



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Die Vielfalt der Entropie –
Eine mathematische Betrachtung ausgewählter
Entropiebegriffe“

verfasst von / submitted by

Michael Putzenlechner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 190 423 406

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Lehramtsstudium UniStG
UF Chemie UniStG
UF Mathematik UniStG

Betreut von / Supervisor:

ao. Univ.-Prof. Mag. Dr. Peter Raith

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Wien, Januar 2020

Michael Putzenlechner

Danksagung

Mit der Abgabe meiner Diplomarbeit und der Absolvierung der Diplomprüfung endet meine Studienzeit. Deshalb gilt mein besonderer Dank meinen Eltern für die mentale und finanzielle Unterstützung. Dadurch war mir ein sorgenfreies Studium möglich. Danken möchte ich auch recht herzlich meiner Freundin, die mir viel Geduld und Zeit entgegenbrachte.

Ebenfalls ganz herzlich bedanke ich mich bei meinem Betreuer Herrn Univ. Prof. Dr. Peter Raith. Durch seine hervorragend strukturierten Vorlesungen wurde meine Leidenschaft für Mathematik entfacht. Seine kritischen und vor allem schnellen Antworten gewährleisteten ein optimales Vorankommen dieser Arbeit.

Kurzfassung

In Teilgebieten der Mathematik und Physik werden unterschiedliche Ausdrücke als Entropie bezeichnet. Die Entropiebegriffe, die in dieser Arbeit betrachtet werden, sind die Shannon-Entropie der Wahrscheinlichkeits- bzw. Informationstheorie, die Boltzmann-Entropie der statistischen Thermodynamik sowie die topologische Entropie und die metrische Entropie der Theorie der dynamischen Systeme. Diese Ausdrücke unterscheiden sich sowohl in ihrer mathematischen Definition als auch in ihrer kontextbezogenen Interpretation.

Ziel dieser Arbeit ist, trotz der mathematischen und kontextbezogenen Unterschiede diese Entropieausdrücke mit Hilfe der Betrachtung eines Teilchensystems aus unabhängigen und nicht unterscheidbaren Teilchen auf Gemeinsamkeiten und formale Zusammenhänge hin zu untersuchen.

Die Ergebnisse zeigen, dass Gemeinsamkeiten hinsichtlich der Logarithmusfunktion, die in den Definitionen verwendet wird und der Maximumeigenschaften der Ausdrücke bestehen. In Korollar 4.4.3.1., Satz 5.3.5., Satz 5.4.2. und Satz 5.4.3. werden Zusammenhänge zwischen den Entropieausdrücken festgestellt.

Abstract

In subareas of both mathematics and physics there are different expressions which are termed as entropy. The entropy expressions, which will be considered in this thesis, are Shannon's entropy of probability- resp. information theory, Boltzmann's entropy of statistical thermodynamics as well as the metric entropy and the topological entropy of the theory of dynamical systems. These expressions differ from each other in both their mathematical definition and contextual interpretation.

The aim of this thesis is to analyse these entropy expressions for similarities and coherences despite the mathematical and contextual differences with the aid of a particle system of independent and indistinguishable particles.

The results show that there are similarities regarding the logarithmic function used in the definitions and the maximum properties of the expressions. In corollary 4.4.3.1., theorem 5.3.5., theorem 5.4.2. and theorem 5.4.3. there are coherences between the entropy expressions ascertained.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Zielsetzung und Fragestellung	1
1.2. Methodik	1
1.3. Aufbau der Arbeit	2
2. Grundlagen	3
2.1. Wahrscheinlichkeitstheorie	3
2.2. Dynamische Systeme	13
3. Entropiebegriff der Informationstheorie	17
3.1. Informationsgehalt	17
3.2. Shannon-Entropie	21
3.3. Shannon-Entropie eines Produktraums	24
3.4. Shannon-Entropie für kontinuierliche Ergebnismengen	27
3.5. Maximale Shannon-Entropie	31
4. Entropiebegriff der statistischen Thermodynamik	36
4.1. Boltzmann-Entropie	36
4.2. Herleitung von Boltzmanns Entropieformel	38
4.3. Maximale Boltzmann-Entropie	39
4.4. Zusammenhang mit der Shannon-Entropie	41
5. Entropiebegriffe der dynamischen Systeme	47
5.1. Topologische Entropie	47
5.2. Metrische Entropie	53
5.3. Verbindung mit der Shannon-Entropie	56
5.4. Maximale metrische Entropie der Shift-Abbildung	65

6. Conclusio	68
6.1. Gemeinsamkeiten	68
6.2. Zusammenhänge	69
6.3. Ausblick	71
Literaturverzeichnis	72
Abbildungsverzeichnis	75

