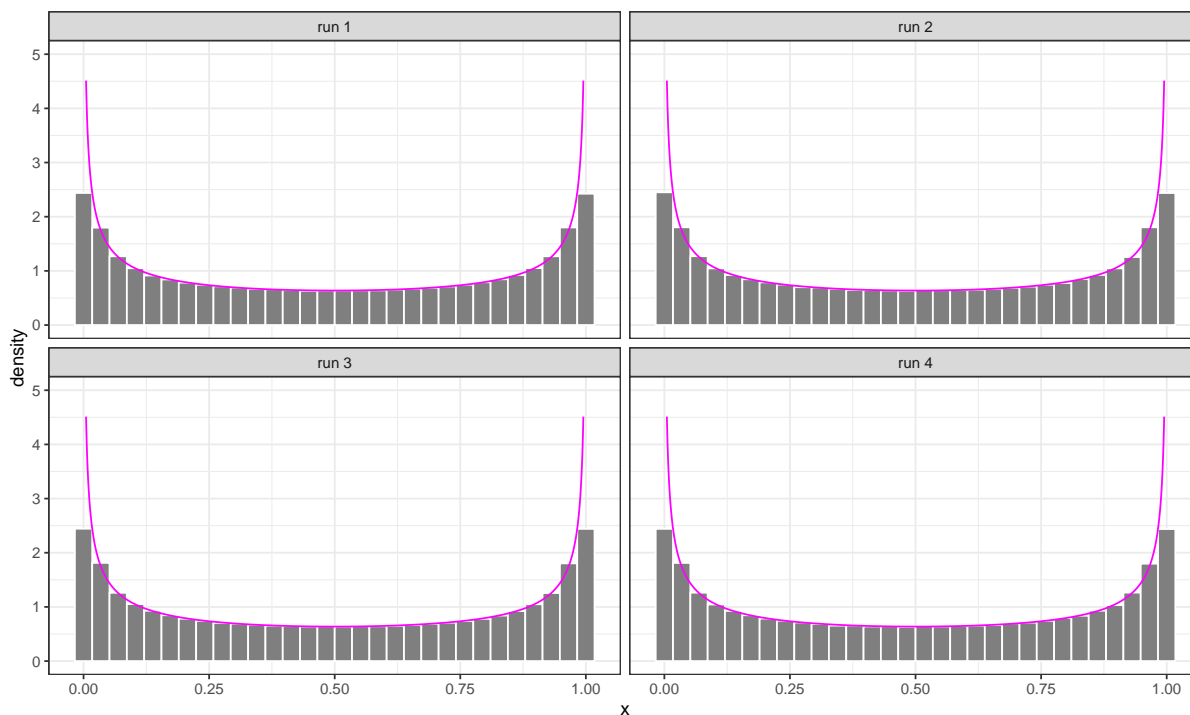


Abbildung 2.1: Iterationen von L_4 : Die in 4 runs erhaltenen Punkte $\{L_a(x_i), L_a^2(x_i), \dots, L_a^n(x_i)\}$ gemäß (iii) in Beispiel 1.1 zusammen mit der Ulam von Neumann Dichte f_* gemäß Gleichung (2.3)

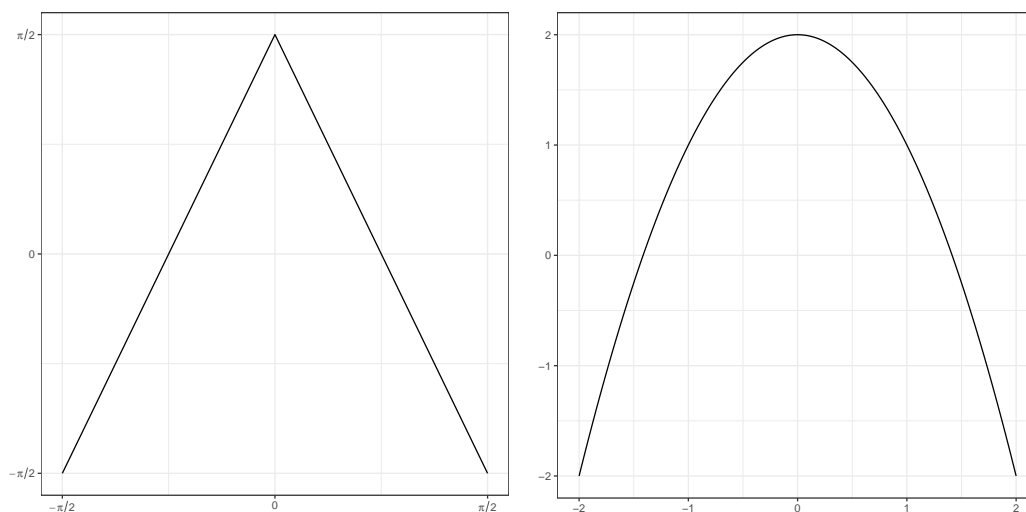


Abbildung 2.2: Die Funktionen \hat{t} und \hat{L}_4 .

motivieren damit nicht zuletzt die Verwendung von topologischen Konjugationen und Isomorphismen zur Analyse dynamischer System, die später noch genauer behandelt wird). Abbildung

2.7 zeigt f_* zusammen mit den in Beispiel 1.1 diskutierten Histogrammen. Zur (eleganten) Herleitung verwenden wir [2] und gehen wie folgt vor: (S1) Wir betrachten skalierte Versionen von t und L_4 , (S2) zeigen, dass sich die Transformation ineinander ‘übersetzen’ lassen, und (S3) leiten aus dieser Tatsache und $\lambda^t = \lambda$ dann f_* ab.

(S1) Definiere $\hat{t} : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$ und $\hat{L}_4 : [-2, 2] \rightarrow [-2, 2]$ durch (siehe Abbildung 2.7)

$$\hat{t}(x) = \begin{cases} 2x + \frac{\pi}{2} & \text{if } x \in [-\frac{\pi}{2}, 0] \\ -2x + \frac{\pi}{2} & \text{if } x \in [0, \frac{\pi}{2}], \end{cases}$$

and $\hat{L}_4(x) = -x^2 + 2$.

(S2) Definieren wir $h : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-2, 2]$ durch $h(x) = 2 \sin(x)$, dann ist h offensichtlich ein Homöomorphismus (also bijektiv und stetig in beide Richtungen). Zusätzlich ist das folgende Diagramm kommutativ, i.e., es gilt $h \circ \hat{t} = \hat{L}_4 \circ h$ (direktes Nachrechnen unter Verwendung der Additionstheoreme für Sinus und Cosinus)

$$\begin{array}{ccc} [-\frac{\pi}{2}, \frac{\pi}{2}] & \xrightarrow{\hat{t}} & [-\frac{\pi}{2}, \frac{\pi}{2}] \\ h \downarrow & & \downarrow h \\ [-2, 2] & \xrightarrow{\hat{L}_4} & [-2, 2] \end{array} \quad (2.4)$$

Anders ausgedrückt, für jedes $x \in [-2, 2]$ gilt die Gleichheit

$$h \circ \hat{t} \circ h^{-1}(x) = \hat{L}_4(x) \quad (2.5)$$

und für jedes $z \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ die Gleichheit

$$\hat{t}(z) = h^{-1} \circ \hat{L}_4 \circ h(z), \quad (2.6)$$

\hat{L}_4 ist also (in der Sprache der Linearen Algebra) \hat{t} in einem anderen ‘Koordinatensystem’ und vice versa. Wir nennen daher \hat{t} und \hat{L}_4 *topologisch konjugiert* (also topologisch gleich). Beachten Sie, dass aus Gleichungen (2.5) und (2.6) sofort $h \circ \hat{t}^n \circ h^{-1} = \hat{L}_4^n$ und $\hat{t}^n = h^{-1} \circ \hat{L}_4^n \circ h$ für jedes $n \in \mathbb{N}$ folgt, auch die Orbits von Punkte können also ineinander übersetzt werden.

Zusätzlich ist folgende wichtige Eigenschaft leicht nachzurechnen: Für jedes Maß μ , das \hat{t} invariant ist, ist μ^h invariant bezügl. \hat{L}_4 :

$$(\mu^h)^{\hat{L}_4} = \mu^{\hat{L}_4 \circ h} = \mu^{h \circ \hat{t}} = (\mu^{\hat{t}})^h = \mu^h$$

Wir wissen, dass $\mu := \frac{1}{\pi} \lambda$ invariant bezügl. \hat{t} ist, das Wahrscheinlichkeitsmaß μ^h muss daher invariant bezügl. \hat{L}_4 sein.

(S3) Wir berechnen μ^h und dessen Dichte: Für $x \in [-2, 2]$ gilt offensichtlich

$$\mu^h([-2, x]) = \mu(h^{-1}([-2, x])) = \mu\left(\left[\frac{\pi}{2}, \arcsin\left(\frac{x}{2}\right)\right]\right) = \frac{1}{\pi} \int_{\frac{\pi}{2}, \arcsin\left(\frac{x}{2}\right)} 1 d\lambda,$$

Anwendung des Hauptsatzes der Differential- und Integralrechnung liefert daher sofort, dass die Dichte f von μ^h gegeben ist durch

$$f(x) = \frac{d}{dx} \mu^h([-2, x]) = \frac{1}{\pi} \frac{1}{\sqrt{1 - \frac{x^2}{4}}} \frac{1}{2} = \frac{1}{\pi \sqrt{4 - x^2}}.$$

f sieht f_* schon verdächtig ähnlich: tatsächlich erhalten wir f_* direkt aus f durch Übertragung von $[-2, 2]$ auf $[0, 1]$: Definieren wir $\varphi : [-2, 2] \rightarrow [0, 1]$ durch $\varphi(x) = \frac{x}{4} + \frac{1}{2}$ und damit $\varphi^{-1}(y) = 4y - 2$, dann erhalten wir sofort

$$(\mu^h)^\varphi([0, x]) = \mu^h(\varphi^{-1}([0, x])) = \mu^h([-2, 4x - 2]) = \frac{1}{\pi} \int_{[-2, 4x-2]} \frac{1}{\sqrt{4 - z^2}} d\lambda(z),$$

woraus sofort (Differentiation der Stammfunktion plus Kettenregel)

$$\frac{d}{dx} (\mu^h)^\varphi([0, x]) = \frac{1}{\pi \sqrt{4 - (4x - 2)^2}} \cdot 4 = \frac{1}{\pi \sqrt{4 - 16x^2 + 16x - 4}} \cdot 4 = \frac{1}{\pi \sqrt{x(1 - x)}} = f_*(x)$$

folgt.

Die für uns in der Vorlesung wichtigste maßtreue Transformation ist der Shift-Operator σ , den wir schon im ersten Kapitel kennengelernt haben. Wir definieren daher als nächsten Schritt eine spezielle Metrik sowie invariante Maße auf dem Code space.

Definition 2.8 (Code space Σ_N). Für $2 \leq N \in \mathbb{N}$ heißt die Menge $\Sigma_N := \{1, 2, \dots, N\}^{\mathbb{N}_0}$ der Code space mit N Symbolen. Für $\mathbf{k} \in \Sigma_N$ schreiben wir $\mathbf{k} = (k_0, k_1, k_2, \dots)$.

Auf Σ_N können verschiedene Metriken definiert werden, die gebräuchlichste ist die folgende Metrik ρ , definiert durch

$$\rho(\mathbf{k}, \mathbf{l}) = \begin{cases} 0 & \text{für } \mathbf{k} = \mathbf{l} \\ 2^{-\min\{i \in \mathbb{N}_0 : k_i \neq l_i\}} & \text{für } \mathbf{k} \neq \mathbf{l}. \end{cases} \quad (2.7)$$

Ebenfalls gerne verwendet wird die ‘klassische’ Produktmetrik δ , definiert durch

$$\delta(\mathbf{k}, \mathbf{l}) = \sum_{i=0}^{\infty} \frac{|k_i - l_i|}{2^i}. \quad (2.8)$$

Es lässt sich unschwer nachrechnen, dass sowohl ρ als auch δ Metriken auf Σ_N (Übungsaufgabe) sind. Der folgende Satz fasst die wichtigsten Eigenschaften von ρ und δ zusammen:

Satz 2.9. (Σ_N, ρ) ist ein ultrametrischer Raum, i.e., ρ ist eine Metrik, die zusätzlich für alle $\mathbf{k}, \mathbf{l}, \mathbf{m} \in \Sigma_N$

$$\rho(\mathbf{k}, \mathbf{l}) \leq \max\{\rho(\mathbf{k}, \mathbf{m}), \rho(\mathbf{m}, \mathbf{l})\} \quad (2.9)$$

erfüllt. Zusätzlich sind ρ und δ Metrisierungen der Produkttopologie, i.e., die folgenden drei Aussagen sind äquivalent für alle $\mathbf{k}, \mathbf{k}^1, \mathbf{k}^2, \dots \in \Sigma_N$,

1. $\lim_{n \rightarrow \infty} \rho(\mathbf{k}^n, \mathbf{k}) = 0$
2. $\lim_{n \rightarrow \infty} \delta(\mathbf{k}^n, \mathbf{k}) = 0$
3. $\lim_{n \rightarrow \infty} |k_i^n - k_i| = 0$ für jedes $i \in \mathbb{N}_0$,

und die metrischen Räume (Σ_N, ρ) und (Σ_N, δ) sind kompakt.

Beweis: Wir geben nur eine Beweisskizze, die Ausführung der Details ist eine nützliche Übungsaufgabe.

(i) Ungleichung (2.9) ist eine direkte Konsequenz aus folgender Beobachtung für den Fall $\mathbf{k} \neq \mathbf{l}$ und $\mathbf{k} \neq \mathbf{m} \neq \mathbf{l}$: Sei $i_0 := \min\{i \in \mathbb{N}_0 : k_i \neq l_i\}$. Setzen wir

$$j_0 := \min\{\min\{j \in \mathbb{N}_0 : k_j \neq m_j\}, \min\{j \in \mathbb{N}_0 : m_j \neq l_j\},$$

dann folgt sofort $k_j = m_j = l_j$ für alle $j < j_0$ und damit $j_0 \leq i_0$.

(ii) Angenommen, die dritte der obigen Aussagen gilt. Weiters sei $\varepsilon > 0$ beliebig aber fest. Wähle j so, dass $2^{-j} < \varepsilon$ gilt. Für $i = 0$ existiert ein Index $n_0 \in \mathbb{N}$ mit $k_i^n = k_i$ für alle $n \geq n_0$; Für $i = 1$ existiert ein Index $n_1 \in \mathbb{N}$ mit $n_0 \leq n_1$ mit $k_i^n = k_i$ für alle $n \geq n_1$ nach endlich vielen Schritten erhalten wir $n_j \in \mathbb{N}$ so, dass $k_i^n = k_i$ für all $n \geq n_j$ und $i \leq j$. Mit anderen Worten, für $n \geq n_j$ gilt $\rho(\mathbf{k}^n, \mathbf{k}) < 2^{-j} < \varepsilon$. Nachdem $\varepsilon > 0$ beliebig war, haben wir also $\lim_{n \rightarrow \infty} \rho(\mathbf{k}^n, \mathbf{k}) = 0$ gezeigt.

(iii) Gelte $\lim_{n \rightarrow \infty} \rho(\mathbf{k}^n, \mathbf{k}) = 0$, weiters sei $i \in \mathbb{N}_0$ beliebig aber fest. Für $\varepsilon < 2^{-i}$ existiert ein Index $n_0 \in \mathbb{N}$, sodass $\rho(\mathbf{k}^n, \mathbf{k}) < \varepsilon < 2^{-i}$ und damit $k_i^n = k_i$ für jedes $n \geq n_0$, i.e., $\lim_{n \rightarrow \infty} |k_i^n - k_i| = 0$. Nachdem i beliebig war, folgt die dritte Aussage, und wir haben die Äquivalent der ersten und der dritten Aussagen des Satzes. Die verbleibende Äquivalenz kann analog gezeigt werden.

(iv) Die Kompaktheit von (Σ_N, ρ) folgt direkt aus der Kompaktheit von $\{1, \dots, N\}$ und dem Satz von Tychonov. Alternativ kann die Kompaktheit auch unschwer zu Fuß nachgerechnet werden indem man für jede Folge $(\mathbf{k}^n)_{n \in \mathbb{N}}$ die Existenz einer konvergenten Teilfolge zeigt.

Ultrametrische Räume haben sehr spezielle Eigenschaften, insbesondere die folgende:

- Für $\mathbf{k} \in B(\mathbf{l}, r) = \{\mathbf{m} \in \Sigma_N : \rho(\mathbf{l}, \mathbf{m}) < r\}$ gilt $B(\mathbf{l}, r) = B(\mathbf{k}, r)$.
Mit anderen Worten: Jeder Punkt in einer offenen Kugel ist Mittelpunkt der Kugel.

(Ω, ρ) hat außerdem die folgenden ungewohnten, aber nützlichen Eigenschaften:

- Die folgende Identität gilt für alle $\mathbf{k} \in \Sigma_N$ und $i_0 \in \mathbb{N}_0$:

$$\begin{aligned} B(\mathbf{k}, 2^{-i_0}) &= \{\mathbf{l} \in \Sigma_N : \rho(\mathbf{k}, \mathbf{l}) < 2^{-i_0}\} \\ &= \{\mathbf{l} \in \Sigma_N : k_i = l_i \text{ für jedes } i \leq i_0\} \\ &= \{k_0\} \times \{k_1\} \times \dots \times \{k_{i_0}\} \times \{1, \dots, N\} \times \dots \times \{1, \dots, N\} \times \dots \end{aligned}$$

Mit anderen Worten: Offene Kugeln sind ‘dünne’ Rechtecke.

- Nachdem die Menge $R = \{k_0\} \times \{k_1\} \times \dots \times \{k_{i_0}\} \times \{1, \dots, N\} \times \dots \times \{1, \dots, N\} \times \dots$ offensichtlich abgeschlossen ist, ist damit jede Kugel in (Σ_N, ρ) offen und gleichzeitig abgeschlossen (‘clopen’...closed and open).

Basierend auf der von ρ erzeugten Topologie auf Σ_N können wir die Borel’sche σ -Algebra $\mathcal{B}(\Sigma_N)$ auf Σ_N betrachten. Das folgende (hoffentlich aus der Maßtheorie bekannte) Lemma behandelt einen einfachen Erzeuger \mathfrak{h}_N von $\mathcal{B}(\Sigma_N)$, der sich in der Folge als sehr nützlich erweisen wird:

Lemma 2.10. *Definieren wir das Mengensystem \mathfrak{h}_N durch*

$$\mathfrak{h}_N := \left\{ \{k_0\} \times \{k_1\} \times \cdots \times \{k_n\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \cdots : n \in \mathbb{N}_0 \text{ and } k_0, k_1, \dots, k_n \in \{1, \dots, N\} \right\},$$

dann ist \mathfrak{h}_N ein Halbring (a.k.a. Halbring der dünnen Rechtecke) und es gilt $\mathcal{A}_\sigma(\mathfrak{h}_N) = \mathcal{B}(\Sigma_N)$.