1. Einleitung

Der Begriff *Entropie* wurde 1865 von Rudolf Clausius eingeführt, um das Phänomen der Irreversibilität von physikalischen Systemen zu erklären [ES14, S. 89]. In den folgenden Jahrzehnten wurden weitere Entropieausdrücke definiert, die sich aus anderen physikalischen bzw. mathematischen Fragestellungen heraus entwickelten. Dazu zählen unter anderem die Shannon-Entropie, die Boltzmann-Entropie, die metrische Entropie und die topologische Entropie. Im Zuge meines Lehramtstudiums für Mathematik und Chemie wurden diese Entropieausdrücke zwar behandelt, jedoch blieb eine gemeinsame Betrachtung dieser Begriffe aus. Die Frage nach den Gemeinsamkeiten und formalen Zusammenhängen zwischen diesen Entropieausdrücken blieb unbeantwortet.

1.1. Zielsetzung und Fragestellung

Ziel dieser Arbeit ist die Entropiebegriffe von Boltzmann, Shannon und der Theorie dynamischer Systeme zunächst einzeln zu beschreiben, um anschließend mit Hilfe eines Teilchensystems aus unabhängigen und nicht unterscheidbaren Teilchen Gemeinsamkeiten zu erkennen sowie formale Zusammenhänge zwischen diesen Begriffen festzustellen. Die Forschungsfrage lautet:

Welche Gemeinsamkeiten und formale Zusammenhänge lassen sich zwischen den Entropiebegriffen nach Shannon, Boltzmann und der Theorie dynamischer Systeme feststellen?

1.2. Methodik

Die Gemeinsamkeiten und formalen Zusammenhänge werden durch Betrachtung eines Teilchensystems aus unabhängigen und nicht unterscheidbaren Teilchen ermittelt.

Dieses Teilchensystem wurde gewählt, da einerseits für Boltzmanns Entropieausdruck nicht unterscheidbare Teilchen betrachtet werden und andererseits durch die Annahme der Unabhängigkeit endliche und unendliche Produkträume herangezogen werden können.

1.3. Aufbau der Arbeit

Die vorliegende Arbeit ist in sechs Hauptkapitel unterteilt. Die Einleitung beschreibt die Motivation des Autors und die Hintergrundsituation des Problems. Mit Hilfe der Beschreibung der Situation wird in Kapitel 1 die wissenschaftliche Fragestellung der Arbeit formuliert.

Im folgenden Kapitel 2. werden die Grundlagen der Wahrscheinlichkeitstheorie und der Theorie der dynamischen Systeme, die zur Behandlung der Forschungsfrage verwendet werden, wiederholt.

Kapitel 3 enthält die Analyse der Shannon-Entropie, die als Bindeglied für die in den folgenden Kapiteln analysierten Entropieausdrücken dient.

In Kapitel 4 wird zunächst die Boltzmann-Entropie einzeln betrachtet und anschließend wird ein formaler Zusammenhang mit der Shannon-Entropie ermittelt.

Kapitel 5 ist inhaltlich analog wie Kapitel 4 aufgebaut: Zu Beginn werden die topologische- und die metrische Entropie einzeln analysiert und anschließend werden diese Entropien mit der Shannon-Entropie in Verbindung gebracht.

Das abschließende Kapitel 6 gibt eine Übersicht der gefunden Gemeinsamkeiten und formalen Zusammenhänge zwischen den Entropiebegriffen.

2. Grundlagen

Für die Definitionen der Shannon-, Boltzmann-, metrischen- und der topologischen Entropie werden Begriffe aus der Wahrscheinlichkeitstheorie und der Theorie dynamischer Systeme, wobei diese in diesem Kapitel näher erläutert werden, verwendet.

2.1. Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeitstheorie behandelt die mathematische Beschreibung von Zufallsgeschehnissen [Geo09, S. 7].

Dabei werden die *Ergebnisse* ω eines Zufallsgeschehens zu einer Menge, der *Ergebnismenge* Ω , zusammengefasst. Eine Teilmenge $A \subseteq \Omega$ wird *Ereignis* genannt und beschreibt das Ereignis, dass das Ergebnis einer Durchführung des Zufallsgeschehens in dieser Teilmenge liegt [Tap13, S. 1 und 3].

Zum besseren Verständnis ein aus dem Alltag bekanntes Beispiel:

Beispiel. Beim Würfeln eines sechsseitigen Würfels sind die Ergebnisse die Zahlen von 1 bis 6. Die daraus resultierende Ergebnismenge lautet $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis „Eine gerade Zahl wird gewürfelt.“ wird durch die Menge $A = \{2, 4, 6\} \subseteq \Omega$ beschrieben.