Wir konstruieren nun ausgehend von einem Wahrscheinlichkeitsmaß ϑ auf $\{1, \dots, N\}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\Sigma_N)$ - der Konstruktion liegt die folgende einfache Idee zugrunde: Wir ziehen unendlich oft gemäß ϑ zufällig ein Element aus $\{1, \dots, N\}$ (mit Zurücklegen), die (diskrete) Zufallsvariable Z_i bezeichne das Ergebnis der i -ten Ziehung. Offensichtlich gilt dann für jedes $i \in \mathbb{N}_0^\dagger$ und für $s \in \{1, \dots, N\}$

$$\mathbb{P}(Z_i = s) = \vartheta(\{s\}) =: \vartheta_s$$

und die Folge $(Z_i)_{i \in \mathbb{N}_0}$ ist i.i.d. Für jedes dünne Rechteck $R = \{k_0\} \times \{k_1\} \times \cdots \times \{k_n\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \cdots$ definieren wir μ_ϑ durch

$$\mu_\vartheta(R) = \mathbb{P}(Z_0 = k_0, Z_1 = k_1, \dots, Z_n = k_n)$$

und erhalten damit wegen der Unabhängigkeit von $(Z_i)_{i \in \mathbb{N}_0}$ sofort

$$\mu_\vartheta(R) = \prod_{i=0}^n \vartheta(\{k_i\}) = \prod_{i=0}^n \vartheta_{k_i}. \quad (2.10)$$

Damit ist eine Mengenfunktion $\mu_\vartheta : \mathfrak{h}_N \rightarrow [0, 1]$ definiert, von der sich (mehr oder weniger unschwer) zeigen lässt, dass sie σ -additiv auf \mathfrak{h}_N ist, und damit eindeutig zu einem Wahrscheinlichkeitsmaß auf ganz $\mathcal{A}_\sigma(\mathfrak{h}_N) = \mathcal{B}(\Sigma_N)$ fortgesetzt werden kann. Das resultierende Wahrscheinlichkeitsmaß μ_ϑ heisst das von ϑ auf Σ_N induzierte Produktmaß.

Satz 2.11. *Sei ϑ ein beliebiges Wahrscheinlichkeitsmaß auf $\{1, \dots, N\}$, μ_ϑ bezeichne das von ϑ induzierte Produktmaß auf $\mathcal{B}(\Sigma_N)$. Dann ist der Linksshift Operator σ maßtreu auf $(\Sigma_N, \mathcal{B}(\Sigma_N), \mu_\vartheta)$.*

Beweis: Gemäß Lemma 2.2 reicht es zu zeigen, dass für jedes dünne Rechteck R die Gleichheit $\mu_\vartheta^\sigma(R) = \mu_\vartheta(R)$ gilt. Selbiges folgt für $R = \{k_0\} \times \{k_1\} \times \cdots \times \{k_n\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \cdots$ aber sofort aus[†]

$$\begin{aligned} \mu_\vartheta^\sigma(R) &= \mu_\vartheta(\sigma^{-1}(R)) = \mu_\vartheta(\{\mathbf{k} \in \Sigma_N : \sigma(\mathbf{k}) \in R\}) = \mu_\vartheta\left(\bigcup_{l=1}^N \{l\} \times R\right) \\ &= \sum_{l=1}^N \mu_\vartheta(\{l\} \times R) = \sum_{l=1}^N \vartheta(\{l\}) \cdot \mu_\vartheta(R) = \mu_\vartheta(R). \end{aligned}$$

[†]damit die Konstruktion durchgeht nennen wir die erste Ziehung die Ziehung 0.

[†]wir schreiben kurz $\{l\} \times R = \{(l, k_0, k_1, k_2, \dots) : (k_0, k_1, k_2, \dots) \in R\}$

Kapitel 3

Ergodentheorie - Grundlagen

Wie wir bereits gesehen haben, können auch sehr einfach strukturierte Systeme ein sehr komplexes dynamisches Verhalten aufweisen. Es ist daher nur eingeschränkt möglich weitreichende Vorhersagen zu treffen. Eine Möglichkeit bestände klarerweise darin den Orbit einer Funktion an jedem beliebigen Punkt auszuwerten.

Anstatt sich auf einzelne Orbits zu konzentrieren, werden wir in dieser Vorlesung zu einer wahrscheinlichkeitstheoretischen Beschreibung dynamischer Systeme übergehen, die üblicherweise als *Ergodentheorie* bezeichnet wird. Auch wenn das Verhalten komplizierter Systeme vollständig deterministisch abläuft, kann das Langzeitverhalten am besten mit stochastischen Begriffen beschrieben werden.

Um das Verhalten einer Abbildung $T : \Omega \rightarrow \Omega$ auf einer Menge Ω zu analysieren, müssen wir zunächst entscheiden welche Mengen von Ω groß sind, welche klein sind und wie sich diese Größe entwickelt. Wir statten daher Ω mit einer probabilistischen Struktur aus - im Folgenden bezeichne $(\Omega, \mathcal{A}, \mu)$ den zugrundeliegenden Maßraum. Wir werden uns hauptsächlich mit endlichen Maßräumen und da im Speziellen mit Wahrscheinlichkeitsräumen beschäftigen. Interpretationen des Maßraums:

- (i) deterministische Interpretation: Wir denken uns μ als eine Verteilung der Masse auf Ω , wobei die Masse des gesamten Raums auf eins normiert wurde. Bei einer Menge $A \subseteq \Omega$ betrachten wir $\mu(A)$ als die Masse von A , genauer gesagt, den Prozentsatz der Gesamtmasse, der auf A konzentriert ist.
- (ii) stochastische Interpretation: Wir betrachten Ω als die Menge aller möglichen Ergebnisse eines Experiments, das dem Zufall unterliegt. Gegeben $A \in \mathcal{A}$, ist die Größe $\mu(A)$ einfach die Wahrscheinlichkeit des Experiments, ein Ergebnis in A zu produzieren.

Wir möchten die Abbildung $T : \Omega \rightarrow \Omega$ gern in die messbare Struktur von Ω einbinden um die Hintereinanderausführung von T als ein neues Experiment mit Werten in Ω interpretieren zu können. Um dies zu gewährleisten muss für jedes Ereignis $A \in \mathcal{A}$ auch $T^{-1}(A)$ ein Ereignis, d.h. $\in \mathcal{A}$, sein; mit anderen Worten: T muss eine messbare Abbildung sein, und damit $T^{-1}(A) \subseteq \mathcal{A}$ gelten. Eine messbare Abbildung auf einem Maßraum bewahrt dessen Struktur genauso wie eine lineare Abbildung auf einem linearen Raum oder eine stetige Abbildung auf einem topologischen Raum.

Interpretieren wir μ als Massenverteilung auf Ω , dann beschreibt das Bildmaß μ^T die Massenverteilung (die probabilistische Struktur) auf dem Bild $T(\Omega)$. Ist die Abbildung maßtreu, dann sind die Massenverteilungen auf Ω und $T(\Omega)$ identisch.

3.1 Rekurrenz

In diesem Abschnitt beschäftigen wir uns mit dem dynamischen Verhalten von Abbildungen und interessieren uns dafür unter welchen Umständen ein Punkt x aus einer Menge A - nach (einem eventuellen) Verlassen dieser Menge - wieder nach A zurückkehrt. Diese Eigenschaft heißt Rekurrenz. Wir werden zeigen, dass auf einem endlichen Maßraum die Eigenschaft Maßtreue bereits hinreichend für Rekurrenz ist (Satz 3.4).

Definition 3.1. (Rekurrenz) Eine maßtreue Abbildung T auf dem Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt rekurrent, falls für jede Menge $A \in \mathcal{A}$ mit $\mu(A) > 0$ eine Nullmenge $N \subseteq A$ existiert so, dass für jedes $x \in A \setminus N$ ein $n \in \mathbb{N}$ existiert mit $T^n(x) \in A$.

Mit anderen Worten: T heißt rekurrent, wenn für jede Menge A mit positivem Maß fast jeder Punkt aus A nach endlicher Zeit wieder nach A zurückkehrt (zur Illustration siehe Abbildung 3.1). n kann als Rückkehrzeitpunkt interpretiert werden.

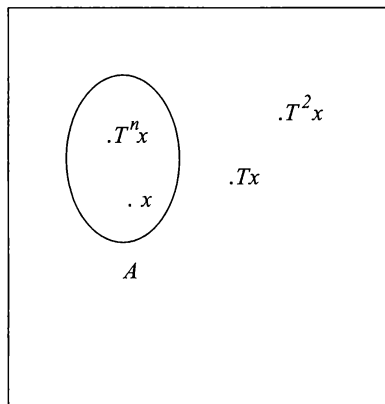


Abbildung 3.1: Illustration Rekurrenz

Wir zeigen zunächst, dass für jede rekurrente Abbildung unendliche viele Rückkehrzeitpunkte existieren.

Lemma 3.2. Sei T eine rekurrente und maßtreue Abbildung auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann gibt es für jede Menge $A \in \mathcal{A}$ mit $\mu(A) > 0$ eine Nullmenge $N \subseteq A$ so, dass für jedes $x \in A \setminus N$ eine monoton wachsende Folge $(n_k)_{k \in \mathbb{N}}$ existiert mit $T^{n_k}(x) \in A$ für alle $k \in \mathbb{N}$.

Beweis: Nach Definition existiert eine Nullmenge N_1 so, dass für jedes $y \in A \setminus N_1$ ein $n(y) \in \mathbb{N}$ existiert mit $T^{n(y)}(y) \in A$. Wähle nun $N = \bigcup_{k \in \mathbb{N}} T^{-k}(N_1)$. Dann ist wegen

$$\mu(N) = \mu \left(\bigcup_{k \in \mathbb{N}} T^{-k}(N_1) \right) \leq \sum_{k \in \mathbb{N}} \mu(T^{-k}(N_1)) = \sum_{k \in \mathbb{N}} \mu(N_1) = 0$$

auch N eine Nullmenge (die für uns interessante). Wegen $T^{-k}(N) \subseteq N$ folgt weiter mit $y \notin N$ auch $T^k(y) \notin N$. Wählt man also $x \in A \setminus N \subseteq A \setminus N_1$, erhält man $T^{n(x)}(x) \in A \setminus N$. Setze $m_1 := n(x)$.

Wendet man das eben gezeigte auf den Punkt $z = T^{m_1}(x) \in A \setminus N$ an, dann existiert wiederum ein $n_1 = n_1(z) \in \mathbb{N}$ mit $T^{n_1}(z) \in A \setminus N$. Mit $m_2 := m_1 + n_1$ folgt schließlich $T^{m_2}(x) = T^{n_1}(z) \in A \setminus N$. Induktiv erhält man nun die Behauptung.

Das nun folgende Resultat ermöglicht eine anschauliche Charakterisierung von Rekurrenz.

Lemma 3.3. *Sei T eine maßtreue Abbildung auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann sind folgende Aussagen äquivalent:*

(a) T ist rekurrent.

(b) Für jede Menge $A \in \mathcal{A}$ mit $\mu(A) > 0$ existiert ein $n \in \mathbb{N}$ mit $\mu(A \cap T^{-n}(A)) > 0$.

Beweis: Für $A \in \mathcal{A}$ mit $\mu(A) > 0$ definieren wir die Menge A^* aller Punkte in A , die A sofort und für immer verlassen, d.h.

$$A^* = A \setminus \bigcup_{n \in \mathbb{N}} T^{-n}(A) = A \cap \bigcap_{n \in \mathbb{N}} T^{-n}(A^c)$$

Offensichtlich ist T genau dann rekurrent, wenn $\mu(A^*) = 0$ für alle $A \in \mathcal{A}$ mit $\mu(A) > 0$.

(a) \Rightarrow (b): Sei nun T rekurrent und $A \in \mathcal{A}$ mit $\mu(A) > 0$. Dann gilt

$$\mu \left(A \setminus \bigcup_{n \in \mathbb{N}} (A \cap T^{-n}(A)) \right) = \mu \left(A \setminus \bigcup_{n \in \mathbb{N}} T^{-n}(A) \right) = \mu(A^*) = 0$$

Wegen $\mu(A) > 0$ existiert nun ein $n \in \mathbb{N}$ mit $\mu(A \cap T^{-n}(A)) > 0$.

(b) \Rightarrow (a): Wir zeigen nun (b) nach (a) durch Widerspruch. Dazu nehmen wir an, dass $\mu(A^*) > 0$ gilt (d.h. T ist nicht rekurrent). Dann existiert nach (b) ein $n \in \mathbb{N}$ mit

$$\mu(A^* \cap T^{-n}(A^*)) > 0$$

woraus $A^* \cap T^{-n}(A^*) \neq \emptyset$ folgt. Somit existiert ein Punkt $x \in A^* \subseteq A$ mit $T^n(x) \in A^* \subseteq A$. Dies steht in Widerspruch zur Definition von A^* .

Die Eigenschaft (b) in Lemma 3.3 wird auch *konservativ* genannt.

Das folgende Resultat stammt aus dem Jahr 1899 und gilt als erster Satz der Ergodentheorie.

Satz 3.4. (*Poincaré*)

Sei T eine maßtreue Abbildung auf einem endlichen Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann ist T rekurrent.

Beweis: Wegen Lemma 3.3 reicht es zu zeigen, dass für jede Menge $A \in \mathcal{A}$ mit $\mu(A) > 0$ ein $n \in \mathbb{N}$ existiert mit $\mu(A \cap T^{-n}(A)) > 0$.

Im Folgenden betrachten wir eine Menge $A \in \mathcal{A}$ mit $\mu(A) > 0$ und nehmen das Gegenteil an, d.h. $\mu(A \cap T^{-n}(A)) = 0$ für alle $n \in \mathbb{N}$. Dann gilt für alle $i \neq j$ (wir schreiben $j = i + k$, $k > 0$)

$$\begin{aligned} \mu(T^{-i}(A) \cap T^{-j}(A)) &= \mu(T^{-i}(A) \cap T^{-(i+k)}(A)) \\ &= \mu(T^{-i}(A \cap T^{-k}(A))) \\ &= \mu(A \cap T^{-k}(A)) \\ &= 0 \end{aligned}$$

[Damit sind die Mengen $\{T^{-n}(A)\}_{n \in \mathbb{N}}$ so etwas wie fast sicher paarweise disjunkt.] Wir erhalten

$$\mu \left(\bigcup_{n \in \mathbb{N}} T^{-n}(A) \right) = \sum_{n \in \mathbb{N}} \mu(T^{-n}(A)) = \sum_{n \in \mathbb{N}} \mu(A) = \infty$$

Dies steht im Widerspruch zur Endlichkeit des Maßraums.

Beispiel 3.5.

1. Die Zelt-Abbildung t aus Beispiel 2.5 ist maßtreu auf dem endlichen Maßraum $([0, 1], \mathcal{B}([0, 1]), \lambda|_{[0,1]})$. Damit ist sie nach dem Satz von Poincaré auch rekurrent, d.h. jede messbare Menge aus $[0, 1]$ mit positivem Maß wird unendlich oft von fast allen ihren Punkten aufgesucht.
2. Obwohl die Transformation $Tx = x + 1$ aus Beispiel 2.6 maßtreu auf dem σ -endlichen Maßraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ ist, ist leicht ersichtlich, dass T nicht rekurrent ist (wähle beispielsweise $A := [0, 1)$).
3. Die logistische Abbildung L_4 aus Beispiel 2.7 ist maßtreu auf dem endlichen Maßraum $([0, 1], \mathcal{B}([0, 1]), \mu)$, wobei μ das Wahrscheinlichkeitsmaß mit λ -Dichte

$$f_*(x) = \frac{1}{\pi \sqrt{x(1-x)}} \quad (3.1)$$

aus Beispiel 2.7 beschreibt. Damit ist L_4 nach dem Satz von Poincaré auch rekurrent, d.h. jede messbare Menge aus $[0, 1]$ mit positivem Maß wird unendlich oft von fast allen ihren Punkten aufgesucht.

4. Der Linksshift Operator σ ist nach Satz 2.11 maßtreu auf dem endlichen Maßraum $(\Sigma_N, \mathcal{B}(\Sigma_N), \mu_\vartheta)$. Damit ist σ nach dem Satz von Poincaré auch rekurrent, d.h. jede messbare Menge aus Σ_N mit positivem Maß wird unendlich oft von fast allen ihren Punkten aufgesucht.

Eine weitere interessante Charakterisierung von Rekurrenz liefert das folgende Resultat, welches für σ -endlich Räume formuliert werden kann:

Lemma 3.6. Sei T eine maßtreue Abbildung auf einem σ -endlichen Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann sind folgende Aussagen äquivalent:

- (a) T ist rekurrent.
- (b) Es existiert keine Menge $A \in \mathcal{A}$, die sowohl

$$T^{-1}(A) \subseteq A \quad \text{als auch} \quad \mu(A \setminus T^{-1}(A)) > 0$$

erfüllt.

Beweis: (a) \Rightarrow (b): Sei T rekurrent, $A \in \mathcal{A}$ mit $T^{-1}(A) \subseteq A$ und setze $B := A \setminus T^{-1}(A)$. Wir zeigen zunächst

$$B \cap T^{-n}(B) = \emptyset$$

für alle $n \in \mathbb{N}$. Wir nehmen an, dass $x \in B$. Dann ist $x \in A$, aber $T(x) \notin A$ und damit nach Voraussetzung $T(x) \notin T^{-1}(A)$ oder äquivalent $T^2(x) \notin A$. Folgt man dieser Denkweise induktiv erhält man $T^n(x) \notin A$ und damit $T^n(x) \notin B$ oder äquivalent $x \notin T^{-n}(B)$ für alle $n \in \mathbb{N}$. Damit ist die obige Behauptung gezeigt. Lemma 3.3 impliziert nun

$$\mu(A \setminus T^{-1}(A)) = \mu(B) = 0$$

also ist (b) erfüllt.

(b) \Rightarrow (a): Sei $A \in \mathcal{A}$ mit $\mu(A) > 0$ und setze $B := \bigcup_{n \in \mathbb{N}_0} T^{-n}(A)$. Dann gilt $T^{-1}(B) \subseteq B$ und (einsetzen und umformen)

$$B \setminus T^{-1}(B) = A \setminus \bigcup_{n \in \mathbb{N}} T^{-n}(A)$$

(vgl. Menge A^* im Beweis von Lemma 3.3). Aus (b) erhält man nun $\mu(A \setminus \bigcup_{n \in \mathbb{N}} T^{-n}(A)) = 0$ womit T rekurrent ist.

Beispiel 3.7. Zeigen Sie mit Lemma 3.6, dass die Transformation $Tx = x + 1$ aus Beispiel 2.6 nicht rekurrent ist.

3.2 Ergodizität

Rekurrenz stellt sicher, dass jede Menge A mit positivem Maß unendlich oft von fast allen ihren Punkten besucht wird. Können wir aber auch Aussagen treffen über die mittlere Rückkehrzeit? Können wir unendliche mittlere Rückkehrzeiten ausschließen?

Um diese Fragen zu beantworten benötigen wir eine *Unzerlegbarkeitseigenschaft* - die sogenannte Ergodizität. Dabei beschränken wir uns im Folgenden auf Wahrscheinlichkeitsräume.