Im obigen Beispiel ist die Ergebnismenge endlich. Die durch Experimente bestätigte Erfahrungstatsache, dass sich die relative Häufigkeit des Eintretens z.B. des Ereignisses $\{2, 4, 6\}$ beim fortlaufenden Würfeln dem Wert $\frac{1}{2}$ annähert, führt zum Begriff der Wahrscheinlichkeit: Der beobachtete „Grenzwert“, gegen den die relative Häufigkeit des Eintretens des Ereignisses $\{2, 4, 6\}$ „strebt“, wird als Wahrscheinlichkeit dieses Ereignisses bezeichnet. Nach dieser intuitiven Festlegung ist die Wahrscheinlichkeit von

$\{2, 4, 6\}$ gleich $\frac{1}{2}$. Allgemein würde in diesem Beispiel ein Ereignis $A \subseteq \Omega$ der Wahrscheinlichkeit $P(A) = \frac{|A|}{|\Omega|}$ zugeordnet werden. Dabei steht $|\cdot|$ für die Mächtigkeit einer Menge [Bau68, S. 110].

Dieses Vorgehen soll nun formalisiert werden: Sei n die Anzahl an Versuchsdurchführungen und $N_n(A)$ die Anzahl der Versuche, in denen A eingetreten ist, dann wird die Wahrscheinlichkeit $P(A)$ definiert als „ $P(A) := \lim_{n \rightarrow \infty} \frac{N_n(A)}{n}$ “ [Hen08, S. 20]. Folgendes Problem tritt auf: Es wird hier eine Folge betrachtet, die erhalten wird, in dem in der Praxis ein Versuch immer wieder ausgeführt wird. Dies geht jedoch nur endlich oft, man bräuchte jedoch unendlich viele Folgenglieder für eine Folge. Selbst wenn man eine Folge erhalten würde, müsste diese nicht konvergieren [Hen08, S. 20]. Diese Definition ist also mathematisch nicht exakt.

Der nächste Versuch, den Wahrscheinlichkeitsbegriff exakt zu definieren ist, diesen zwar nicht durch „ $P(A) := \lim_{n \rightarrow \infty} \frac{N_n(A)}{n}$ “ zu definieren, aber wie beim Würfelbeispiel durch diese Idee geleitet, Ausdrücke wie $P(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathfrak{P}(\Omega)$ zu definieren. Dieser Ansatz kommt der heute verwendeten Definition der Wahrscheinlichkeit schon näher. Jedoch muss folgendes bedacht werden: Bei überabzählbar unendlichen Mengen wie \mathbb{R} oder $[0,1)$ können Teilmengen konstruiert werden, die keiner intuitiv „sinnvollen“ Wahrscheinlichkeit zugeordnet werden können [Els18, S. 5], [Els18, S. 109-110].

In solchen Fällen betrachtet man nicht die ganze *Potenzmenge* $\mathfrak{P}(\Omega)$, sondern beschränkt sich auf die Ereignisse einer speziellen Familie an Teilmengen der Menge Ω , einer σ -Algebra [Tap13, S. 9]:

Definition 2.1.1. Sei $\Omega \neq \emptyset$. Man nennt eine Familie $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ von Teilmengen von Ω eine σ -Algebra, falls:

- (i) $\emptyset \in \mathcal{A}$
- (ii) $A \in \mathcal{A} \Rightarrow A^c := \Omega \setminus A \in \mathcal{A}$
- (iii) $(A_n)_{n \in \mathbb{N}}$ eine Folge in $\mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

Für folgende Proposition vergleiche [Tap13, S. 9, Lemma 2.6]:

Proposition 2.1.2. *Es sei \mathcal{A} eine σ -Algebra über Ω . Für jede Folge $(A_n)_{n \in \mathbb{N}}$ in \mathcal{A} gilt $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$.*

Das Problem einer allgemeinen, exakten mathematischen Definition des Wahrscheinlichkeitsbegriffes löste Kolmogoroff 1933 [Bau68, S. 11]. Er führte den Begriff über formale Eigenschaften ein, die eine Wahrscheinlichkeit intuitiv erfüllen soll und berücksichtigte, dass es, wie oben genannt, Beispiele gibt, in denen es nicht möglich ist, die Wahrscheinlichkeit für alle Mengen aus $\mathfrak{P}(\Omega)$ festzulegen [Bau68, S. 111].

Dies führte zu folgender Definition [Tap13, S. 4 und 15]:

Definition 2.1.3 (Axiome von Kolmogoroff). Sei $\Omega \neq \emptyset$, $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ eine σ -Algebra auf Ω . Dann nennt man $P : \mathcal{A} \rightarrow \mathbb{R}$ ein *Wahrscheinlichkeitsmaß*, falls:

- (i) $P(A) \geq 0 \forall A \in \mathcal{A}$
- (ii) $P(\Omega) = 1$
- (iii) Falls $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}$ und paarweise disjunkt ist, dann gilt: $P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

Das Tripel (Ω, \mathcal{A}, P) wird dann *Wahrscheinlichkeitsraum* genannt.

Beispiel. Für $\Omega := [0,1)$ mit der *Borelschen σ -Algebra* $\mathcal{B}([0,1))$ bezüglich dieser Menge und mit $P := \lambda$, wobei λ hier das *Lebesguemaß* auf $\mathcal{B}([0,1))$ bezeichnet, ist P ein Wahrscheinlichkeitsmaß (Die Punkte (i) und (iii) gelten, da λ ein Maß ist, der Punkt (ii) ist erfüllt, da $P([0,1)) = \lambda([0,1)) = 1 - 0 = 1$).

Einige elementare Rechenregeln werden in folgender Proposition angeführt [Tap13, S. 13, Satz 2.18]:

Proposition 2.1.4. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gelten folgende Aussagen:

- (i) Für zwei beliebige Ereignisse $A, B \in \mathcal{A}$ gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

- (ii) Für zwei beliebige Ereignisse $A, B \in \mathcal{A}$ mit $A \subseteq B$ gilt

$$P(A) = P(B) - P(B \setminus A).$$

Insbesondere gilt $P(A) \leq P(B)$.

(iii) Für zwei beliebige Ereignisse $A, B \in \mathcal{A}$ gilt

$$P(A \cap (\Omega \setminus B)) = P(A) - P(A \cap B).$$

(iv) Für jedes Ereignis $A \in \mathcal{A}$ gilt $P(A) = 1 - P(A^c)$.

(v) Es gilt $P(\emptyset) = 0$.

Definition 2.1.5. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge und $(A_i)_{i \in I}$ eine beliebige Familie von Ereignissen aus \mathcal{A} . Die Familie $(A_i)_{i \in I}$ heißt *unabhängig*, falls für jede endliche Teilmenge $J \subseteq I$ gilt, dass

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j),$$

vergleiche [Kle13, S.51].

Liegt das Interesse beim Würfelbeispiel daran, ob ein Sechser gewürfelt wird oder nicht, so kann dies mit folgender Abbildung beschrieben werden:

$$X : \{1, 2, \dots, 6\} \rightarrow \{0, 1\}, \quad X(\omega) := \begin{cases} 1 & \text{für } \omega = 6, \\ 0 & \text{sonst.} \end{cases}$$

Solche Abbildungen können allgemein definiert werden und werden *Zufallsvariablen* genannt [Kle13, S. 43]:

Definition 2.1.6. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann wird eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ eine (reelle) Zufallsvariable genannt, falls $\forall \alpha \in \mathbb{R} : \{\omega \in \Omega : X(\omega) \leq \alpha\} \in \mathcal{A}$, diese also $\mathcal{A} - \mathcal{B}(\mathbb{R})$ messbar ist.

Ist $M \in \mathcal{B}(\mathbb{R})$, so wird $\{X \in M\} := X^{-1}(M)$ und $P(X \in M) := P(X^{-1}(M))$ gesetzt. Speziell wird für ein beliebiges $b \in \mathbb{R}$ der Ausdruck $\{X \leq b\}$ definiert durch $\{X \leq b\} := X^{-1}((-\infty, b])$.

Definition 2.1.7. Sei X eine Zufallsvariable.

- (i) Das Wahrscheinlichkeitsmaß $P_X := P \circ X^{-1}$ heißt *Verteilung* von X , wobei X^{-1} die Urbildfunktion von X bezeichnet.
- (ii) Ist X eine reelle Zufallsvariable, so heißt die Abbildung $F_X : \mathbb{R} \rightarrow [0, 1]$, $x \mapsto P(X \leq x)$ die *Verteilungsfunktion* von X .

(iii) Eine Familie $(X_i)_{i \in I}$ heißt *gleich verteilt*, falls $P_{X_i} = P_{X_j}$ für alle $i, j \in I$. vergleiche [Kle13, S. 43].

Definition 2.1.8. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge und $(X_i)_{i \in I}$ eine beliebige Familie von Zufallsvariablen. Die Familie $(X_i)_{i \in I}$ heißt *unabhängig*, falls für jede endliche Teilmenge $J \subseteq I$ und für jede Wahl $M_j \in \mathcal{B}(\mathbb{R})$, $j \in J$ gilt, dass

$$P\left(\bigcap_{j \in J} \{X_j \in M_j\}\right) = \prod_{j \in J} P(X_j \in M_j),$$

vergleiche [Kle13, S. 57].