Das Konzept Ergodizität wurde zuerst 1928 von Birkhoff und Smith unter dem Namen *metric transitivity* eingeführt. Es ist ein zentrales Konzept dieser Vorlesung. Wir werden einige äquivalente Beschreibungen von Ergodizität kennenlernen, welche jede ihren eigenen Reiz hat.

Im Folgenden wollen wir auf dem Raum Ω (oder einer Zerlegung dessen) die Dynamik einer (maßtreuen) Transformation T untersuchen. Angenommen, es existiert eine Menge $A \subseteq \Omega$ mit der Eigenschaft, dass $x \in A$ genau dann wenn $T(x) \in A$. Dann gilt sofort $T(A) \subseteq A$ und gleichzeitig $T(A^c) \subseteq A^c$ und die Restriktionen $T|_A$ und $T|_{A^c}$ können als eigenständige dynamische Systeme aufgefasst werden. Damit kann die Dynamik der Transformation T allein anhand der Dynamiken der Restriktionen $T|_A$ und $T|_{A^c}$ untersucht werden - in diesem Fall lässt sich das Problem also *zerlegen*. Systeme, in denen solche Mengen nicht vorkommen können dann als *nicht zerlegbar* angesehen werden.

Definition 3.8. Eine maßtreue Abbildung T auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ heißt ergodisch, falls für jede Menge $A \in \mathcal{A}$ mit $T^{-1}(A) = A$ entweder $\mu(A) = 0$ oder $\mu(A^c) = 0$ gilt.

Ein sehr einfaches Beispiel zu Beginn:

Beispiel 3.9. Betrachte den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ bestehend aus der Grundmenge $\Omega = \{z\}$, der trivialen σ -Algebra $\mathcal{A} = \{\emptyset, \Omega\}$ und dem Dirac Maß $\mu = \delta_z$. Sei weiter T die Identität. Dann ist T (offensichtlich) maßtreu und ergodisch. Die Menge A mit $T^{-1}(A) = A$ entspricht hier entweder der Grundmenge oder der leeren Menge.

Da jede maßtreue Abbildung auf einem Wahrscheinlichkeitsraum rekurrent ist (Satz von Poincaré) und Ergodizität nur für maßtreue Abbildungen auf einem Wahrscheinlichkeitsraum definiert ist, ist klarerweise jede ergodische Abbildung (auf einem Wahrscheinlichkeitsraum) rekurrent. Das folgende Beispiel verdeutlicht, dass Ergodizität (auf Wahrscheinlichkeitsräumen) stärker als Rekurrenz ist:

Beispiel 3.10. Betrachte den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}([0, 1]), \mu := \lambda|_{[0,1]})$ und die Abbildung

$$T(x) := x$$

Dann ist T rekurrent, aber nicht ergodisch (wähle beispielsweise $A := [0, 0.5]$).

Übungsaufgabe 1. Betrachte den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}([0, 1]), \mu := \lambda|_{[0,1]})$ und die Abbildung

$$T(x) := x + 3/4 \pmod{1}$$

Zeigen Sie, dass T maßtreu, aber nicht ergodisch ist.

Um Charakterisierungen von Ergodizität anzugeben, benötigen wir noch den Begriff der symmetrischen Differenz:

Übungsaufgabe 2. Die Menge $A \Delta B := (A \setminus B) \cup (B \setminus A)$ heißt die symmetrische Differenz von A und B . Zeigen Sie folgende Eigenschaften der symmetrischen Differenz:

1. $A \Delta \emptyset = A$ (Neutrales Element)
2. $A \Delta A = \emptyset$ (Inverses Element)
3. $A \Delta B = B \Delta A$ (Kommutativität)
4. $A \Delta C = (A \Delta B) \Delta (B \Delta C)$
5. $A \Delta (B \Delta C) = (A \Delta B) \Delta C$ (Assoziativität)
6. $(A \Delta B) \cap C = (A \cap C) \Delta (B \cap C)$ (Distributivität)
7. $A \Delta C \subseteq (A \Delta B) \cup (B \Delta C)$ (Dreiecksungleichung)

Wir können nun ergodische Abbildungen charakterisieren:

Satz 3.11. Sei T eine maßtreue Abbildung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$. Dann sind folgende Aussagen äquivalent:

- (a) T ist ergodisch.
- (b) (Unzerlegbarkeitsbedingung)
Existieren $A_1, A_2 \in \mathcal{A}$ mit $T(A_1) \subseteq A_1$, $T(A_2) \subseteq A_2$ und $\mu(A_1 \cap A_2) = 0$, dann gilt $\mu(A_1)\mu(A_2) = 0$.

- (c) Aus $\mu(T^{-1}(A)\Delta A) = 0$ folgt $\mu(A) \in \{0, 1\}$.
- (d) Aus $\mu(A) > 0$ folgt $\mu\left(\bigcup_{n \in \mathbb{N}} T^{-n}(A)\right) = 1$.
- (e) Aus $\mu(A_1)\mu(A_2) > 0$ folgt $\mu(T^{-n}(A_1) \cap A_2) > 0$ für ein $n \in \mathbb{N}_0$.
- (f) Für jedes $f \in L^1$ mit $f \circ T = f$ gilt $f = \text{const} [\mu]$.

Beweis: (a) nach (b): Wir setzen zunächst

$$B_i := \bigcup_{n \in \mathbb{N}_0} T^{-n}(A_i) \quad i \in \{1, 2\}$$

Dann gilt $A_i \subseteq B_i$, $T^{-1}(B_i) = B_i$ und

$$B_1 \cap B_2 \subseteq \bigcup_{n \in \mathbb{N}_0} T^{-n}(A_1 \cap A_2) \quad (\text{Selbst!})$$

Aufgrund der Ergodizität erhalten wir daher $\mu(B_i) \in \{0, 1\}$.

1: Angenommen $\mu(B_i) = 0$ für ein $i \in \{1, 2\}$, dann gilt $0 \leq \mu(A_i) \leq \mu(B_i) = 0$ und damit $\mu(A_1)\mu(A_2) = 0$.

2: Angenommen $\mu(B_1) = \mu(B_2) = 1$. Dann gilt

$$0 \leq \mu(B_1 \cap B_2) \leq \sum_{n \in \mathbb{N}_0} \mu(T^{-n}(A_1 \cap A_2)) = \sum_{n \in \mathbb{N}_0} \mu(A_1 \cap A_2) = 0$$

Aber dies steht im Widerspruch zu $2 = \mu(B_1) + \mu(B_2) - \mu(B_1 \cap B_2) = \mu(B_1 \cup B_2) \leq \mu(B_1) = 1$.
Damit ist Fall 2 ausgeschlossen und wir erhalten (b).

(b) nach (c): Dazu setzen wir

$$\begin{aligned} B &:= \bigcap_{n \in \mathbb{N}_0} \bigcup_{k \geq n} T^{-k}(A) \\ &= \{x \in \Omega : x \in T^{-n}(A) \text{ für unendlich viele } n \in \mathbb{N}_0\} \\ &= \{x \in \Omega : T^n(x) \in A \text{ für unendlich viele } n \in \mathbb{N}_0\} \end{aligned}$$

Dann gilt zunächst $T(B) \subseteq B$ (Selbst!). Wir wollen nun zeigen, dass aufgrund der Voraussetzung auch $\mu(B) = \mu(A)$ gilt. Mit der Dreiecksungleichung der symmetrischen Differenz erhalten wir zunächst $T^{-2}(A)\Delta A \subseteq T^{-2}(A)\Delta T^{-1}(A) \cup T^{-1}(A)\Delta A$ und damit

$$0 \leq \mu(T^{-2}(A)\Delta A) \leq \mu(T^{-2}(A)\Delta T^{-1}(A)) + \mu(T^{-1}(A)\Delta A) = \mu(T^{-1}(A)\Delta A) + \mu(T^{-1}(A)\Delta A) = 0$$

woraus induktiv $\mu(T^{-n}(A)\Delta A) = 0$ für alle $n \in \mathbb{N}$ folgt. Damit erhalten wir

$$\begin{aligned} 0 &\leq \mu(A \Delta B) \\ &= \mu\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A \setminus T^{-k}(A) \cup \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} T^{-k}(A) \setminus A\right) \\ &\leq \sum_{n \in \mathbb{N}} \mu(A \setminus T^{-n}(A)) + \sum_{n \in \mathbb{N}} \mu(T^{-n}(A) \setminus A) \\ &= \sum_{n \in \mathbb{N}} \mu(T^{-n}(A)\Delta A) \\ &= 0 \end{aligned}$$

und schließlich

$$\mu(B) = \mu(B \cap A) + \mu(B \setminus A) \leq \mu(A) = \mu(A \cap B) + \mu(A \setminus B) \leq \mu(B)$$

Mittels analogem Vorgehen für A^c und (b) folgt $0 = \mu(A)\mu(A^c) = \mu(A) - \mu(A)^2$ und damit $\mu(A) \in \{0, 1\}$.

(c) nach (d): Wir wählen $\mu(A) > 0$ und setzen wieder

$$B := \bigcup_{n \in \mathbb{N}_0} T^{-n}(A)$$

und wollen $\mu(B) = 1$ zeigen.

Wie bereits im Beweisteil (a) nach (b) erwähnt, gilt $T^{-1}(B) = B$, also $\mu(T^{-1}(B) \Delta B) = 0$ und wegen (c) $\mu(B) \in \{0, 1\}$. Aufgrund der Voraussetzung $\mu(A) > 0$ und wegen $A \subseteq B$ folgt die Behauptung.

(d) nach (e): Wir zeigen, dass aus $\mu(T^{-n}(A_1) \cap A_2) = 0$ für alle $n \in \mathbb{N}$ folgt, dass $\mu(A_1)\mu(A_2) = 0$ gilt. Falls $\mu(A_1) = 0$, dann ist nichts zu zeigen. Falls $\mu(A_1) > 0$, dann ist wegen (d) $\bigcup_{n \in \mathbb{N}} T^{-n}(A_1)$ eine μ -Eins-Menge und wir erhalten

$$\begin{aligned} \mu(A_2) &= \mu\left(\left(\bigcup_{n \in \mathbb{N}} T^{-n}(A_1)\right) \cap A_2\right) \\ &= \mu\left(\bigcup_{n \in \mathbb{N}} (T^{-n}(A_1) \cap A_2)\right) \\ &\leq \sum_{n \in \mathbb{N}} \mu(T^{-n}(A_1) \cap A_2) \\ &= 0 \end{aligned}$$

und damit $\mu(A_1)\mu(A_2) = 0$.

(e) nach (f): Für ein integrierbares f (messbar) mit $f \circ T = f$ $[\mu]$ setzen wir für jedes $n \in \mathbb{N}$ und jedes $l \in \mathbb{Z}$

$$C_{n,l} := \left\{x \in \Omega : \frac{l}{2^n} \leq f(x) < \frac{l+1}{2^n}\right\} = f^{-1}\left(\left[\frac{l}{2^n}, \frac{l+1}{2^n}\right)\right)$$

Offensichtlich gilt dann $\mu\left(\bigcup_{l \in \mathbb{Z}} C_{n,l}\right) = 1$ für alle $n \in \mathbb{N}$. Des Weiteren erhalten wir

$$T^{-1}(C_{n,l}) = \left\{x \in \Omega : \frac{l}{2^n} \leq f(T(x)) < \frac{l+1}{2^n}\right\} = \left\{x \in \Omega : \frac{l}{2^n} \leq f(x) < \frac{l+1}{2^n}\right\} = C_{n,l} \quad [\mu]$$

sowie induktiv $T^{-k}(C_{n,l}) = C_{n,l}$ $[\mu]$ für alle $k \in \mathbb{N}$. Aufgrund der Disjunktheit der Mengen $(C_{n,l})_{l \in \mathbb{Z}}$ gilt für $l_1 \neq l_2$ weiter

$$\mu(T^{-k}(C_{n,l_1}) \cap C_{n,l_2}) = \mu(C_{n,l_1} \cap C_{n,l_2}) = 0$$

für alle $k \in \mathbb{N}$ und wegen (e) $\mu(C_{n,l_1})\mu(C_{n,l_2}) = 0$. Damit existiert für jedes $n \in \mathbb{N}$ genau ein $l_n \in \mathbb{Z}$ mit $\mu(C_{n,l_n}) = 1$. Auf $\Omega^* := \bigcap_{n \in \mathbb{N}} C_{n,l_n}$ ist dann f konstant und wegen $\mu(\Omega^*) = 1$ gilt damit (f).

(f) nach (a): Wähle $f = \mathbf{1}_A$ für $A \in \mathcal{A}$ mit $T^{-1}(A) = A$.

Bevor wir weitergehen, wollen wir die eben gezeigten Charakterisierungen interpretieren:

- Die Bedeutung von Eigenschaft (b) ist leicht zu erfassen: Es besagt gerade, dass eine Situation wie in Abbildung 3.2 nicht auftreten kann, wenn nicht (mindestens) eine der Mengen Maß Null hat.

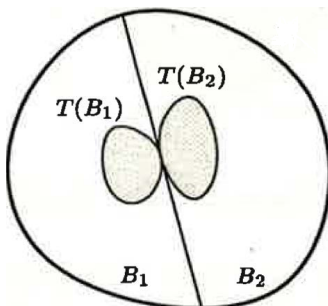


Abbildung 3.2: Illustration Ergodizität; Bedingung (b).

- Für eine maßtreue Abbildung T ist das Mengensystem

$$\mathcal{A}_T := \{A \in \mathcal{A} : \mu(T^{-1}(A) \Delta A) = 0\}$$

eine Unter- σ -Algebra von \mathcal{A} (ÜA!) und heißt die σ -Algebra der fast sicher invarianten Mengen. Gilt für eine Menge $A \in \mathcal{A}$ entweder $T^{-1}(A) \subseteq A$ oder $A \subseteq T^{-1}(A)$, dann gilt $A \in \mathcal{A}_T$ (ÜA!).

- Eigenschaft (d) besagt, dass jedes A mit $\mu(A) > 0$ irgendwann von fast allen Punkten besucht wird. Damit ist Ergodizität eine stärkere Eigenschaft als Rekurrenz.
- Aus (e) leiten wir ab, dass Ergodizität als Analogon der topologischen Transitivität betrachtet werden kann: Es ist immer möglich, sich von einer beliebigen Menge mit positivem Maß zu einer beliebigen anderen solchen Menge zu bewegen.

Beispiel 3.12. Sowohl die Zelt-Abbildung t als auch die logistische Abbildung L_4 als auch der Linksshift σ sind ergodisch. Wir umschiffen den direkten Nachweis und zeigen die Ergodizität der drei Abbildungen zu einem späteren Zeitpunkt unter Ausnutzung der Tatsache, dass die drei Abbildungen isomorph zueinander sind.

Unsere Hauptmotivation für die Einführung der Ergodizität bestand darin, unendliche mittlere Rückkehrzeiten auszuschließen, die nicht a priori durch das Rekurrenz-Theorem (Poincaré) ausgeschlossen werden. Es ist noch nicht sicher, ob wir die richtige Definition gewählt haben um dieses Ziel zu erreichen. Das folgende Theorem, das üblicherweise als Kac's Lemma bezeichnet wird, zeigt, dass wir tatsächlich erfolgreich waren.

Wir definieren die Rückkehrzeit $t_A : A \rightarrow \mathbb{N} \cup \{\infty\}$ mittels

$$t_A(x) := \inf\{n \in \mathbb{N} : T^n(x) \in A\}$$

Wie üblich wird dabei t_A als ∞ gewählt, falls die Menge auf der rechten Seite leer ist. (Aufgrund des Satzes von Poincaré ist t_A fast überall auf A endlich).

Satz 3.13. (*Lemma von Kac, 1947*)

Sei T eine ergodische Abbildung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$. Dann gilt für jedes $A \in \mathcal{A}$ mit $\mu(A) > 0$

$$E(t_A | A) = \frac{1}{\mu(A)} \int_A t_A \, d\mu = \frac{1}{\mu(A)}$$

Beweis: Die Abbildung t_A ist messbar und mit der disjunkten Zerlegung im Bildbereich $A_n := \{x \in A : t_A(x) = n\} = t_A^{-1}(\{n\})$ erhalten wir zunächst

$$\int_A t_A \, d\mu = \sum_{n \in \mathbb{N}} \int_{A_n} t_A \, d\mu = \sum_{n \in \mathbb{N}} n \mu(A_n) = \sum_{k \in \mathbb{N}} \sum_{n \geq k} \mu(A_n)$$

Für jedes $k \in \mathbb{N}$ setzen wir

$$B_k := T^{-1}(A^c) \cap \dots \cap T^{-(k-1)}(A^c) \cap T^{-k}(A)$$

Dann gilt nach Definition $B_k \cap A = A_k$ und $T^{-1}(B_k \setminus A) = T^{-1}(B_k) \cap T^{-1}(A^c) = B_{k+1}$. Aufgrund der Ergodizität (Bedingung (d)) erhalten wir nun (nachrechnen)

$$\mu\left(\bigcup_{k \in \mathbb{N}} B_k\right) = \mu\left(\bigcup_{k \in \mathbb{N}} T^{-k}(A)\right) = 1$$

Darüberhinaus gilt

$$\mu(B_{k+1}) = \mu(T^{-1}(B_k \setminus A)) = \mu(B_k \setminus A) = \mu(B_k) - \mu(B_k \cap A) = \mu(B_k) - \mu(A_k)$$

damit $\mu(B_k) = \sum_{n \geq k} \mu(A_n)$ und schließlich

$$\int_A t_A \, d\mu = \sum_{k \in \mathbb{N}} \sum_{n \geq k} \mu(A_n) = \sum_{k \in \mathbb{N}} \mu(B_k) = \mu\left(\bigcup_{k \in \mathbb{N}} B_k\right) = 1$$

Damit ist die Behauptung gezeigt.

Ergodizität erzwingt also nicht nur die Endlichkeit der mittleren Wiederkehrzeiten (∞ nur auf einer Nullmenge), sondern stellt auch sicher, dass diese sich sehr eindeutig verhalten: je kleiner das Maß einer Menge A ist, desto länger braucht ein Punkt (im Mittel), um zu A zurückzukehren. Mit anderen Worten: Ein ergodisches Maß weist den schnell zurückkehrenden Punkten das größte Gewicht zu. Beachten Sie, dass diese Beobachtung nicht trivial ist (auch wenn es so klingen mag). Es bestätigt lediglich, dass unsere Definitionen sauber zusammenpassen und einige grundlegende dynamische Eigenschaften genau so wiedergeben, wie wir es wollten.

3.3 Ergodensatz

Dieses Kapitel ist dem Ergodensatz und einigen seiner Anwendungen gewidmet. Der Ergodensatz, der 1931 zunächst von J. von Neumann für den Fall der Konvergenz im Mittel (L^1) und dann von G. D. Birkhoff für den stärkeren Fall der punktweisen Konvergenz bewiesen wurde, löste eine wichtige Frage, die sich in der statistischen Mechanik stellte: unter welcher Bedingung stimmen in einem dynamischen System Zeitmittel und Raummittel überein.

Sei T eine messbare Abbildung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$. Für $A \in \mathcal{A}$ und $x \in \Omega$ definieren wir die *relative Häufigkeit* der ersten n Iterationen von x in A

$$h_A(x, n) := \frac{1}{n} \sum_{k=0}^{n-1} (\mathbf{1}_A \circ T^k)(x) = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A)}(x)$$

Wir werden uns hauptsächlich für das Verhalten von $h_A(x, n)$ für $n \rightarrow \infty$ interessieren. Konvergiert diese Folge? Wenn ja, wie hängt der Grenzwert von x und A ab?

Die punktweise Version des Ergodensatzes, bekannt als Birkhoffs Ergodensatz besagt in seiner einfachsten Form, dass für T ergodisch

$$\lim_{n \rightarrow \infty} h_A(x, n) = \mu(A)$$

für fast alle $x \in \Omega$. Der Wert auf der linken Seite entspricht dem Grenzwert der durchschnittlichen Anzahl von Besuchen von $T^n(x)$ in A . Wir können uns diese Größe als das Zeitmittel der Besuche des Orbits von x in A vorstellen. Die Größe rechts kann als das Raummittel von A gedacht werden. Der Birkhoffsche Ergodensatz impliziert, dass für jede messbare Menge A und für fast jeden Punkt x der Zeitmittelwert asymptotisch gleich dem Raummittelwert ist.

Als geringfügige Verallgemeinerung werden wir auch das asymptotische Verhalten der Summen

$$S_n f(x) := \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k)(x) \quad \text{für } f \in L^1$$

untersuchen. Um diesen Ausdruck zu motivieren, sei daran erinnert, dass wir $(\Omega, \mathcal{A}, \mu)$ wiederholt als die probabilistische Struktur interpretiert haben, die das Ergebnis eines Experiments beschreibt. Ist $\{X_n\}_{n \in \mathbb{N}_0}$ eine Folge von unabhängigen Experimenten/Zufallsvariablen, die alle der gleichen Verteilung μ unterliegen, garantieren die Gesetze der großen Zahlen die fast sichere Konvergenz von

$$\bar{X} := \frac{1}{n} \sum_{k=0}^{n-1} X_k$$

gegen den Erwartungswert $E(X)$. Leider sind in unserem Fall die Folgenglieder

$$x, T(x), T^2(x), \dots$$

definitiv nicht unabhängig: Sie sind durch die deterministische Abbildung T gekoppelt. Die Untersuchung der asymptotischen Eigenschaften von $S_n f$ beschäftigt sich daher implizit mit der Frage, ob Orbits unter T (zumindest asymptotisch) als eine Folge von unabhängigen Experimenten betrachtet werden können.

Das folgende einfache Beispiel zeigt, dass das Verhalten von $S_n f$ nicht von vornherein offensichtlich sein muss. Betrachte dazu $\Omega = \{1, \dots, 6\}$ mit $\mathcal{A} = \mathcal{P}(\Omega)$ und dem Wahrscheinlichkeitsmaß $\mu : \mathcal{P}(\Omega) \rightarrow [0, 1]$ gegeben durch

$$\mu(A) := \frac{|A|}{|\Omega|} = \frac{|A|}{6}$$

Wir definieren die zwei maßtreuen Abbildungen

$$T_1(x) := x + 1 - \lfloor \frac{x}{6} \rfloor \cdot 6 \quad \text{und} \quad T_2 := T_1 \circ T_1$$

Was T_1 betrifft, so ist leicht zu erkennen, dass für jede Funktion $f : \Omega \rightarrow \mathbb{R}$ der Grenzwert

$$\lim_{n \rightarrow \infty} S_n f(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T_1^k)(x) = \frac{1}{6} \sum_{i=1}^6 f(i) = \int_{\Omega} f(x) d\mu(x)$$

unabhängig von der Wahl von x ist. Schaut man auf T_2 werden die Dinge komplizierter. Wählen wir beispielsweise $B := \{1, 2, 3\}$, $E := \{2, 4, 6\}$ und $O := \{1, 3, 5\}$, dann erhalten wir

$$\lim_{n \rightarrow \infty} h_B(x, n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} (\mathbf{1}_B \circ T_2^k)(x) = \begin{cases} \frac{2}{3} & x \in O \\ \frac{1}{3} & x \in E \end{cases}$$

womit der Grenzwert von der Wahl von x abhängt. Letzterer hängt auch von der gewählten Menge ab, denn für $C := \{1, 3\}$ gilt

$$\lim_{n \rightarrow \infty} h_C(x, n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} (\mathbf{1}_C \circ T_2^k)(x) = \begin{cases} \frac{2}{3} & x \in O \\ 0 & x \in E \end{cases}$$

In diesem einfachen Beispiel kann natürlich jede interessierende Größe explizit berechnet werden. In der allgemeinen Situation muss das Verhalten von $S_n f$ jedoch weiter analysiert werden. Für letzteres erweisen sich zwei Werkzeuge als entscheidend: das Konzept der bedingten Erwartungen und die σ -Algebra

$$\mathcal{A}_T = \{A \in \mathcal{A} : \mu(T^{-1}(A) \Delta A) = 0\}$$

der fast sicher invarianten Mengen. Beachte: \mathcal{A}_T ist eine Unter- σ -Algebra von \mathcal{A} .

Einschub: Bedingte Erwartung

Wir betrachten den Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und eine Unter- σ -Algebra \mathcal{G} von \mathcal{A} .

Beachte: Eine Zufallsvariable X ist im Allgemeinen nicht bezüglich einer beliebigen Unter- σ -Algebra messbar.

Frage: Können wir eine X beschreibende und \mathcal{G} -messbare Zufallsvariable U finden?

→ Informationsverlust aufgrund größerer Struktur!

Satz: Sei X eine positive Zufallsvariable. Dann gibt es eine positive \mathcal{G} -messbare Zufallsvariable U mit

$$\int_G U \, dP = \int_G X \, dP$$

für alle $G \in \mathcal{G}$. Sind U und V positive \mathcal{G} -messbare Zufallsvariable mit

$$\int_G U \, dP = \int_G X \, dP = \int_G V \, dP$$

für alle $G \in \mathcal{G}$, so gibt es eine Nullmenge $N \in \mathcal{G}$ mit $U(\omega) = V(\omega)$ für alle $\omega \in \Omega \setminus N$, d.h. U ist $P|_{\mathcal{G}}$ -fast sicher eindeutig bestimmt.

Beweis: Die Existenz erhält man mit dem Satz von Radon-Nikodym.

Beachte: U beschreibt X nur im Sinne der Integration über Mengen aus \mathcal{G} . Wären die Integrale für alle $A \in \mathcal{A}$ gleich, dann wären X und U P -fast sicher identisch. (Warum?)

Ist X eine positive Zufallsvariable, so nennen wir jede positive \mathcal{G} -messbare Zufallsvariable U mit

$$\int_G U \, dP = \int_G X \, dP$$

für alle $G \in \mathcal{G}$ eine *Version der \mathcal{G} -bedingten Erwartung* von X ; wir bezeichnen mit $E(X|\mathcal{G})$ eine beliebige Version der \mathcal{G} -bedingten Erwartung von X und nennen $E(X|\mathcal{G})$ die *\mathcal{G} -bedingte Erwartung* von X .

Da die σ -Algebra \mathcal{G} außer der leeren Menge weitere Nullmengen enthalten kann, besitzt die \mathcal{G} -bedingte Erwartung einer positiven Zufallsvariable im allgemeinen mehrere Versionen, von denen aber je zwei auf dem Komplement einer Nullmenge in \mathcal{G} übereinstimmen. Für eine beliebige Zufallsvariable Y gilt daher $U(\omega) = Y(\omega)$ fast sicher entweder für jede oder für keine Version U der \mathcal{G} -bedingten Erwartung von X ; gilt sie für jede Version der \mathcal{G} -bedingten Erwartung von X , so schreiben wir kurz $E(X|\mathcal{G}) = Y$.

Eigenschaften:

1. Wegen $\Omega \in \mathcal{G}$ gilt $E(E(X|\mathcal{G})) = E(X)$.
2. Ist $\mathcal{G} = \{\emptyset, \Omega\}$, dann ist jede \mathcal{G} -messbare Zufallsvariable X konstant und es gilt $E(X|\mathcal{G}) = E(X)$.
3. Ist $\mathcal{G} = \mathcal{A}$, dann ist jede positive Zufallsvariable X \mathcal{G} -messbar und es gilt $E(X|\mathcal{G}) = X$ P -fast sicher.
4. $E(E(X|\mathcal{G})|\mathcal{G}) = E(X|\mathcal{G})$.
5. $E(X + Y|\mathcal{G}) = E(X|\mathcal{G}) + E(Y|\mathcal{G})$.
6. Für jede positive \mathcal{G} -messbare Zufallsvariable U gilt $E(UX|\mathcal{G}) = U E(X|\mathcal{G})$.
7. Im Fall $X \leq Y$ gilt $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$.

Das folgende Ergebnis zeigt, dass die Definition der \mathcal{G} -bedingten Erwartung einer positiven Zufallsvariable mit der Definition ihres bedingten Erwartungswerts bezüglich einem Ereignis im Einklang steht:

Lemma: (Fourier-Entwicklung)

Sei $\{G_i\}_{i \in \{1, \dots, m\}} \subseteq \mathcal{A}$ eine disjunkte Familie von Ereignissen mit $\sum_{i=1}^m G_i = \Omega$ und $P(G_i) > 0$ für alle $i \in \{1, \dots, m\}$ und sei \mathcal{G} die von der Familie $\{G_i\}_{i \in \{1, \dots, m\}}$ erzeugte σ -Algebra. Dann gilt für jede positive Zufallsvariable X

$$E(X|\mathcal{G}) = \sum_{i=1}^m E(X|G_i) \mathbf{1}_{G_i} = \sum_{i=1}^m \left(\frac{1}{P(G_i)} \int_{G_i} X \, dP \right) \mathbf{1}_{G_i}$$

Das Lemma deutet darauf hin, dass der Übergang von einer positiven Zufallsvariable X zu ihrer \mathcal{G} -bedingten Erwartung als ein lokales Mitteln verstanden werden kann: Wenn die σ -Algebra durch endlich viele Ereignisse erzeugt wird, dann nimmt die \mathcal{G} -bedingte Erwartung $E(X|\mathcal{G})$ nur endlich viele Werte an, und im Extremfall $\mathcal{G} = \{\emptyset, \Omega\}$ ist die \mathcal{G} -bedingten Erwartung von X sogar konstant.

Die Definition der \mathcal{G} -bedingten Erwartung für integrierbare Zufallsvariable erfolgt über die übliche Zerlegung in Positiv- und Negativteil.

Die bedingte Erwartung einer quadratisch integrierbaren Zufallsvariable kann als Projektion auf die Menge aller quadratisch integrierbaren und \mathcal{G} -messbaren Zufallsvariablen interpretiert werden:

Satz: (Projektion)

Sei $X \in L^2(\Omega, \mathcal{A}, P)$. Dann sind für eine quadratisch integrierbare und \mathcal{G} -messbare Zufallsvariable Y folgende Aussagen äquivalent:

- (a) $Y = E(X|\mathcal{G})$.
- (b) Die Ungleichung $E((X - Y)^2) \leq E((X - Z)^2)$ gilt für alle \mathcal{G} -messbaren Zufallsvariablen $Z \in L^2(\Omega, \mathcal{A}, P)$.

Ende Einschub

Lemma 3.14. Sei T eine maßtreue Abbildung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ und $f \in L^1$. Dann gilt $E(f|\mathcal{A}_T)(T(\omega)) = E(f|\mathcal{A}_T)(\omega)$ für μ -fast alle $\omega \in \Omega$.

Beweis: Für alle $G \in \mathcal{A}_T$ gilt

$$\begin{aligned} \int_G E(f|\mathcal{A}_T)(T(\omega)) \, d\mu(\omega) &= \int_{T^{-1}(G)} E(f|\mathcal{A}_T)(T(\omega)) \, d\mu(\omega) \\ &= \int_G E(f|\mathcal{A}_T)(\omega) \, d\mu_T(\omega) \\ &= \int_G E(f|\mathcal{A}_T)(\omega) \, d\mu(\omega) \end{aligned}$$

und damit die Behauptung.

Satz 3.15. (*Ergodensätze*)

Sei T eine maßtreue Abbildung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ und $f \in L^1$. Dann existiert eine Funktion $f^* \in L^1$ mit $f^* \circ T = f^*$ $[\mu]$ und

$$\int_{\Omega} f^* \, d\mu = \int_{\Omega} f \, d\mu$$

so dass

1. $\lim_{n \rightarrow \infty} S_n f(x) = f^*(x)$ punktweise für fast alle $x \in \Omega$. (*Birkhoff Ergodensatz*)
2. $\lim_{n \rightarrow \infty} S_n f = f^*$ in L^1 , d.h.

$$\lim_{n \rightarrow \infty} \int_{\Omega} |S_n f(x) - f^*(x)| \, d\mu(x) = 0$$

(*von Neumann Ergodensatz*)

Beweis: Wir werden zunächst zeigen, dass $f^* := E(f|\mathcal{A}_T)$ definiert als die bedingte Erwartung von f unter \mathcal{A}_T die erste Eigenschaft erfüllt. Wir erhalten

$$E(f^*) = E(E(f|\mathcal{A}_T)) = E(f) < \infty$$

und damit $f^* \in L^1$. Weiter gilt mit Lemma 3.14 $f^* \circ T = f^*$ $[\mu]$ und nach Definition der bedingten Erwartung auch

$$\int_{\Omega} f^* \, d\mu = \int_{\Omega} f \, d\mu$$

Sei nun $g \in L^1$ und wähle

$$G_n(x) := \max_{1 \leq k \leq n} \sum_{i=0}^{k-1} g \circ T^i(x)$$

Dann ist die Folge $\{G_n(x)\}_{n \in \mathbb{N}}$ monoton wachsend und entweder konvergiert diese oder divergiert gegen $+\infty$. Wir definieren nun die Menge

$$B := \{x \in \Omega : \sup_{n \in \mathbb{N}} G_n(x) < \infty\} \in \mathcal{A}$$

aller Konvergenzpunkte. Wegen

$$\sum_{i=0}^k g \circ T^i(x) = g(x) + \sum_{i=1}^k g \circ T^i(x) = g(x) + \sum_{i=0}^{k-1} g \circ T^i(T(x))$$

erhalten wir zunächst

$$\begin{aligned} (G_n \circ T)(x) &= \max_{1 \leq k \leq n} \sum_{i=0}^{k-1} g \circ T^i(T(x)) \\ &= \max_{1 \leq k \leq n} \left(\sum_{i=0}^k g \circ T^i(x) - g(x) \right) \\ &= \max_{2 \leq k \leq n+1} \sum_{i=0}^{k-1} g \circ T^i(x) - g(x) \end{aligned}$$

und damit

$$\begin{aligned}
 G_{n+1}(x) - (G_n \circ T)(x) &= \max \left\{ \max_{2 \leq k \leq n+1} \sum_{i=0}^{k-1} g \circ T^i(x), g(x) \right\} - (G_n \circ T)(x) \\
 &= \max \{ (G_n \circ T)(x) + g(x), g(x) \} - (G_n \circ T)(x) \\
 &= g(x) + \max \{ (G_n \circ T)(x), 0 \} - (G_n \circ T)(x) \\
 &= g(x) - \min \{ 0, (G_n \circ T)(x) \}
 \end{aligned}$$

Für jedes $x \in B$ gilt wegen obiger Gleichung

$$\sup_{n \in \mathbb{N}} (G_n \circ T)(x) \leq \sup_{n \in \mathbb{N}} (G_{n+1}(x) - g(x)) < \infty$$

und damit $T(x) \in B$, d.h. $x \in T^{-1}(B)$, also $B \subseteq T^{-1}(B)$. Daraus erhalten wir $B \in \mathcal{A}_T$ und damit $B^c \in \mathcal{A}_T$.

Für Punkte $x \in B^c$ divergiert die Folge $\{G_n(x)\}_{n \in \mathbb{N}}$ und wegen

$$\begin{aligned}
 \int_{B^c} G_n(x) \, d\mu(x) &= \int_{B^c} G_n(x) \, d\mu_T(x) \\
 &= \int_{T^{-1}(B^c)} (G_n \circ T)(x) \, d\mu(x) \\
 &= \int_{B^c} (G_n \circ T)(x) \, d\mu(x)
 \end{aligned}$$

und damit

$$\begin{aligned}
 0 &\leq \int_{B^c} G_{n+1}(x) - G_n(x) \, d\mu(x) \\
 &= \int_{B^c} G_{n+1}(x) - (G_n \circ T)(x) \, d\mu(x) \\
 &= \int_{B^c} g(x) - \min \{ 0, (G_n \circ T)(x) \} \, d\mu(x)
 \end{aligned}$$

erhalten wir mit dem Satz über die majorisierte/dominierte Konvergenz und der Definition der bedingten Erwartung

$$0 \leq \limsup_{n \rightarrow \infty} \int_{B^c} G_{n+1} - G_n \, d\mu = \int_{B^c} g \, d\mu = \int_{B^c} E(g|\mathcal{A}_T) \, d\mu$$

Aus der letzten Ungleichung lässt sich schließen, dass wenn $E(g|\mathcal{A}_T) < 0$ auf Ω , dann muss $\mu(B^c) = 0$ und damit $\mu(B) = 1$ gelten. Für $x \in B$ gilt weiter

$$\liminf_{n \rightarrow \infty} S_n g(x) \leq \limsup_{n \rightarrow \infty} S_n g(x) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i(x) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} G_n(x) = 0$$

Nun werden wir den Ergodensatz von Birkhoff beweisen:

Sei $f \in L^1$, $\varepsilon > 0$ und setze zunächst $g_1 := f - E(f|\mathcal{A}_T) - \varepsilon$. Dann gilt $E(g_1|\mathcal{A}_T) = -\varepsilon < 0$

sowie für μ -fast alle $x \in \Omega$

$$\begin{aligned} S_n g_1(x) &= \frac{1}{n} \sum_{i=0}^{n-1} g_1 \circ T^i(x) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x) - \frac{1}{n} \sum_{i=0}^{n-1} E(f|\mathcal{A}_T) \circ T^i(x) - \varepsilon \\ &= S_n f(x) - E(f|\mathcal{A}_T) - \varepsilon \end{aligned}$$

und schließlich

$$0 \geq \limsup_{n \rightarrow \infty} S_n g_1(x) = \limsup_{n \rightarrow \infty} S_n f(x) - E(f|\mathcal{A}_T) - \varepsilon$$

Für $g_2 := E(f|\mathcal{A}_T) - f - \varepsilon$ erhalten wir nach analogem Vorgehen $E(g_2|\mathcal{A}_T) = -\varepsilon < 0$ sowie für μ -fast alle $x \in \Omega$

$$S_n g_2(x) = E(f|\mathcal{A}_T) - S_n f(x) - \varepsilon$$

und schließlich

$$0 \geq \liminf_{n \rightarrow \infty} S_n g_2(x) = E(f|\mathcal{A}_T) - \limsup_{n \rightarrow \infty} S_n f(x) - \varepsilon$$

Zusammenfassend erhalten wir

$$\limsup_{n \rightarrow \infty} S_n f(x) \in [E(f|\mathcal{A}_T) - \varepsilon, E(f|\mathcal{A}_T) + \varepsilon] [\mu]$$

und analog

$$\liminf_{n \rightarrow \infty} S_n f(x) \in [E(f|\mathcal{A}_T) - \varepsilon, E(f|\mathcal{A}_T) + \varepsilon] [\mu]$$

Da ε beliebig war gilt also

$$\lim_{n \rightarrow \infty} S_n f = E(f|\mathcal{A}_T) [\mu]$$

Damit haben wir den Ergodensatz von Birkhoff bewiesen.