Definition 2.1.9. Eine reelle Zufallsvariable X heißt *integrierbar*, falls $X \in \mathcal{L}^1(P) := \mathcal{L}^1(\Omega, \mathcal{A}, P) := \{X : \Omega \rightarrow \overline{\mathbb{R}} : X \text{ ist messbar und } \int |X| dP < \infty\}$. Der Ausdruck

$$E(X) := \int X dP$$

wird der *Erwartungswert* von X genannt [Kle13, S.90, 103]. Für die Struktur auf $\overline{\mathbb{R}}$ und die Bedeutung eines Integrals bezüglich eines Maßes siehe [Kle13, Kapitel 4].

Die intuitive Idee, dass die Wahrscheinlichkeit dem „Grenzwert der relativen Häufigkeit bei unendlich vielen Versuchsdurchführungen“ entspricht, lässt sich formal in der auf dem axiomatisch definierten Wahrscheinlichkeitsbegriff von Kolmogoroff und damit der Maßtheorie aufbauenden Wahrscheinlichkeitstheorie wiederfinden. Dafür wird ein Wahrscheinlichkeitsraum benötigt, mit dem das fortlaufende Ausführen eines Zufallsgeschehens modelliert wird. Das heißt, dass ein Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ zu einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) gesucht wird, der beschreibt, dass ein durch (Ω, \mathcal{A}, P) beschriebenes Zufallsgeschehen unabhängig voneinander immer wieder ausgeführt wird [Bau68, S. 135]. Durch $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ soll z.B. das Immer-wieder-Würfeln eines Würfels [Bau68, S. 135] beschrieben werden. Dieser Wahrscheinlichkeitsraum wird später verwendet, um Zusammenhänge zwischen den in dieser Arbeit untersuchten Entropieausdrücken zu zeigen. Folgende Definitionen werden verwendet [Kle13, S. 280]:

Definition 2.1.10. Sei Ω eine nichtleere Menge. Für $\Omega^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_n \in \Omega \forall n \in \mathbb{N}\}$ wird die Abbildung $\pi_i : \Omega^{\mathbb{N}} \rightarrow \Omega$, $\omega \mapsto \omega_i$ die *i-te Koordinatenabbildung* genannt.

Definition 2.1.11. Sei $\Omega \neq \emptyset$ und \mathcal{A} eine σ -Algebra auf Ω . Die *Produkt- σ -Algebra* $\mathcal{A}^{\otimes \mathbb{N}}$ ist die kleinste σ -Algebra auf $\Omega^{\mathbb{N}}$, sodass für jedes $i \in \mathbb{N}$ die Abbildung π_i messbar bezüglich $\mathcal{A}^{\otimes \mathbb{N}} - \mathcal{A}$ ist, also

$$\mathcal{A}^{\otimes \mathbb{N}} := \sigma(\{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\}).$$

Sei nun (Ω, \mathcal{A}, P) ein beliebiger Wahrscheinlichkeitsraum. $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ kann wie folgt konstruiert werden [Bau68, S. 135]:

Wird ein Zufallsgeschehen immer wieder ausgeführt, so erhält man eine Folge an Ergebnissen dieser Versuchsdurchführungen. Daher soll $\tilde{\Omega}$ die Menge aller Folgen $\omega : \mathbb{N} \rightarrow \Omega$ sein, weshalb $\tilde{\Omega} := \Omega^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_n \in \Omega \forall n \in \mathbb{N}\}$ gesetzt wird.

Für $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathcal{A}$ setze

$$[A_1, \dots, A_n] := \{(\omega_1, \omega_2, \dots) : \omega_1 \in A_1, \omega_2 \in A_2, \dots, \omega_n \in A_n\}.$$

Solche Mengen werden Zylindermengen genannt, vergleiche hierfür die analog definierten Mengen in [Kle13, S. 18]. $[A_1, \dots, A_n]$ wird als das Ereignis interpretiert, dass beim fortlaufenden Ausführen des Versuchs bei der j -ten Durchführung, $j \in \{1, 2, \dots, n\}$, das Ereignis A_j eintritt. Da die Versuche unabhängig voneinander durchgeführt werden, wird die Wahrscheinlichkeit für eine Zylindermenge $[A_1, \dots, A_n]$ definiert durch

$$\tilde{P}([A_1, \dots, A_n]) := \prod_{j=1}^n P(A_j).$$

Von $\tilde{\mathcal{A}}$ wird verlangt, dass alle Zylindermengen enthalten sind und \tilde{P} soll ein Wahrscheinlichkeitsmaß auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ sein, welche obige Festlegung für Zylindermengen erfüllt. Daher sei $\tilde{\mathcal{A}}$ die kleinste σ -Algebra, die alle Zylindermengen enthält:

$$\tilde{\mathcal{A}} := \sigma(\{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\}).$$

Im Folgenden wird gezeigt, dass die bisher nur auf der Menge aller Zylindermengen $\{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\}$ definierte Abbildung \tilde{P} eindeutig zu einem Maß auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ fortgesetzt werden kann.