Wir werden abschließend den von Neumann Ergodensatz beweisen:

Sei zunächst $f \in L^1$ beschränkt. In diesem Fall ist auch $S_n f$ beschränkt und die fast sichere Konvergenz aus dem Ergodensatz von Birkhoff liefert, zusammen mit dem Satz über die majorisierte/dominierte Konvergenz,

$$\lim_{n \rightarrow \infty} \|S_n f - E(f|\mathcal{A}_T)\|_1 = \lim_{n \rightarrow \infty} \int_{\Omega} |S_n f(x) - E(f|\mathcal{A}_T)(x)| d\mu(x) = 0$$

Sei nun $f \in L^1$ beliebig und $\varepsilon > 0$. Dann existiert ein $f_\varepsilon \in L^1$ beschränkt mit $\|f - f_\varepsilon\|_1 < \varepsilon$ und wir erhalten

$$\begin{aligned} \|S_n f - E(f|\mathcal{A}_T)\|_1 &\leq \|S_n f - S_n f_\varepsilon\| + \|S_n f_\varepsilon - E(f_\varepsilon|\mathcal{A}_T)\| + \|E(f_\varepsilon|\mathcal{A}_T) - E(f|\mathcal{A}_T)\| \\ &\leq \varepsilon + \|S_n f_\varepsilon - E(f_\varepsilon|\mathcal{A}_T)\| + \varepsilon \end{aligned}$$

Damit erhalten wir

$$0 \leq \liminf_{n \rightarrow \infty} \|S_n f - E(f|\mathcal{A}_T)\|_1 \leq \limsup_{n \rightarrow \infty} \|S_n f - E(f|\mathcal{A}_T)\|_1 < 2\varepsilon$$

Da ε beliebig war gilt also

$$\lim_{n \rightarrow \infty} \|S_n f - E(f|\mathcal{A}_T)\|_1 = 0$$

Damit haben wir den von Neumann Ergodensatz bewiesen.

Nun ein paar Anmerkungen zu Satz 3.15:

Ist T ergodisch, dann hat nach Satz 3.11 jede Menge aus \mathcal{A}_T entweder Maß 0 oder 1. In diesem Fall gilt

$$E(f|\mathcal{A}_T)(x) = E(f)$$

für fast alle $x \in \Omega$ und wir erhalten das folgende Korollar:

Korollar 3.16. (*Ergodensätze für ergodische Abbildungen*)

Sei T eine maßtreue und ergodische Abbildung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ und $f \in L^1$. Dann gilt

1. $\lim_{n \rightarrow \infty} S_n f(x) = E(f) = \text{const}$ für fast alle $x \in \Omega$. (*Birkhoff Ergodensatz*)
2. $\lim_{n \rightarrow \infty} S_n f = E(f)$ in L^1 . (*von Neumann Ergodensatz*)

Physiker formulieren das Birkhoff'sche Ergodentheorem üblicherweise um, indem sie sagen, dass das Zeitmittel und das Raummittel zusammenfallen. Insbesondere impliziert diese Tatsache im ergodischen Fall

$$\lim_{n \rightarrow \infty} h_A(x, n) = \lim_{n \rightarrow \infty} S_n \mathbf{1}_A(x) = E(\mathbf{1}_A) = \mu(A)$$

für fast alle x . Vor dem Hintergrund des Lemmas von Kac, sollte das letzte Ergebnis nicht überraschen. In nicht-technischer Sprache besagt es einfach, dass der Orbit eines typischen Punktes gemäß μ verteilt ist: Das Maß ist genau dort konzentriert, wo sich die Orbits der meisten Punkte konzentrieren.

Wir kehren noch einmal zu unserem Eingangsbeispiel mit den Abbildungen T_1 und T_2 zurück. Da T_1 ergodisch ist, erhalten wir

$$\lim_{n \rightarrow \infty} S_n f(x) = E(f) = \frac{1}{6} \sum_{i=1}^6 f(i)$$

für alle $x \in \Omega = \{1, \dots, 6\}$. Während dieses Ergebnis schon bekannt war, können wir nun leicht ein analoges Ergebnis für T_2 angeben. Die Schwierigkeiten, die im Zusammenhang mit T_2 auftraten, kamen eindeutig von der Tatsache, dass die Abbildung nicht ergodisch ist. Nachdem man sich klar macht, dass $\mathcal{A}_{T_2} = \{\emptyset, O, E, \Omega\}$ liefert der Ergodensatz von Birkhoff

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n f(x) &= E(f|\mathcal{A}_{T_2}) \\ &= E(f|O) \mathbf{1}_O(x) + E(f|E) \mathbf{1}_E(x) \\ &= \frac{1}{3} \sum_{i \in O} f(i) \mathbf{1}_O(x) + \frac{1}{3} \sum_{i \in E} f(i) \mathbf{1}_E(x) \end{aligned}$$

für alle $x \in \Omega = \{1, \dots, 6\}$.

3.4 Mischungseigenschaften

Als wir das Konzept der Ergodizität einführten, befassten wir uns hauptsächlich mit sehr langen Wiederkehrzeiten in bestimmten maßerhaltenden Systemen. Wir haben uns eine ergodische Abbildung als eine Unzerlegbarkeitseigenschaft vorgestellt, welche den ganzen Raum mischt.

Nachdem wir einige der verschiedenen Aspekte der Ergodizität kennengelernt haben, wollen wir nun kurz diskutieren, ob eine ergodische Abbildung wirklich eine Mischung im Sinne des allgemeinen Sprachgebrauchs erzeugt. Um dem Begriff Mischung eine exakte Bedeutung zu geben, denken Sie an ein Glas Wasser (formalisiert als Grundmenge Ω), das einen Tropfen Tinte $A \subseteq \Omega$ enthält. Die Aufgabe von $T : \Omega \rightarrow \Omega$ bestehe darin, einmal mit einem Löffel umzurühren. Welches Verhalten erwarten wir von der Tinte? Nimmt man eine beliebige Menge $B \subseteq \Omega$, so wird der Anteil der Tinte in B - nach einer ausreichend großen Anzahl von Rührvorgängen - mehr oder weniger dem Prozentsatz der Tinte im gesamten Glas entsprechen.

Definition 3.17. Eine maßtreue Abbildung T auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ heißt stark mischend, falls für jede Wahl von Mengen $A, B \in \mathcal{A}$

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

gilt.

Die Eigenschaft stark mischend zu sein kann auch, falls $\mu(B) > 0$, äquivalent formuliert werden:

$$\lim_{n \rightarrow \infty} \frac{\mu(T^{-n}(A) \cap B)}{\mu(B)} = \mu(A)$$

D.h. nach einiger Zeit ist der Anteil von A , der sich in B befindet (unter Iteration durch T), nahe an der Größe von A (das ist der Anteil von A in Ω).

Starkes Mischen kann auch als eine Aussage über asymptotische Unabhängigkeit gesehen werden. Ist T stark mischend, dann ist $T^{-n}(A)$ schließlich unabhängig von B .

Definition 3.18. Eine maßtreue Abbildung T auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ heißt schwach mischend, falls für jede Wahl von Mengen $A, B \in \mathcal{A}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)| = 0$$

gilt.

Es ist leicht ersichtlich, dass jede stark mischende Abbildung schwach mischend ist. Die Umkehrung gilt i.A. nicht.

Auch die Ergodizität lässt sich mittels dieser Terminologie umformulieren.

Satz 3.19. Sei T eine maßtreue Abbildung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$. Dann sind die folgenden Aussagen äquivalent:

- (a) T ist ergodisch.
- (b) Für jede Wahl von Mengen $A, B \in \mathcal{A}$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B)$$

Beweis: Wir nehmen zunächst an, dass T ergodisch ist und wählen $A, B \in \mathcal{A}$. Dann ist $\mathbf{1}_A$ integrierbar und mit dem Birkhoff Ergodensatz erhalten wir

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A \circ T^i(x) = \mu(A) \quad \text{fast sicher}$$

und damit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(T^i(x)) \mathbf{1}_B(x) = \mu(A) \mathbf{1}_B(x) \quad \text{fast sicher}$$

Da für alle $n \in \mathbb{N}$

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(T^i(x)) \mathbf{1}_B(x) \right| \leq 1$$

folgt aus dem Satz über die majorisierte Konvergenz

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) &= \lim_{n \rightarrow \infty} \int_{\Omega} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(T^i(x)) \mathbf{1}_B(x) \, d\mu(x) \\ &= \int_{\Omega} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(T^i(x)) \mathbf{1}_B(x) \, d\mu(x) \\ &= \int \mu(A) \mathbf{1}_B(x) \, d\mu(x) \\ &= \mu(A) \mu(B) \end{aligned}$$

Wir nehmen nun an, dass (b) gilt und zeigen (a). Sei dazu A T -invariant, d.h. $T^{-1}(A) = A$. Dann gilt

$$\mu(A)^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap A) = \mu(A)$$

und damit $\mu(A) \in \{0, 1\}$. Damit ist (a) gezeigt.

Offensichtlich ist jede schwach mischende Abbildung ergodisch. Die Umkehrung gilt i.A. nicht.

Die Zusammenhänge sind in folgendem Satz zusammengefasst:

Satz 3.20. *Sei T eine maßtreue Abbildung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$.*

1. *Ist T stark mischend, dann ist T schwach mischend.*
2. *Ist T schwach mischend, dann ist T ergodisch.*

Interessanterweise reicht es aus die Mischungseigenschaften auf einem die σ -Algebra erzeugenden Halbring zu zeigen.

Lemma 3.21. *Sei $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und \mathcal{H} ein \mathcal{A} erzeugender Halbring. Für jedes $A \in \mathcal{A}$ und jedes $\varepsilon > 0$ existiert eine Menge H darstellbar als endliche Vereinigung disjunkter Mengen aus \mathcal{H} mit*

$$\mu(A \Delta H) < \varepsilon$$

Beweis: (Skizze)

Sei $A \in \mathcal{A}$ und $\varepsilon > 0$. Nach der Konstruktion des Satzes von Caratheodory existiert eine Menge $H^*(\varepsilon) = \bigcup_{n \in \mathbb{N}} H_n \supseteq A$ mit $(H_n)_{n \in \mathbb{N}} \subseteq \mathcal{H}$ und $\mu(H^*(\varepsilon)) < \mu(A) + \varepsilon/2$ und damit $\mu(H^*(\varepsilon) \setminus A) < \varepsilon/2$. Mit der Stetigkeit von unten gilt

$$\lim_{n \rightarrow \infty} \mu \left(\bigcup_{i=1}^n H_i \right) = \mu \left(\bigcup_{j \in \mathbb{N}} H_j \right) = \mu(H^*(\varepsilon))$$

womit ein $n_0 \in \mathbb{N}$ existiert mit

$$0 \leq \mu(H^*(\varepsilon)) - \mu \left(\bigcup_{i=1}^{n_0} H_i \right) < \varepsilon/2$$

Mittels $H := \bigcup_{i=1}^{n_0} H_i$ erhalten wir schließlich

$$\mu(A \Delta H) = \mu(H \setminus A) + \mu(A \setminus H) \leq \mu(H^*(\varepsilon) \setminus A) + \mu(H^*(\varepsilon) \setminus H) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Aus der Maßtheorie ist bekannt (siehe z.B. Silva [Proposition 2.7.3]), dass die Menge H auch als endliche Vereinigung disjunkter Mengen aus \mathcal{H} geschrieben werden kann.

Lemma 3.22. *Sei T eine maßtreue Abbildung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ und \mathcal{H} ein \mathcal{A} erzeugender Halbring.*

1. *Gilt für jede Wahl von $H_1, H_2 \in \mathcal{H}$*

$$\lim_{n \rightarrow \infty} \mu(T^{-n}(H_1) \cap H_2) = \mu(H_1)\mu(H_2)$$

dann ist T stark mischend.

2. *Gilt für jede Wahl von $H_1, H_2 \in \mathcal{H}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}(H_1) \cap H_2) - \mu(H_1)\mu(H_2)| = 0$$

dann ist T schwach mischend.

3. *Gilt für jede Wahl von $H_1, H_2 \in \mathcal{H}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(H_1) \cap H_2) = \mu(H_1)\mu(H_2)$$

dann ist T ergodisch.

Beweis: (Skizze)

Wir beschränken uns auf den Nachweis von (1). Die beiden anderen Aussagen lassen sich auf analoge Weise zeigen.

Seien $A, B \in \mathcal{A}$ mit positivem Maß. Nach Lemma 3.21 existieren für $\varepsilon > 0$ disjunkte Mengen $I_j \in \mathcal{H}$, $j \in \{1, \dots, p\}$, und disjunkte Mengen $J_k \in \mathcal{H}$, $k \in \{1, \dots, q\}$, mit

$$\mu(A \Delta H_A) < \varepsilon \quad \text{und} \quad \mu(A \Delta H_B) < \varepsilon$$

wobei $H_A = \bigcup_{j=1}^p I_j$ und $H_B = \bigcup_{k=1}^q J_k$. Für jedes $k \in \{1, \dots, q\}$ gilt dann

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \mu(T^{-n}(H_A) \cap J_k) &= \lim_{n \rightarrow \infty} \mu \left(T^{-n} \left(\bigcup_{j=1}^p I_j \right) \cap J_k \right) \\
 &= \lim_{n \rightarrow \infty} \sum_{j=1}^p \mu(T^{-n}(I_j) \cap J_k) \\
 &= \sum_{j=1}^p \lim_{n \rightarrow \infty} \mu(T^{-n}(I_j) \cap J_k) \\
 &= \sum_{j=1}^p \mu(I_j) \mu(J_k) \\
 &= \mu(H_A) \mu(J_k)
 \end{aligned}$$

und auf analoge Art und Weise

$$\lim_{n \rightarrow \infty} \mu(T^{-n}(H_A) \cap H_B) = \mu(H_A) \mu(H_B)$$

Damit existiert ein $n_0 \in \mathbb{N}$ mit

$$|\mu(T^{-n}(H_A) \cap H_B) - \mu(H_A) \mu(H_B)| < \varepsilon$$

für alle $n \geq n_0$. Die Behauptung folgt nun durch mehrmalige Anwendung der Dreiecksungleichung.

Wir haben weiterhin das Ziel zu zeigen, dass unsere 3 Beispiele (Zelt, L_4 und Linksshift) ergodisch sind. Im folgendem Zwischenschritt zeigen wir zunächst, dass der Linksshift stark mischend und damit ergodisch ist.

Beispiel 3.23. Der Linksshift Operator σ ist stark mischend und damit schwach mischend und ergodisch.

Dazu betrachten wir zunächst die Rechtecke

$$\begin{aligned}
 R_A &= \{k_0\} \times \{k_1\} \times \cdots \times \{k_{n_A}\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \dots && \text{und} \\
 R_B &= \{l_0\} \times \{l_1\} \times \cdots \times \{l_{n_B}\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \dots
 \end{aligned}$$

Dann gilt für $n \geq l_{n_B}$

$$\begin{aligned}
 &\sigma^{-n}(R_A) \cap R_B \\
 &= \bigcup_{m_1=1}^N \cdots \bigcup_{m_n=1}^N (\{m_1\} \times \cdots \times \{m_n\} \times R_A) \cap R_B \\
 &= \bigcup_{m_{l_{n_B}+1}=1}^N \cdots \bigcup_{m_n=1}^N \{l_0\} \times \{l_1\} \times \cdots \times \{l_{n_B}\} \times \{m_{l_{n_B}+1}\} \times \cdots \times \{m_n\} \times R_A
 \end{aligned}$$

wobei die rechte Seite einer Vereinigung über disjunkte Mengen entspricht. Damit gilt

$$\begin{aligned}
 & \mu_{\vartheta}(\sigma^{-n}(R_A) \cap R_B) \\
 &= \sum_{m_{l_{n_B}+1}=1}^N \cdots \sum_{m_n=1}^N \mu_{\vartheta}(\{l_0\} \times \{l_1\} \times \cdots \times \{l_{n_B}\} \times \{m_{l_{n_B}+1}\} \times \cdots \times \{m_n\} \times R_A) \\
 &= \sum_{m_{l_{n_B}+1}=1}^N \cdots \sum_{m_n=1}^N \left(\prod_{i_B=0}^{n_B} \vartheta(\{l_{i_B}\}) \right) \left(\prod_{j=l_{n_B}+1}^n \vartheta(\{m_j\}) \right) \left(\prod_{i_A=0}^{n_A} \vartheta(\{k_{i_A}\}) \right) \\
 &= \left(\prod_{i_B=0}^{n_B} \vartheta(\{l_{i_B}\}) \right) \left(\prod_{i_A=0}^{n_A} \vartheta(\{k_{i_A}\}) \right) \\
 &= \mu_{\vartheta}(R_A) \mu_{\vartheta}(R_B)
 \end{aligned}$$

Mit Lemma 3.22 folgt die Behauptung.

Bevor wir mit Hilfe des Linksshiftoperators σ zeigen, dass auch die Zelt-Abbildung t und die logistische Abbildung L_4 stark mischend bezügl. den entsprechenden invarianten Maßen λ bzw. $f_* d\lambda^\dagger$ sind, betrachten wir ein Beispiel, das Sie (hoffentlich) schon aus den Analysis Grundvorlesungen kennen, die dyadische Transformation $D_2 : [0, 1) \rightarrow [0, 1)$, definiert durch

$$D_2(x) = 2x \pmod{1}.$$

Wir überlegen uns der Reihe nach einige Eigenschaften von D_2 und ihren Zusammenhang mit dem Linksshiftoperator σ auf Σ_2 :

- D_2 ist maßtreu auf $([0, 1), \mathcal{B}([0, 1)), \lambda)$ da für jedes $x \in [0, 1)$

$$D_2^{-1}([0, x]) = \{z \in [0, 1) : D_2(z) \in [0, x]\} = [0, \frac{x}{2}] \cup [\frac{1}{2}, \frac{x+1}{2}],$$

und daher auch

$$\lambda^{D_2}([0, x]) = \lambda([0, \frac{x}{2}] \cup [\frac{1}{2}, \frac{x+1}{2}]) = x$$

gilt. Zusätzlich gilt offensichtlich $D_2^n(x) = 2^n x \pmod{1}$ für jedes $n \in \mathbb{N}$.

- Wir setzen $I_0 = [0, \frac{1}{2}), I_1 = [\frac{1}{2}, 1)$ und codieren den D_2 -Orbit von $x \in [0, 1)$ mittels I_0, I_1 :

$$h(x) := \mathbf{k} \in \Sigma_2 \iff D_2^n(x) \in I_{k_n} \text{ für jedes } n \in \mathbb{N}_0$$

Die damit induzierte Abbildung $h : [0, 1) \rightarrow \Sigma_2$ ist offensichtlich wohldefiniert und liefert genau die Binärentwicklung von x . Obwohl die Transformation h bekannt sein sollte, diskutieren wir hier kurz ihre wichtigsten Eigenschaften und überlegen uns zuerst, wie $h^{-1}(R)$ für dünne Rechtecke $R \in \mathfrak{h}_2$ aussieht. Zwecks Vereinfachung der Notation schreiben wir kurz

$$\{k_0\} \times \{k_1\} \times \cdots \times \{k_n\} \times \star := \{k_0\} \times \{k_1\} \times \cdots \times \{k_n\} \times \{1, \dots, N\} \times \{1, \dots, N\} \times \cdots,$$

[†]soll heißen: invariant bezüglich des Wahrscheinlichkeitsmaßes μ das absolut stetig bezügl. λ mit Dichte f_* ist

wobei der Stern andeuten soll, dass es ab der $n+2$ -ten Koordinaten keine Einschränkungen gibt. Damit erhalten wir nun offensichtlich für $n = 0$

$$h^{-1}(\{0\} \times \star) = [0, \frac{1}{2}), \quad h^{-1}(\{1\} \times \star) = [\frac{1}{2}, 1),$$

also

$$h^{-1}(\{k_0\} \times \star) = [\frac{k_0}{2}, \frac{k_0}{2} + \frac{1}{2}) = [\frac{k_0}{2}, \frac{k_0}{2} + \frac{1}{2^{n+1}}) =: I_0(\mathbf{k}).$$

Für $n = 1$ gilt

$$\begin{aligned} h^{-1}(\{0\} \times \{0\} \times \star) &= [0, \frac{1}{4}), & h^{-1}(\{0\} \times \{1\} \times \star) &= [\frac{1}{4}, \frac{1}{2}), \\ h^{-1}(\{1\} \times \{0\} \times \star) &= [\frac{1}{2}, \frac{3}{4}), & h^{-1}(\{1\} \times \{1\} \times \star) &= [\frac{3}{4}, 1), \end{aligned}$$

und damit

$$h^{-1}(\{k_0\} \times \{k_1\} \times \star) = [\frac{k_0}{2} + \frac{k_1}{4}, \frac{k_0}{2} + \frac{k_1}{4} + \frac{1}{4}) = [\frac{k_0}{2} + \frac{k_1}{4}, \frac{k_0}{2} + \frac{k_1}{4} + \frac{1}{2^{n+1}}) =: I_1(\mathbf{k}).$$

Analog folgt für jedes feste $n \in \mathbb{N}_0$ der folgende Zusammenhang:

$$h^{-1}(\{k_0\} \times \{k_1\} \cdots \times \{k_n\} \times \star) = \left[\sum_{i=0}^n \frac{k_i}{2^{i+1}}, \sum_{i=0}^n \frac{k_i}{2^{i+1}} + \frac{1}{2^{n+1}} \right) =: I_n(\mathbf{k}) \quad (3.2)$$

Nachdem der Halbring \mathfrak{h}_2 aller dünnen Rechtecke ein Erzeuger von $\mathcal{B}(\Sigma_2)$ ist, und offensichtlich $I_n(\mathbf{k}) \in \mathcal{B}([0, 1])$ für jedes $n \in \mathbb{N}_0$ und jedes $\mathbf{k} \in \Sigma_2$ gilt, ist die Transformation $h : ([0, 1], \mathcal{B}([0, 1])) \rightarrow (\Sigma_2, \mathcal{B}(\Sigma_2))$ messbar.

Bezeichnen wir mit ϑ_2 die Gleichverteilung auf $[0, 1]$ und mit μ_{ϑ_2} das induzierte Wahrscheinlichkeitsmaß auf $\mathcal{B}(\Sigma_2)$ gemäß Satz 2.11, dann folgt weiters

$$\begin{aligned} \mu_{\vartheta_2}(\{k_0\} \times \{k_1\} \cdots \times \{k_n\} \times \star) &= \frac{1}{2^{n+1}} = \lambda(I_n(\mathbf{k})) \\ &= \lambda(h^{-1}(\{k_0\} \times \{k_1\} \cdots \times \{k_n\} \times \star)) \\ &= \lambda^h(\{k_0\} \times \{k_1\} \cdots \times \{k_n\} \times \star) \end{aligned}$$

für jedes $\mathbf{k} \in \Sigma_2$ und für jedes $n \in \mathbb{N}_0$, es gilt also $\lambda^h = \mu_{\vartheta_2}$.

- Für jedes $\mathbf{k} \in \Sigma_2$ ist die Folge $(I_n(\mathbf{k}))_{n \in \mathbb{N}_0}$ von Intervallen gemäß Gleichung (3.2) strikt fallend, ihr Durchschnitt ist höchstens einpunktig und es gilt

$$\bigcap_{n=0}^{\infty} I_n(\mathbf{k}) = \sum_{i=0}^{\infty} \frac{k_i}{2^{i+1}} =: g(\mathbf{k}).$$

Die damit induzierte Abbildung $g : \Sigma_2 \rightarrow [0, 1]$ erfüllt $g \circ h = id_{[0,1]}$: Für $x \in [0, 1]$ mit $h(x) = \mathbf{k}$ folgt nämlich $x \in h^{-1}(\{k_0\} \times \{k_1\} \cdots \times \{k_n\} \times \star)$ für jedes $n \in \mathbb{N}_0$ und damit $x = g(\mathbf{k})$, also $g(h(x)) = x$. Daraus folgt sofort die Injektivität von h und die Surjektivität von g .

- Sei $x \in [0, 1)$ und gelte $\mathbf{k} = h(x)$. Dann folgt (per definitionem) $D_2^n(x) \in I_{k_n}$ für jedes $n \in \mathbb{N}_0$, und damit natürlich auch $D_2^n(D_2x) \in I_{k_{n+1}}$ für jedes $n \in \mathbb{N}_0$. Letzteres ist aber gleichbedeutend mit $h(D_2x) = \sigma(\mathbf{k}) = \sigma(h(x))$. Mit anderen Worten, das folgende Diagramm ist kommutativ, i.e., es gilt $h \circ D_2 = \sigma \circ h$ (Pfeile mit Doppelspitze symbolisieren Injektivität):

$$\begin{array}{ccc} [0, 1) & \xrightarrow{D_2} & [0, 1) \\ h \downarrow & & \downarrow h \\ \Sigma_2 & \xrightarrow{\sigma} & \Sigma_2 \end{array} \quad (3.3)$$

- Die Abbildung g ist zwar surjektiv, aber nicht injektiv; beispielsweise gilt für $\mathbf{k} = (0, 1, 1, \bar{1})$ und $\mathbf{l} = (1, 0, 0, \bar{0})$ offensichtlich $g(\mathbf{k}) = g(\mathbf{l}) = \frac{1}{2}$. Nichtsdestotrotz ist g aber zumindest injektiv außerhalb einer Nullmenge: Wenn wir mit N die Menge aller Elemente von Σ_2 bezeichnen, die schließlich konstant[†] sind, dann ist N abzählbar unendlich, messbar, erfüllt $\mu_{\vartheta_2}(N) = 0$ und g ist injektiv auf $\Sigma_2 \setminus N$. Weiters gilt offensichtlich $\sigma^{-1}(N) = N$.

Definieren wir andererseits die Menge M als

$$M = \left\{ \frac{i}{2^n} : n \in \mathbb{N}, i \in \{0, 1, \dots, 2^n - 1\} \right\},$$

dann ist M abzählbar unendlich, erfüllt daher $\lambda(M) = 0$, und es gilt $D_2^{-1}(M) = M$. Weiters ist die Abbildung $h : [0, 1) \setminus M \rightarrow \Sigma_2 \setminus N$ sogar bijektiv, ihre Inverse entspricht genau g (eingeschränkt auf $\Sigma_2 \setminus N$).

I.e.: die zwei dynamischen Systeme $([0, 1), \mathcal{B}([0, 1)), \lambda, D_2)$ und $(\Sigma_2, \mathcal{B}(\Sigma), \sigma, \mu_{\vartheta_2})$ entsprechen einander zumindest außerhalb einer Nullmenge (also auf einer 1-Menge) und können daher als dynamisch gleich oder isomorph betrachtet werden.

Wir werden in der Folge zeigen und verwenden, dass isomorphe dynamische Systeme auch die gleichen Mischungseigenschaften haben, und die Mischungseigenschaften von t und L_4 von jenen des Shiftoperators ableiten.

Definition 3.24 (Isomorphie dynamischer System).

Zwei dynamische Systeme $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ heißen isomorph genau dann wenn die folgenden Aussagen gelten:

1. Es existieren Menge $\Lambda_1 \in \mathcal{A}_1, \Lambda_2 \in \mathcal{A}_2$ mit $\mu_1(\Lambda_1) = 1 = \mu_2(\Lambda_2)$ und $T_1(\Lambda_1) \subseteq \Lambda_1$ und $T_2(\Lambda_2) \subseteq \Lambda_2$.
2. Es existiert eine invertierbare maßtreue Transformation[†] $h : \Lambda_1 \rightarrow \Lambda_2$, sodass das folgende Diagramm kommutativ ist:

$$\begin{array}{ccc} \Lambda_1 & \xrightarrow{T_1} & \Lambda_1 \\ h \downarrow & & \downarrow h \\ \Lambda_2 & \xrightarrow{T_2} & \Lambda_2 \end{array} \quad (3.4)$$

[†]i.e., die Menge aller \mathbf{k} für die ein Index $n_0 = n_0(\mathbf{k})$ existiert, sodass k_n konstant für alle $n \geq n_0$

[†]formal sauber: Die Transformation h ist messbar bezüglich der Spur- σ -Algebren $\Lambda_1 \cap \mathcal{A}_1$ und $\Lambda_2 \cap \mathcal{A}_2$ und es gilt $\mu_1^{T_1} = \mu_2$

Mit den obigen Vorüberlegungen haben wir also folgendes Resultat gezeigt:

Satz 3.25. *Das dynamische System $([0, 1), \mathcal{B}([0, 1)), \lambda, D_2)$ ist isomorph zu $(\Sigma_2, \mathcal{B}(\Sigma_2), \mu_{\vartheta_2}, \sigma)$.*

Isomorphe dynamische System haben die selben Mischungseigenschaften - es gilt das folgende Theorem:

Satz 3.26. *Angenommen die dynamische Systeme $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ sind isomorph, dann gilt:*

- $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ ist stark mischend genau dann, wenn $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ stark mischend ist.
- $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ ist schwach mischend genau dann, wenn $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ schwach mischend ist.
- $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ ist ergodisch genau dann, wenn $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ ergodisch ist.

Beweis: Wir beweisen nur die erste Aussage, die Beweise der anderen beiden verlaufen analog. Beachten Sie im Beweis, wie hier die Mengen Λ_i in's Spiel gebracht werden.

Angenommen $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ ist stark mischend, wir wollen zeigen, dass dann auch das System $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ stark mischend ist, i.e., dass für beliebige $A, B \in \mathcal{A}_2$

$$\lim_{n \rightarrow \infty} \mu_2(T_2^{-n}(A) \cap B) = \mu_2(A)\mu_2(B)$$

gilt. Unter Verwendung von

$$\begin{aligned} \mu_2(T_2^{-n}(A) \cap B) &= \mu_1^h(T_2^{-n}(A \cap \Lambda_2) \cap B \cap \Lambda_2) = \mu_1(h^{-1}(T_2^{-n}(A \cap \Lambda_2)) \cap h^{-1}(B \cap \Lambda_2)) \\ &= \mu_1\left(T_1^{-n}\left(\underbrace{h^{-1}(A \cap \Lambda_2)}_{=: C \in \mathcal{A}_1 \cap \Lambda_1} \cap \underbrace{h^{-1}(B \cap \Lambda_2)}_{=: D \in \mathcal{A}_1 \cap \Lambda_1}\right)\right) \end{aligned}$$

liefert die Tatsache, dass T_1 stark mischend ist, sofort

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu_2(T_2^{-n}(A) \cap B) &= \lim_{n \rightarrow \infty} \mu_1(T_1^{-n}(C) \cap D) = \mu_1(C)\mu_1(D) \\ &= \mu_1(h^{-1}(A \cap \Lambda_2)) \cdot \mu_1(h^{-1}(B \cap \Lambda_2)) = \mu_2(A \cap \Lambda_2) \cdot \mu_2(B \cap \Lambda_2) \\ &= \mu_2(A)\mu_2(B) \end{aligned}$$

und die erste Aussage folgt.

Satz 3.27. *Das dynamische System $([0, 1), \mathcal{B}([0, 1)), \lambda, D_2)$ ist stark mischend.*

Wenn wir nochmals in Ruhe auf den Beweis von Satz 3.26 blicken, so sehen wir, dass für den Beweis der ersten Implikation (T_1 stark mischend impliziert T_2 stark mischend) nur verwendet wurde, dass $h : \Lambda_1 \rightarrow \Lambda_2$ maßtreu ist (i.e., dass $\mu_1^h = \mu_2$ gilt). Erst für den Beweis der Rückrichtung (T_2 stark mischend impliziert T_1 stark mischend) benötigt man die Bijektivität von h , weil in diesem Fall dann $h^{-1} : \Lambda_2 \rightarrow \Lambda_1$ ebenfalls maßtreu ist. Dies führt zu folgender schwächeren Form von Isomorphie.

Definition 3.28 (Faktor dynamischer System).

Ein dynamisches System $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ heißt Faktor des dynamischen Systems $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ genau dann, wenn die folgenden Aussagen gelten:

1. Es existieren Mengen $\Lambda_1 \in \mathcal{A}_1, \Lambda_2 \in \mathcal{A}_2$ mit $\mu_1(\Lambda_1) = 1 = \mu_2(\Lambda_2)$ und $T_1(\Lambda_1) \subseteq \Lambda_1$ und $T_2(\Lambda_2) \subseteq \Lambda_2$.
2. Es existiert eine maßtreue Transformation[†] $h : \Lambda_1 \rightarrow \Lambda_2$, sodass das folgenden Diagramm kommutativ ist:

$$\begin{array}{ccc} \Lambda_1 & \xrightarrow{T_1} & \Lambda_1 \\ \downarrow h & & \downarrow h \\ \Lambda_2 & \xrightarrow{T_2} & \Lambda_2 \end{array} \quad (3.5)$$

Wir haben also implizit schon folgendes Resultat bewiesen:

Satz 3.29. Sei $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ ein Faktor von $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$, dann gilt:

- Ist $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ stark mischend, dann auch $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$.
- Ist $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ schwach mischend, dann, wenn $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ schwach mischend ist.
- Ist $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ ergodisch, dann auch $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ ergodisch ist.

Als finalen Schritt betreffend Mischungseigenschaften zeigen wir jetzt, dass die Zelt-Abbildung t und die logistische Abbildung L_4 stark mischend sind.

Satz 3.30. $([0, 1), \mathcal{B}([0, 1)), \lambda, t)$ ist ein Faktor von $([0, 1), \mathcal{B}([0, 1)), \lambda, D_2)$ und als solches stark mischend.

Beweis: Setzen wir $h = t$ dann ist $h : [0, 1) \rightarrow [0, 1)$ maßtreu (und nicht invertierbar). Für jedes $x \in [0, 1)$ ist leicht nachzurechnen, dass $h \circ D_2(x) = t \circ h(x)$ gilt, womit die erste Aussagen bewiesen ist. Die zweite Aussage folgt sofort via Satz 3.29.

Bemerkung 3.31. Die im Zusammenhang mit D_2 verwendete Idee der Codierung von Orbits mittels der Partition $I_0 = [0, \frac{1}{2}), I_1 = [\frac{1}{2}, 1)$ lässt sich auch auf t übertragen, und es kann gezeigt werden, dass $([0, 1), \mathcal{B}([0, 1)), \lambda, t)$ sogar isomorph zu $([0, 1), \mathcal{B}([0, 1)), \lambda, D_2)$. Wir sind oben den Weg über den Faktor gegangen, um, erstens, Faktoren von dynamischen Systemen kennenzulernen und, zweitens, weil die Konstruktion einer direkten Isomorphie (also ohne ‘Umweg’ über Σ_2 wesentlich aufwändiger ist).

Satz 3.32. $([0, 1), \mathcal{B}([0, 1)), f_*d\lambda, L_4)$ ist isomorph zu $([0, 1), \mathcal{B}([0, 1)), \lambda, t)$ und daher stark mischend.

Beweis: Wir haben schon im zweiten Abschnitt (siehe Gleichung 2.4) gezeigt, dass die reskalierten Versionen \hat{t} und \hat{L}_4 der Zelt-Abbildung und der logistischen Abbildung topologisch konjugiert sind und haben darüber das \hat{L}_4 -invariante Maß $\mu_{f_*}^\dagger$ konstruiert. Die behauptete Isometrie von $([0, 1), \mathcal{B}([0, 1)), f_*d\lambda, L_4)$ und $([0, 1), \mathcal{B}([0, 1)), \lambda, t)$ folgt daher sofort. Die zweite Aussage ist abermals eine direkte Konsequenz von Satz 3.29.

[†]formal sauber: Die Transformation h ist messbar bezüglich der Spur- σ -Algebren $\Lambda_1 \cap \mathcal{A}_1$ und $\Lambda_2 \cap \mathcal{A}_2$ und $\mu_1^{T_1} = \mu_2$

[†]also das Maß mit Dichte f_* bezügl. λ auf $[0, 1)$

Bemerkung 3.33. Bemerkung 3.31 und Satz 3.32 implizieren, dass auch die logistische Abbildung L_4 isomorph zum Shiftoperator σ ist, genauer formuliert: die dynamischen Systeme $([0, 1), \mathcal{B}([0, 1)), f_*d\lambda, L_4)$ und $(\Sigma_2, \mathcal{B}(\Sigma_2), \mu_{\vartheta_2}, \sigma)$ sind isomorph.

Bevor wir diskutieren, wie die in Abschnitt 1 formulierten Eigenschaften über geschachtelte Quadratwurzeln von 2 mit Hilfe der bisher in der Vorlesung bewiesenen Resultate hergeleitet werden können, rekapitulieren wir und überlegen uns, was die Ergodizität von D_2, t, L_4 für Konsequenzen für die Orbits ‘typischer’ Punkte hat. Wir betrachten dabei zuerst D_2 , die erhaltenen Aussagen gelten aber auch für t und L_4 .

Aus Analysis I ist bekannt, dass die euklidische Topologie \mathcal{O} in $[0, 1)$ (bzw. in ganz \mathbb{R}) eine abzählbare Basis hat: \mathbb{Q}^2 ist abzählbar, die Menge aller offenen, nichtleeren Intervalle $I_n := (a_n, b_n)$ mit $a_n, b_n \in \mathbb{Q} \cap [0, 1)$ und $a_n < b_n$ ist daher ebenfalls abzählbar. Offensichtlich existiert für jedes $x \in [0, 1)$ und jede offenen Menge $U \subseteq [0, 1)$ mit $x \in U$ ein Intervall (a_n, b_n) mit $x \in (a_n, b_n) \subseteq U$, die Familie $((a_n, b_n))_{n \in \mathbb{N}}$ ist daher eine abzählbare Basis der Topologie.