Satz 2.1.12. *Es gilt $\tilde{\mathcal{A}} = \mathcal{A}^{\otimes \mathbb{N}}$.*

Beweis. (\subseteq) Sei $A \in \mathcal{A}$ beliebig und $i \in \mathbb{N}$. Es gilt

$$\begin{aligned} \pi_i^{-1}(A) &= \{(\omega_1, \omega_2, \dots) : \omega_i \in A\} \\ &= [\Omega, \Omega, \dots, A_i] \subseteq \{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\} \\ &\Rightarrow \{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\} \subseteq \{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\}. \end{aligned}$$

Daraus folgt

$$\begin{aligned} \mathcal{A}^{\otimes \mathbb{N}} &= \sigma(\{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\}) \\ &\subseteq \sigma(\{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\}) = \tilde{\mathcal{A}}. \end{aligned}$$

(\supseteq) Es gilt

$$\begin{aligned} [A_1, \dots, A_n] &= \{(\omega_1, \omega_2, \dots) : \omega_1 \in A_1, \omega_2 \in A_2, \dots, \omega_n \in A_n\} \\ &= \bigcap_{k=1}^n \underbrace{\{(\omega_1, \omega_2, \dots) : \omega_k \in A_k\}}_{=\pi_k^{-1}(A_k)} = \bigcap_{k=1}^n \pi_k^{-1}(A_k). \end{aligned}$$

Ist somit \mathcal{F} eine σ -Algebra mit $\{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\} \subseteq \mathcal{F}$, so ist $[A_1, \dots, A_n] = \bigcap_{k=1}^n \pi_k^{-1}(A_k) \in \mathcal{F}$. Demzufolge ist $\{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\} \subseteq \mathcal{F}$. Daher gilt

$$\begin{aligned} \{\mathcal{F} \text{ ist } \sigma\text{-Algebra} : \{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\} \subseteq \mathcal{F}\} \\ \subseteq \{\mathcal{F} \text{ ist } \sigma\text{-Algebra} : \{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\} \subseteq \mathcal{F}\} \end{aligned}$$

Damit erhalt man

$$\tilde{\mathcal{A}} = \bigcap_{\substack{\{[A_1, \dots, A_n] : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A}\} \subseteq \mathcal{F} \\ \mathcal{F} \text{ ist } \sigma\text{-Algebra}}} \mathcal{F} \subseteq \bigcap_{\substack{\{\pi_i^{-1}(A) : i \in \mathbb{N}, A \in \mathcal{A}\} \subseteq \mathcal{F} \\ \mathcal{F} \text{ ist } \sigma\text{-Algebra}}} \mathcal{F} = \mathcal{A}^{\otimes \mathbb{N}},$$

da im linken Durchschnitt mindestens so viele Mengen geschnitten werden wie im rechten Durchschnitt. \square

Folgendes Korollar des *Satzes von Ionescu-Tulcea* [Kle13, S. 293, Satz 14.32] besagt, dass \tilde{P} eindeutig zu einem Ma auf $\tilde{\mathcal{A}}$ fortgesetzt werden kann [Kle13, Kor. 14.33]:

Proposition 2.1.13 (Kor. d. Satzes von Ionescu-Tulcea). *Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß \tilde{P} auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ mit*

$$\tilde{P}([A_1, \dots, A_n]) = \prod_{j=1}^n P(A_j)$$

für $A_j \in \mathcal{A}$, $j \in \{1, 2, \dots, n\}$ und $n \in \mathbb{N}$.

Sei X eine Zufallsvariable auf (Ω, \mathcal{A}, P) . Betrachte die Zufallsvariable $\tilde{X}_n : \tilde{\Omega} \rightarrow \mathbb{R}$ auf $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ mit $\tilde{X}_n((\omega_1, \omega_2, \dots)) := X(\omega_n)$. Wird für ein $A \in \mathcal{A}$ mit $P(A) \in (0, 1)$ die Zufallsvariable X durch

$$X(\omega) := \begin{cases} 1 & \text{für } \omega \in A, \\ 0 & \text{sonst.} \end{cases}$$

definiert [Bau68, S. 143], so ist \tilde{X}_j gleich 1, falls A beim j -ten Versuch eintritt und 0, falls A beim j -ten Versuch nicht eintritt. Daher entspricht $\sum_{j=1}^n \tilde{X}_j$ anschaulich der Anzahl an Versuchen aus n Durchführungen, bei denen A eingetreten ist und damit $\frac{1}{n} \sum_{j=1}^n \tilde{X}_j$ der relativen Häufigkeit des Ereignisses A unter den ersten n Durchführungen. Obwohl es $(\omega_1, \omega_2, \dots) \in \tilde{\Omega}$ mit $\omega_n \in A \forall n \in \mathbb{N}$ gibt und für solche $(\omega_1, \omega_2, \dots)$ gilt, dass $\frac{1}{n} \sum_{j=1}^n \tilde{X}_j(\omega_1, \omega_2, \dots) = 1 \neq P(A)$ ist, sollte gemäß der Erfahrung die relative Häufigkeit des Eintretens von A , also die Folge $\frac{1}{n} \sum_{j=1}^n \tilde{X}_j$, „mit großer Wahrscheinlichkeit“ gegen $P(A)$ konvergieren [Bau68, S. 143]. Man erhält einen geeigneten Konvergenzbegriff indem „mit großer Wahrscheinlichkeit“ durch „mit Wahrscheinlichkeit 1“ oder gleichbedeutend mit „fast sicher“ übersetzt wird. „Fast sichere“ Konvergenz lässt sich wie folgt formalisieren [Tap13, S. 111]:

Definition 2.1.14. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen auf (Ω, \mathcal{A}, P) und X eine Zufallsvariable auf (Ω, \mathcal{A}, P) . Dann *konvergiert* (X_n) *P -fast sicher gegen* X , falls $\exists A \in \mathcal{A}$ mit $P(A) = 1$, sodass $\forall \omega \in A : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$.

Der nachfolgende Satz [Bau68, S. 151, Satz 37.2] und die beiden daraus folgenden Korollare liefern die Bestätigung, dass die relative Häufigkeit des Eintretens eines Ereignisses „fast sicher“ gegen die Wahrscheinlichkeit des Ereignisses konvergiert.

Satz 2.1.15. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, integrierbaren Zufallsvariablen auf (Ω, \mathcal{A}, P) , die die gleiche Verteilung haben. Setze $\mu := E(X_1)$. Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = \mu \text{ } P\text{-fast sicher.}$$

Es gelten folgende Korollare aus dem starken Gesetz der großen Zahlen:

Korollar 2.1.15.1. Sei X eine integrierbare Zufallsvariable auf (Ω, \mathcal{A}, P) . Für $j \in \mathbb{N}$ definiere die Zufallsvariable \tilde{X}_j auf $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ durch $\tilde{X}_j((\omega_1, \omega_2, \dots)) := X(\omega_j)$. Dann gilt: $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \tilde{X}_j = E(X)$ \tilde{P} -fast sicher.

Beweis. Die Zufallsvariablen \tilde{X}_j sind integrierbar, unabhängig und haben alle die gleiche Verteilung. Aus Satz 2.1.15. folgt daher die Behauptung. □

Korollar 2.1.15.2. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $A \in \mathcal{A}$ und $X : \Omega \rightarrow \{0, 1\}$ mit

$$X(\omega) := \begin{cases} 1 & \text{für } \omega \in A, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \tilde{X}_j = P(A) \text{ } \tilde{P}\text{-fast sicher.}$$

Beweis. $E(X) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$. Wegen Korollar 2.1.15.1. gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \tilde{X}_j = P(A) \text{ } \tilde{P}\text{-fast sicher.}$$

□

Die relative Häufigkeit des Eintretens eines Ereignisses A , $\frac{1}{n} \sum_{j=1}^n \tilde{X}_j$, konvergiert „ (\tilde{P}) -fast sicher“ gegen die Wahrscheinlichkeit von A , $P(A)$. Diese Aussage ist die Formalisierung der intuitiven Interpretation der Wahrscheinlichkeit innerhalb der axiomatisch aufgebauten Wahrscheinlichkeitstheorie.

Für einen Zusammenhang zwischen der Boltzmann- und der Shannon-Entropie wird eine andere Interpretationsmöglichkeit von $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, die aus [Bau68, S. 135] zu entnehmen ist, verwendet: $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ beschreibt das Zufallsgeschehen, bei dem ein Experiment abzählbar unendlich oft, unabhängig nebeneinander, durchgeführt wird.

Sind die abzählbar unendlich viele gleichzeitigen Durchführungen durchnummeriert, so ist $\tilde{\Omega}$ die Menge aller Ergebnisse dieses Zufallsgeschehens. Dabei wird $(\omega_1, \omega_2, \dots) \in \tilde{\Omega}$ so gedeutet, dass ω_i das Ergebnis der Durchführung mit der Nummer $i \in \mathbb{N}$ ist.