Für jedes $k \in \mathbb{N}$ gilt $\lambda(I_k) = b_k - a_k > 0$, der Birkoff’sche Ergodensatz liefert daher die Existenz einer Menge $\Lambda_k \in \mathcal{B}([0, 1))$ mit $\lambda(\Lambda_k) = 1$, sodass für jedes $x \in \Lambda_k$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{I_k}(D_2^i(x)) = \int_{[0,1)} \mathbf{1}_{I_k} d\lambda = \lambda(I_k) > 0$$

gilt. Insbesondere ‘besucht’ x via D_2 die Menge I_k unendlich oft (warum?). Setzen wir $\Lambda = \bigcap_{k=1}^{\infty} \Lambda_k$ dann folgt sofort $\Lambda \in \mathcal{B}([0, 1))$ und $\lambda(\Lambda) = 1$. Jedes $x \in \Lambda$ ‘besucht’ daher via D_2 jedes I_k unendlich oft. Nachdem $(I_k)_{k \in \mathbb{N}}$ eine abzählbare Basis der Topologie ist, und jedes I_k von einem solchen x unendlich oft besucht wird, bedeutet dies insbesondere: Für $x \in \Lambda$ ist der Orbit $O_{D_2}(x)$ dicht in $[0, 1)$. Mit anderen Worten: λ -fast jeder Punkt $x \in [0, 1)$ hat einen dichten D_2 Orbit. Wir haben also folgendes Resultat bewiesen.

Satz 3.34. λ -fast jeder Punkt $x \in [0, 1)$ hat einen dichten D_2 -, einen dichten t - und einen dichten L_4 -Orbit.

Wir schließen dieses Kapitel nun, wie angekündigt, mit den in Abschnitt 1 erwähnten Resultaten betreffend geschachtelte Quadratwurzeln aus 2 ab und rufen uns folgende Eigenschaften in Erinnerung:

(S1) $\hat{t} : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$ und $\hat{L}_4 : [-2, 2] \rightarrow [-2, 2]$ sind definiert durch

$$\hat{t}(x) = \begin{cases} 2x + \frac{\pi}{2} & \text{if } x \in [-\frac{\pi}{2}, 0] \\ -2x + \frac{\pi}{2} & \text{if } x \in [0, \frac{\pi}{2}], \end{cases}$$

und $\hat{L}_4(x) = -x^2 + 2$.

(S2) Definieren wir $h : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-2, 2]$ durch $h(x) = 2 \sin(x)$, dann ist h ein Homöomorphismus und das folgende Diagramm kommutativ, i.e., es gilt $h \circ \hat{t} = \hat{L}_4 \circ h$:

$$\begin{array}{ccc} [-\frac{\pi}{2}, \frac{\pi}{2}] & \xrightarrow{\hat{t}} & [-\frac{\pi}{2}, \frac{\pi}{2}] \\ \downarrow h & & \downarrow h \\ [-2, 2] & \xrightarrow{\hat{L}_4} & [-2, 2] \end{array} \tag{3.6}$$

Anders ausgedrückt, für jedes $x \in [-2, 2]$ gilt die Gleichheit

$$h \circ \hat{t} \circ h^{-1}(x) = \hat{L}_4(x). \quad (3.7)$$

Für die nächsten Schritte definieren wir die Abbildungen $w_{-1} : [-2, 2] \rightarrow [-2, 0]$ und $w_1 : [-2, 2] \rightarrow [0, 2]$ durch $w_{-1}(x) = -\sqrt{2-x}$ and $w_1(x) = \sqrt{2-x}$ und setzen $J_{-1} = [-2, 0]$ sowie $J_1 = (0, 2]$. Weiters definieren wir $\hat{\Sigma}_2 := \{-1, 1\}^{\mathbb{N}_0}$ um die selbe Notation wie im Originalpaper [2] zu verwenden[†]. Der Schlüssel zu den Resultaten in [2] ist folgende, einfach nachzurechnende Relation:

$$x = \begin{cases} w_{-1}(\hat{L}_4(x)) & \text{falls } x \in J_{-1}, \\ w_1(\hat{L}_4(x)) & \text{falls } x \in J_1 \end{cases} \quad (3.8)$$

Tatsächlich gilt für $x \in J_{-1}$

$$w_{-1} \circ \hat{L}_4(x) = -\sqrt{2 - \hat{L}_4(x)} = -\sqrt{x^2} = -|x| = x,$$

und für $x \in J_1$

$$w_1 \circ \hat{L}_4(x) = \sqrt{2 - \hat{L}_4(x)} = \sqrt{x^2} = |x| = x.$$

Codieren wir nun den \hat{t} -Orbit von $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ mittels $I_{-1} = [-\frac{\pi}{2}, 0]$, $I_1 = (0, \frac{\pi}{2}]$ und den \hat{L}_4 -Orbit von $x \in [-2, 2]$ mittels J_{-1}, J_1 und bezeichnen die entsprechende Codierung mit $C_{\hat{t}} : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow \hat{\Sigma}_2$ bzw. $C_{\hat{L}_4} : [-2, 2] \rightarrow \hat{\Sigma}_2$, dann gilt offensichtlich

$$C_{\hat{t}}(x) = C_{\hat{L}_4} \circ h(x) \quad (3.9)$$

Für $C_{\hat{t}}(x) = (k_0, k_1, k_2, \dots) \in \hat{\Sigma}_2$ folgt daher mit Gleichung (3.8) und dem kommutativen Diagramm (3.6)

$$\begin{aligned} h(x) &= w_{k_0}(\hat{L}_4 \circ h(x)) = w_{k_0}(h \circ \hat{t}(x)) = w_{k_0} \circ w_{k_1}(\hat{L}_4 \circ h \circ \hat{t}(x)) \\ &= w_{k_0} \circ w_{k_1}(\hat{L}_4 \circ \hat{L}_4 \circ h(x)) = w_{k_0} \circ w_{k_1}((\hat{L}_4)^2 \circ h(x)). \end{aligned}$$

Mehrmaliges Anwenden derselben Idee liefert nach n Schritten

$$\begin{aligned} h(x) &= w_{k_0} \circ w_{k_1} \circ \dots \circ w_{k_n}((\hat{L}_4)^{n+1}(h(x))) \\ &= k_0 \sqrt{2 - k_1 \sqrt{2 - k_2 \sqrt{2 - k_3 \sqrt{\dots - k_n \sqrt{2 - (\hat{L}_4)^{n+1}(h(x))}}}}} \end{aligned} \quad (3.10)$$

Der letzte technische Schritt ist zu zeigen, dass die verschachtelte Quadratwurzel in Gleichung (3.10) auch tatsächlich konvergiert, selbiges lässt sich abermals problemlos durch Übersetzung auf \hat{t} bewerkstelligen, die Details können in [2] (Stichwort: Address map) nachgelesen werden.

Bezeichnen wir mit $\vartheta_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$ die stetige Gleichverteilung auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$, dann folgt aus der Tatsache, dass die Zelt-Abbildung isomorph zum σ auf Σ_2 (und damit auch auf $\hat{\Sigma}_2$) ist, dass für $\vartheta_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$ -fast jedes $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ die Folge $C_{\hat{t}}(x) = \mathbf{k} \in \hat{\Sigma}_2$ normal ist, i.e., es gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{1\}}(k_i) = \frac{1}{2} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{-1\}}(k_i).$$

[†]Die Linksshift-Operatoren auf den entsprechenden Räumen sind natürlich isomorph

Aus Gleichung (3.9) und der Tatsache, dass der Homöomorphismus h absolut stetig ist, folgt daher, dass für $\vartheta_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$ -fast jedes $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ die Folge $C_{\hat{L}_4}(h(x)) = \mathbf{k} \in \hat{\Sigma}_2$ normal ist. Damit haben wir also folgendes Theorem bewiesen:

Satz 3.35 ([2]). *Für $\vartheta_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$ -fast jedes $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ ist die \hat{L}_4 -Codierung $C_{\hat{L}_4}(h(x)) = \mathbf{k} \in \{-1, 1\}^{\mathbb{N}_0}$ von $h(x)$ normal und es gilt*

$$2 \sin(x) = \lim_{n \rightarrow \infty} k_0 \sqrt{2 - k_1 \sqrt{2 - k_2 \sqrt{2 - k_3 \sqrt{\dots - k_n \sqrt{2}}}}} \quad (3.11)$$

Kapitel 4

Entropie

4.1 Zufall und Information

Die Entropie eines dynamischen Systems kann intuitiv als eine Zahl beschrieben werden, die den Grad der *Unvorhersehbarkeit / Unsicherheit* im System ausdrückt.

Münzwurf I: Angenommen, wir haben eine faire Münze mit den möglichen Ereignissen Wappen (1) und Zahl (2), welche beide mit Wahrscheinlichkeit $1/2$ eintreten; Grundraum $\{1, 2\}$. Beim Wurf von zwei Münzen entspricht unser Grundraum $\{1, 2\} \times \{1, 2\}$ und jedes der vier möglichen Ereignisse tritt mit Wahrscheinlichkeit $(1/2)^2 = 1/4$ ein. Beim Wurf von n Münzen entspricht unser Grundraum $\{1, 2\}^n$ und jedes der 2^n möglichen Ereignisse tritt mit Wahrscheinlichkeit $(1/2)^n$ ein. Werfen wir die Münze unendlich oft (mit Beginn bei Zeitpunkt 0 = Gegenwart und Zukunft = $\{1, 2, \dots\}$), dann können wir den Grundraum

$$\Omega = \Sigma_2 = \{1, 2\}^{\mathbb{N}_0}$$

mit zugehöriger Borel'scher σ -Algebra $\mathcal{B}(\Sigma_2)$ und dem Produktmaß μ_{ϑ_2} mit

$$\mu_{\vartheta_2}(R) = \prod_{i=0}^n \vartheta_2(\{k_i\})$$

für jedes dünne Rechteck $R \in \mathfrak{h}_2$ der Form $R = \{k_0\} \times \{k_1\} \times \dots \times \{k_n\} \times \{1, 2\} \times \{1, 2\} \times \dots$ zur Modellierung verwenden, wobei $\vartheta_2(\{1\}) = 1/2 = \vartheta_2(\{2\})$ [vgl. Notation in Kapitel 2].

Für $\mathbf{k} = (k_0, k_1, k_2, \dots) \in \Sigma_2$ entspricht k_0 dem Resultat des Münzwurfs in der Gegenwart und k_n dem Resultat des Münzwurfs zum Zeitpunkt n .

Betrachten wir ein weiteres Element $\mathbf{l} = (l_0, l_1, l_2, \dots) \in \Sigma_2$, welches mittels $l_n = k_{n+1}$ für alle $n \in \mathbb{N}_0$ mit \mathbf{k} verbunden ist; \mathbf{l} entspricht \mathbf{k} um eine Zeiteinheit in die Zukunft verschoben. Dies legt ein dynamisches System nahe, das jede Sequenz nach links verschiebt (Linksshift), d.h.

$$\sigma(\mathbf{k}) = \mathbf{l}$$

Wie wir bereits gesehen haben, ist der Linksshift ergodisch.

[Durchdenken Sie das obige Beispiel mit beliebig vielen $N \in \mathbb{N}$ anstatt mit $N = 2$ gleichwahrscheinlichen Ausgängen.]

Obwohl das eben beschriebene dynamische System vollkommen deterministisch ist, besteht eine gewisse *Unsicherheit* das Verhalten des dynamischen Systems [für $N > 1$] aus unvollkommenem Wissen vorherzusagen. (Formal spiegelt sich dies in der Tatsache wider, dass die Entropie eines Linksshifts [für $N > 1$] ungleich Null ist.)

Um den Begriff Entropie zu verstehen, ist es wichtig, Unvorhersehbarkeit aus einem anderen Blickwinkel zu betrachten.

Münzwurf II: Angenommen, wir haben eine faire Münze 1000 Mal geworfen. Die Kenntnis der Ergebnisse dieser Würfe hilft uns nicht den nächsten Wurf vorherzusagen: Es gibt immer noch eine 50%-ige Chance auf Wappen und eine 50%-ige Chance auf Zahl. Formaler: Sei $f : \Sigma_2 \rightarrow \{1, 2\}$ eine Funktion gegeben durch

$$f(\mathbf{k}) := k_{1000}$$

Dann liefern die ersten 1000 Würfe k_0, \dots, k_{999} keine Information über den Ausgang des Wurfs zum Zeitpunkt 1000. Damit ist der Linksshift σ zu einem gewissen Grad *unvorhersehbar*.

4.2 Definition und Eigenschaften

Nachdem wir Entropie als *Begriff* motiviert haben, wollen wir nun noch die mathematische Definition der Entropie motivieren; vgl. dazu Kapitel 2 in [3].

Zunächst allgemein: Für eine maßerhaltende Transformation T auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mu)$ wollen wir eine nichtnegative Größe $H = h(T)$ definieren, welche die Unsicherheit darüber misst, wohin T die Punkte von Ω verschiebt. Der Wert von $h(T)$ soll dabei den Wert an *Zufälligkeit* (im Sinne von *Unvorhersehbarkeit* / *Unsicherheit*) widerspiegeln, der durch T erzeugt wird. Wir wollen $h(T)$ so definieren, dass

- (i) die Menge an Informationen, die durch eine Anwendung von T gewonnen wird, proportional zur Menge der entfernten Unsicherheit ist, und dass
- (ii) $h(T)$ isomorphismusinvariant ist, d.h. isomorphen Transformationen wird die gleiche Entropie zugeordnet.

Nun genauer: Angenommen, wir interessieren uns für die Lage eines Punktes ω im Grundraum Ω .

- Wird Ω in 2 Mengen mit gleichem Maß aufgeteilt und wissen wir, zu welcher dieser Mengen der Punkt ω gehört, dann haben wir 1 Bit an Information über die Lage von ω . (Die beiden Hälften von Ω können mit 0 und 1 codiert werden.)
- Wird Ω in 4 Mengen mit gleichem Maß aufgeteilt und wissen wir, zu welcher dieser Mengen der Punkt ω gehört, dann haben wir 2 Bits an Information über die Lage von ω . (Die vier Viertel von Ω können mit 00, 01, 10 und 11 codiert werden.)
- Allgemeiner: Wird Ω in 2^k Mengen mit gleichem Maß aufgeteilt und wissen wir, zu welcher dieser Mengen der Punkt ω gehört, dann haben wir k Bits an Information über die Lage von ω .

- Die vorangegangene Beobachtung legt folgende Interpretation nahe: Wird Ω in n Mengen mit gleichem Maß aufgeteilt und wissen wir, zu welcher dieser Mengen der Punkt ω gehört, dann haben wir $\log_2(n)$ Bits an Information über die Lage von ω .
- Es besteht allerdings keine Notwendigkeit, Ω tatsächlich in Stücke zu teilen: Betrachten wir dazu eine (messbare) Teilmenge A von Ω mit $\mu(A) = 1/n$. Wissen wir, dass ω zu A gehört, dann haben wir $\log_2(n)$ Bits an Information über die Lage von ω . Gilt $\mu(A) = p > 0$ und wissen wir, dass ω zu A gehört, dann haben wir $\log_2(1/p) = -\log_2(p)$ Bits an Information über die Lage von ω .
- Nehmen wir nun an, Ω sei in endlich viele Teilmengen A_1, \dots, A_m mit zugehörigen Wahrscheinlichkeiten p_1, \dots, p_m unterteilt, d.h.

$$p_i \geq 0 \quad \text{und} \quad \sum_{i=1}^m p_i = 1$$

Dann entpricht

$$\sum_{i=1}^m p_i \log_2(1/p_i) = -\sum_{i=1}^m p_i \log_2(p_i)$$

der Menge an Information, die wir (im Sinne der Wahrscheinlichkeitstheorie) über die Lage von ω *erwarten* können.

Sei $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $\mathcal{S} = \{A_1, \dots, A_m\}$ eine Partition von Ω in endlich viele messbare Mengen mit zugehörigen Wahrscheinlichkeiten $p_1, \dots, p_m \in [0, 1]$ (jede der Mengen A_1, \dots, A_m nennen wir *Atom* von \mathcal{S}). Dann definieren wir die *Entropie* dieser Zerlegung mittels

$$H(\mathcal{S}) := H(p_1, \dots, p_m) := -\sum_{i=1}^m p_i \log(p_i) \quad (4.1)$$

Konvention: Für $p = 0$ setzen wir $p \log(p) := 0$.

Da die Entropie aus der Informationstheorie stammt (Informationen werden über binäre Sequenzen übertragen), bevorzugen einige Autoren die Verwendung des Logarithmus zur Basis $b = 2$. Wir werden im Folgenden den natürlichen Logarithmus verwenden – andere Skalierung. Beachten Sie, dass die Entropie niemals negativ ist.

Warum ist dies eine geeignete Definition für Unsicherheit?

Stellen wir uns die Partition \mathcal{S} als ein Experiment mit m (sich gegenseitig ausschließenden) möglichen Ergebnissen vor. (Die Wahrscheinlichkeit für das i -te Ergebnis ist p_i .) Wir möchten, dass $H(\mathcal{S})$ die Menge an Informationen beschreibt, die man durch die Durchführung des Experiments zu gewinnen erwartet. Alternativ kann man $H(\mathcal{S})$ auch als die Menge an Unsicherheit bezüglich des Ergebnisses des Experiments betrachten. Zum Beispiel ist $H(\{\Omega\}) = -1 \cdot \log(1) = 0$, weil wir keine neuen Informationen gewinnen, wenn wir ein Experiment durchführen, dessen Ergebnis im Voraus bekannt ist. Wir listen zunächst einige Eigenschaften von H auf, welche man erwarten würde, wenn man H als Informationsgewinn aus einem Experiment interpretieren würde.

- Die Entropie hängt nicht von den einzelnen Teilmengen ab, die für die Partitionierung gewählt wurden, sondern nur von deren Wahrscheinlichkeiten. Eine Permutation der Wahrscheinlichkeiten p_1, \dots, p_m lässt den Wert $H(p_1, \dots, p_m)$ unverändert.

- Ein Experiment liefert dann und nur dann keine Informationen, wenn wir sein Ergebnis mit Sicherheit vorhersagen können. Es gilt $H(p_1, \dots, p_m) = 0$ genau dann wenn $p_i = 1$ für ein $i \in \{1, \dots, m\}$.
- Wenn ein bestimmtes Ergebnis eines Experiments unmöglich ist, dann kann es nicht schaden, es aus der Beschreibung des Experiments zu streichen. Es gilt $H(p_1, \dots, p_m, 0) = H(p_1, \dots, p_m)$.
- Das am wenigsten vorhersagbare Experiment ist eines, bei dem alle Ergebnisse gleich wahrscheinlich sind. H ist maximal genau dann wenn $p_i = 1/m$ für alle $i \in \{1, \dots, m\}$. Letzteres lässt sich leicht zeigen: Dazu betrachten wir die Funktion $f : [0, 1] \rightarrow [0, \infty)$ gegeben durch

$$f(t) := \begin{cases} 0 & t = 0 \\ -t \log(t) & t \in (0, 1] \end{cases}$$

Die Funktion f ist stetig und konkav und für p_1, \dots, p_m mit $p_i \geq 0$ und $\sum_{i=1}^m p_i = 1$ liefert die Ungleichung von Jensen

$$\frac{1}{m} H(p_1, \dots, p_m) = \frac{1}{m} \sum_{i=1}^m f(p_i) \leq f\left(\frac{1}{m} \sum_{i=1}^m p_i\right) = f\left(\frac{1}{m}\right) = -\frac{1}{m} \log(m)$$

woraus für jede Wahl von p_1, \dots, p_m mit $p_i \geq 0$ und $\sum_{i=1}^m p_i = 1$ die Ungleichung

$$H(p_1, \dots, p_m) \leq -\log(m)$$

folgt. Andererseits gilt bei Wahl von $p_i = 1/m$ für alle i

$$H(p_1, \dots, p_m) = -\log(m)$$

Damit erreicht $H(p_1, \dots, p_m)$ das Maximum für $p_i = 1/m$ für alle i .

- Die Abbildung

$$(p_1, \dots, p_m) \mapsto H(p_1, \dots, p_m)$$

ist stetig.

- Die letzte Eigenschaft ist etwas anspruchsvoller, aber immer noch intuitiv: Angenommen, wir haben zwei endliche Partitionen \mathcal{S}_1 und \mathcal{S}_2 und sei

$$\mathcal{S}_1 \vee \mathcal{S}_2 := \{A_1 \cap A_2 : A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2\}$$

deren Vereinigung. Die Vereinigung entspricht der Durchführung beider Experimente. Wir würden erwarten, dass wir die gleiche Menge an Informationen erhalten, wenn wir die beiden Experimente nacheinander durchführen, also

$$H(\mathcal{S}_1 \vee \mathcal{S}_2) = H(\mathcal{S}_1) + H(\mathcal{S}_2 | \mathcal{S}_1)$$

wobei $H(\mathcal{S}_2 | \mathcal{S}_1)$ die erwartete (oder bedingte) Entropie von \mathcal{S}_2 gegeben \mathcal{S}_1 beschreibt, d.h. die gewonnene Menge an Informationen, die von der Durchführung des Experiments \mathcal{S}_2 erwartet wird, vorausgesetzt, dass das Experiment \mathcal{S}_1 bereits durchgeführt wurde. Etwas genauer: Wenn das Experiment \mathcal{S}_1 durchgeführt wurde, dann ist ein Ereignis A_1 beobachtet worden. Die Menge an Information $H_{A_1}(\mathcal{S}_2)$, die nun vom Experiment $\mathcal{S}_2 = \{A_{2,1}, \dots, A_{2,m_2}\}$ erwartet wird, ist gegeben durch die Entropie der Partition

$\{A_{2,j} \cap A_1\}_{j \in \{1, \dots, m_2\}}$ von A_1 , also

$$H_{A_1}(\mathcal{S}_2) = H\left(\frac{\mu(A_{2,1} \cap A_1)}{\mu(A_1)}, \dots, \frac{\mu(A_{2,m_2} \cap A_1)}{\mu(A_1)}\right)$$

Dies wird noch gewichtet durch die Wahrscheinlichkeit $\mu(A_1)$ womit wir

$$H(\mathcal{S}_2 | \mathcal{S}_1) = \sum_{A_1 \in \mathcal{S}_1} H_{A_1}(\mathcal{S}_2) \mu(A_1)$$

und schließlich

$$H(\mathcal{S}_1 \vee \mathcal{S}_2) = H(\mathcal{S}_1) + H(\mathcal{S}_2 | \mathcal{S}_1) = H(\mathcal{S}_1) + \sum_{A_1 \in \mathcal{S}_1} H_{A_1}(\mathcal{S}_2) \mu(A_1)$$

erhalten.

Man kann nun zeigen, dass jede Funktion H , die alle diese Bedingungen erfüllt, so sein muss, wie in Gleichung (4.1) beschrieben (für eine beliebige Wahl der Basis des Logarithmus).

Eine Partition $\mathcal{S}_2 = \{A_{2,1}, \dots, A_{2,m_2}\}$ heißt *Verfeinerung* der Partition $\mathcal{S}_1 = \{A_{1,1}, \dots, A_{1,m_1}\}$, falls für jedes $j \in \{1, \dots, m_2\}$ ein $i \in \{1, \dots, m_1\}$ existiert mit $A_{2,j} \subseteq A_{1,i}$ (bis auf Nullmengen). In diesem Fall schreiben wir $\mathcal{S}_1 \leq \mathcal{S}_2$.

Übungsaufgabe 3. Seien \mathcal{S}_1 und \mathcal{S}_2 Partitionen von Ω .

- Es gilt $\mathcal{S}_1 \leq \mathcal{S}_1 \vee \mathcal{S}_2$ und $\mathcal{S}_2 \leq \mathcal{S}_1 \vee \mathcal{S}_2$.
- Aus $\mathcal{S}_1 \leq \mathcal{S}_2$ folgt $H(\mathcal{S}_1) \leq H(\mathcal{S}_2)$.
- Es gilt $H(\mathcal{S}_2 | \mathcal{S}_1) \leq H(\mathcal{S}_2)$.
- Es gilt $H(\mathcal{S}_1 \vee \mathcal{S}_2) \leq H(\mathcal{S}_1) + H(\mathcal{S}_2)$.

Beispiel 4.1. Sei $\Omega = \Sigma_2 = \{1, 2\}^{\mathbb{N}_0}$ und betrachte die Mengen

$$B := \{\mathbf{k} \in \Omega : k_0 = 1\} \quad \text{und} \quad B^* := \{\mathbf{k} \in \Omega : k_1 = 1\}$$

sowie die Partitionen $\mathcal{S} := \{B, \Omega \setminus B\}$ und $\mathcal{S}^* := \{B^*, \Omega \setminus B^*\}$. Dann gilt

$$\Omega \setminus B := \{\mathbf{k} \in \Omega : k_0 = 2\} \quad \text{und} \quad \Omega \setminus B^* := \{\mathbf{k} \in \Omega : k_1 = 2\}$$

Für $n \in \mathbb{N}$ definieren wir die Mengen (diese benötigen wir später noch)

$$A_{a_0, a_1, \dots, a_{n-1}} = \{\mathbf{k} \in \Omega : k_m = a_m, 0 \leq m \leq n-1\}$$

Dann gilt

$$\mathcal{S} \vee \mathcal{S}^* = \{A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}\}$$

Damit ist $\mathcal{S}_1 \vee \mathcal{S}_2$ eine Verfeinerung von \mathcal{S}_1 und eine Verfeinerung von \mathcal{S}_2 .

Wegen $\sigma^{-1}(\mathcal{S}) = \{\sigma^{-1}(B), \sigma^{-1}(\Omega \setminus B)\} = \{B^*, \Omega \setminus B^*\} = \mathcal{S}^*$ gilt demnach auch

$$\mathcal{S} \vee \sigma^{-1}(\mathcal{S}) = \{A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}\}$$

Induktiv lässt sich daher schließen, dass die Partition $\mathcal{S} \vee \sigma^{-1}(\mathcal{S}) \vee \dots \vee \sigma^{-(n-1)}(\mathcal{S})$ aus 2^n Mengen der Form $A_{a_0, a_1, \dots, a_{n-1}}$ mit $a_i \in \{1, 2\}$, $i \in \{0, 1, \dots, n-1\}$, besteht.

Die Entropie einer Partition führt direkt zur Definition der Entropie eines dynamischen Systems. Um dies zu motivieren, nehmen wir an, dass wir das Experiment zu verschiedenen Zeitpunkten wiederholen. Bei der ersten Durchführung erhalten wir vermutlich einige Informationen aus dem Ergebnis des Experiments. Der zweite Durchgang kann einige zusätzliche Informationen liefern (das Ergebnis eines Experiments kann sich von Durchführung zu Durchführung ändern). Und so weiter. Ist das dynamische System *vorhersehbar*, dann liefern spätere Experimente nicht viel neue Informationen. Andererseits lernen wir in einem *unvorhersehbaren* System, wie z.B. einem Münzwurf, bei jeder Durchführung etwas Neues - die erwartete Gesamtmenge an Informationen ist direkt proportional zur Anzahl der Wiederholungen des Experiments. Die Gesamtmenge an Informationen nach n Wiederholungen entspricht

$$E^n(T, \mathcal{S}) = H(\mathcal{S} \vee T^{-1}(\mathcal{S}) \vee T^{-2}(\mathcal{S}) \vee \dots \vee T^{-(n-1)}(\mathcal{S}))$$

wobei $T^{-m}(\mathcal{S}) = \{T^{-m}(A) : A \in \mathcal{S}\}$. Da der Grenzwert

$$\lim_{n \rightarrow \infty} \frac{1}{n} E^n(T, \mathcal{S})$$

existiert (siehe Lemma 4.2; für die Subadditivität siehe Übungsaufgabe 3), definieren wir

$$h(T, \mathcal{S}) := \lim_{n \rightarrow \infty} \frac{1}{n} E^n(T, \mathcal{S})$$

als die Entropie von T zur Partition \mathcal{S} . Als Entropie von T bezeichnen wir schließlich den Grenzwert für das effektivste Experiment:

$$h(T) := \sup_{\mathcal{S} \text{ endliche Partition von } \Omega} h(T, \mathcal{S})$$

Übungsaufgabe 4. Es gilt $H(T^{-1}(\mathcal{S})) = H(\mathcal{S})$.

Lemma 4.2. Für jede subadditive Folge nicht-negativer reeller Zahlen $\{a_n\}_{n \in \mathbb{N}}$ existiert der Grenzwert $\lim_{n \rightarrow \infty} a_n/n$.

Beweis: Subadditivität bedeutet hier, dass $a_{p+q} \leq a_p + a_q$ für alle $p, q \in \mathbb{N}$ gilt.

Sei nun $m \in \mathbb{N}$. Für $n \in \mathbb{N}$ mit $n > m$ existieren dann $k, i \in \mathbb{N}$ mit $1 \leq i \leq m$ und $n = km + i$. Die Subadditivität liefert dann

$$\frac{a_n}{n} = \frac{a_{km+i}}{km+i} \leq \frac{a_{km+i}}{km} \leq \frac{a_{km}}{km} + \frac{a_i}{km} \leq k \frac{a_m}{km} + \frac{a_i}{km} = \frac{a_m}{m} + \frac{a_i}{km}$$

Aus $n \rightarrow \infty$ folgt $k \rightarrow \infty$ woraus sich

$$\limsup_{n \rightarrow \infty} \frac{a_n}{n} \leq \frac{a_m}{m} + \limsup_{k \rightarrow \infty} \frac{a_i}{km} = \frac{a_m}{m}$$

ergibt. Da m beliebig war, folgt damit

$$\limsup_{n \rightarrow \infty} \frac{a_n}{n} \leq \inf_{n \in \mathbb{N}} \frac{a_n}{n} \leq \liminf_{n \rightarrow \infty} \frac{a_n}{n}$$

Also existiert der Grenzwert $\lim_{n \rightarrow \infty} \frac{a_n}{n}$ und entspricht $\inf_{n \in \mathbb{N}} \frac{a_n}{n}$.

4.3 Berechnung der Entropie

Die hergeleitete Definition der Entropie ist, aufgrund des Supremums über alle möglichen endlichen Partitionen, praktisch nur bedingt handhabbar. Der folgende Satz beseitigt diese Schwierigkeit, indem er uns erlaubt, nur eine einzige (*erzeugende*) Partition zu betrachten.

Satz 4.3. (*ohne Beweis*)

1. (*Kolmogorov & Sinai 1958*)
Ist \mathcal{S} eine endliche, erzeugende Partition für T , d.h.

$$\bigcup_{n \in \mathbb{N}_0} T^{-n}(\mathcal{S})$$

ist ein Erzeuger der σ -Algebra \mathcal{A} (bis auf Nullmengen), dann gilt $h(T) = h(T, \mathcal{S})$.

2. (*Krieger 1970*)
Ist T eine maßerhaltende und ergodische Abbildung, dann besitzt T eine endliche, erzeugende Partition.

Wir bestimmen im Folgenden die Entropy unseres Münzwurf Beispiels:

Korollar 4.4. (*Münzwurf III*)

Wir betrachten das dynamische System $(\Sigma_2, \mathcal{B}(\Sigma_2), \sigma, \mu_{\vartheta_2})$ zum Linksshift σ wie zuvor. Dann gilt $h(\sigma) = \log(2)$.

Beweis: Wir definieren die Menge (vergleich Beispiel 4.1)

$$B := \{\mathbf{k} \in \Omega : k_0 = 1\}$$

und die Partition $\mathcal{S} := \{B, \Omega \setminus B\}$; es gilt $\Omega \setminus B = \{\mathbf{k} \in \Omega : k_0 = 2\}$. Dann besteht

$$\mathcal{S} \vee \sigma^{-1}(\mathcal{S}) \vee \sigma^{-2}(\mathcal{S}) \vee \dots \vee \sigma^{-(n-1)}(\mathcal{S})$$

aus 2^n Mengen der Form

$$A_{a_0, a_1, \dots, a_{n-1}} = \{\mathbf{k} \in \Omega : k_m = a_m, 0 \leq m \leq n-1\}$$

mit $\mu(A_{a_0, a_1, \dots, a_{n-1}}) = 1/2^n$. Damit erhalten wir

$$\begin{aligned} h(\sigma, \mathcal{S}) &= \lim_{n \rightarrow \infty} \frac{1}{n} E^n(\sigma, \mathcal{S}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{a_0, a_1, \dots, a_{n-1} \in \{1, 2\}} -\mu(A_{a_0, a_1, \dots, a_{n-1}}) \log(\mu(A_{a_0, a_1, \dots, a_{n-1}})) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} 2^n \frac{1}{2^n} \log(2^n) \\ &= \log(2) \end{aligned}$$

Wegen

$$\bigcup_{n \in \mathbb{N}_0} T^{-n}(\mathcal{S}) = \mathfrak{h}_2$$

und Lemma 2.10 gilt schließlich $h(\sigma) = \log(2)$.

Analog zum Beweis von Korollar 4.4 lässt sich zeigen, dass für das dynamische System $(\Sigma_N, \mathcal{B}(\Sigma_N), \sigma, \mu_{\vartheta_N})$ mit $N \in \mathbb{N}$ gilt $h(\sigma) = \log(N)$.

Satz 4.5. *Die Entropie ist isomorphieinvariant.*

Beweis: Seien $(\Omega_1, \mathcal{A}_1, \mu_1, T_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2, T_2)$ zwei isomorphe dynamische Systeme. Dann existieren Mengen $\Lambda_1 \in \mathcal{A}_1, \Lambda_2 \in \mathcal{A}_2$ mit $\mu_1(\Lambda_1) = 1 = \mu_2(\Lambda_2)$ und $T_1(\Lambda_1) \subseteq \Lambda_1$ und $T_2(\Lambda_2) \subseteq \Lambda_2$, und eine invertierbare maßtreue Transformation $\phi : \Lambda_1 \rightarrow \Lambda_2$, sodass

$$T_2 \circ \phi = \phi \circ T_1$$

Sei $\mathcal{S}_2 = \{B_1, \dots, B_m\}$ eine Partition von Λ_2 . Dann ist $\phi^{-1}(\mathcal{S}_2) = \{\phi^{-1}(B_1), \dots, \phi^{-1}(B_m)\}$ eine Partition von Λ_1 . Wir setzen $A_i := \phi^{-1}(B_i)$ für alle $i \in \{1, \dots, m\}$. Wir erhalten zunächst für beliebige Elemente $B_{i_0}, \dots, B_{i_{n-1}} \in \mathcal{S}_2$

$$\begin{aligned} & \mu_2(B_{i_0} \cap T_2^{-1}(B_{i_1}) \cap \dots \cap T_2^{-(n-1)}(B_{i_{n-1}})) \\ &= \mu_1(\phi^{-1}(B_{i_0}) \cap \phi^{-1}(T_2^{-1}(B_{i_1})) \cap \dots \cap \phi^{-1}(T_2^{-(n-1)}(B_{i_{n-1}}))) \\ &= \mu_1(\phi^{-1}(B_{i_0}) \cap T_1^{-1}(\phi^{-1}(B_{i_1})) \cap \dots \cap T_1^{-(n-1)}(\phi^{-1}(B_{i_{n-1}}))) \\ &= \mu_1(A_{i_0} \cap T_1^{-1}(A_{i_1}) \cap \dots \cap T_1^{-(n-1)}(A_{i_{n-1}})) \end{aligned}$$

und mit $A(n) := A_{i_0} \cap T_1^{-1}(A_{i_1}) \cap \dots \cap T_1^{-(n-1)}(A_{i_{n-1}})$ und $B(n) := B_{i_0} \cap T_2^{-1}(B_{i_1}) \cap \dots \cap T_2^{-(n-1)}(B_{i_{n-1}})$

$$\begin{aligned} h(T_2) &= \sup_{\mathcal{S}_2} h(T_2, \mathcal{S}_2) \\ &= \sup_{\mathcal{S}_2} \lim_{n \rightarrow \infty} \frac{1}{n} E^n(T_2, \mathcal{S}_2) \\ &= \sup_{\mathcal{S}_2} \lim_{n \rightarrow \infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T_2^{-i}(\mathcal{S}_2)) \\ &= \sup_{\mathcal{S}_2} \lim_{n \rightarrow \infty} \frac{-1}{n} \sum_{B(n) \in \bigvee_{i=0}^{n-1} T_2^{-i}(\mathcal{S}_2)} \mu_2(B(n)) \log(\mu_2(B(n))) \\ &= \sup_{\phi^{-1}(\mathcal{S}_2)} \lim_{n \rightarrow \infty} \frac{-1}{n} \sum_{A(n) \in \bigvee_{i=0}^{n-1} T_1^{-i}(\phi^{-1}(\mathcal{S}_2))} \mu_1(A(n)) \log(\mu_1(A(n))) \\ &= \sup_{\phi^{-1}(\mathcal{S}_2)} h(T_1, \phi^{-1}(\mathcal{S}_2)) \\ &\leq \sup_{\mathcal{S}_1} h(T_1, \mathcal{S}_1) \\ &= h(T_1) \end{aligned}$$

Damit ist die Behauptung gezeigt.

Aus Korollar 4.4, Satz 4.5 und den Resultaten in Abschnitt 3.4 erhalten wir abschließend folgendes Resultat:

Korollar 4.6. *Die Entropie der dynamische Systeme*

- $([0, 1), \mathcal{B}([0, 1)), \lambda, t)$
- $([0, 1), \mathcal{B}([0, 1)), f_* d\lambda, L_4)$

beträgt $\log(2)$.

Literatur

- [1] C.J. Efthimiou: A class of continued radicals. *The American Mathematical Monthly* 120(5), 459–461 (2013)
- [2] J. Fernández Sánchez, W. Trutschnig: Nested Square Roots of 2 Revisited, *The American Mathematical Monthly* 127(4), 344-351 (2020)
- [3] D.W. Morris: Ratner’s Theorem on Unipotent Flows. Available at <https://arxiv.org/abs/math/0310402v5>. (2005)
- [4] C.E. Silva: *Invitation to Ergodic Theory*. American Mathematical Society, Providence RI. (2008)
- [5] S.M. Ulam, J. von Neumann: On combination of stochastic and deterministic processes, *Bulletin of the American Mathematical Society* 53 (11), 1120- (1947)