Werden endlich viele der Versuchsdurchführungen (mit den Nummern $j_1 < j_2 < \dots < j_m$) betrachtet, so soll, da sich die Versuchsdurchführungen nicht gegenseitig beeinflussen, folgendes gelten: Die Wahrscheinlichkeit, dass beim j_k -ten Versuch das Ereignis A_{j_k} , $k \in \{1, \dots, m\}$, eintritt, soll gleich $\prod_{k=1}^m P(A_{j_k})$ sein. Daher wird als σ -Algebra

$$\mathcal{A}^* := \sigma(\{(\omega_1, \omega_2, \dots) \in \tilde{\Omega} : \omega_{j_1} \in A_{j_1}, \omega_{j_2} \in A_{j_2}, \dots, \omega_{j_m} \in A_{j_m}\} : \\ m \in \mathbb{N}, A_{j_1}, A_{j_2}, \dots, A_{j_m} \in \mathcal{A}\})$$

gewählt. Es wird ein Maß P^* auf \mathcal{A}^* gesucht, welches

$$P^*(\{(\omega_1, \omega_2, \dots) \in \tilde{\Omega} : \omega_{j_1} \in A_{j_1}, \omega_{j_2} \in A_{j_2}, \dots, \omega_{j_m} \in A_{j_m}\}) = \prod_{k=1}^m P(A_{j_k})$$

erfüllt. Aufgrund von

$$\{(\omega_1, \omega_2, \dots) \in \tilde{\Omega} : \omega_{j_1} \in A_{j_1}, \omega_{j_2} \in A_{j_2}, \dots, \omega_{j_m} \in A_{j_m}\} \\ = [\underbrace{\Omega, \dots, \Omega}_{(j_1-1)\text{-mal}}, A_{j_1}, \underbrace{\Omega, \dots, \Omega}_{(j_2-1)\text{-mal}}, A_{j_2}, \dots, \Omega, A_{j_m}]$$

folgt, dass die Familie aller Mengen $\{(\omega_1, \omega_2, \dots) \in \tilde{\Omega} : \omega_{j_1} \in A_{j_1}, \omega_{j_2} \in A_{j_2}, \dots, \omega_{j_m} \in A_{j_m}\}$ gleich der Menge aller Zylindermengen ist. Daher ist $\mathcal{A}^* = \tilde{\mathcal{A}}$. Außerdem gilt

$$\tilde{P}(\{(\omega_1, \omega_2, \dots) \in \tilde{\Omega} : \omega_{j_1} \in A_{j_1}, \omega_{j_2} \in A_{j_2}, \dots, \omega_{j_m} \in A_{j_m}\}) = \prod_{k=1}^m P(A_{j_k}).$$

Es gilt folgende Proposition:

Proposition 2.1.16. *Die Menge der Zylindermengen ist schnittstabil: Der Durchschnitt zweier Zylindermengen ist eine Zylindermenge.*

Beweis. Seien $[A_1, \dots, A_r]$ und $[B_1, \dots, B_m]$ zwei Zylindermengen. Dann gilt

$$\begin{aligned} r = m & : [A_1, \dots, A_r] \cap [B_1, \dots, B_r] = [A_1 \cap B_1, \dots, A_r \cap B_r] \\ m > r & : [A_1, \dots, A_r] \cap [B_1, \dots, B_m] = [A_1 \cap B_1, \dots, A_r \cap B_r, B_{r+1}, \dots, B_m] \\ r > m & : [A_1, \dots, A_r] \cap [B_1, \dots, B_m] = [A_1 \cap B_1, \dots, A_r \cap B_r, A_{r+1}, \dots, A_m] \end{aligned}$$

□

Aus dem Eindeutigkeitsatz für Maße [Geo09, S. 16, Satz 1.12] folgt daher, dass P^* eindeutig durch \tilde{P} bestimmt ist. Insgesamt lässt sich also $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ als der Wahrscheinlichkeitsraum interpretieren, der beschreibt, dass ein Zufallsgeschehen unendlich oft unabhängig nebeneinander ausgeführt wird.

Für die Definition des Boltzmann'schen Entropiebegriffes wird zusätzlich ein Begriff aus der Kombinatorik benötigt:

Definition 2.1.17. Sei $n \in \mathbb{N}$ und $(k_1, k_2, \dots, k_r) \in \mathbb{N}^r$ mit $k_1 + k_2 + \dots + k_r = n$. Dann heißt

$$\binom{n}{k_1, k_2, \dots, k_r} := \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!}$$

Multinomialkoeffizient [KW08, S.26].

$\binom{n}{k_1, k_2, \dots, k_r}$ entspricht der Anzahl an Möglichkeiten, n Objekte so auf r Fächer zu verteilen, dass das j -te Fach genau k_j Objekte enthält [KW08, S.26].

2.2. Dynamische Systeme

Zu Beginn werden die Begriffe definiert, die bei der Behandlung dynamischer Systeme von Bedeutung sind:

Definition 2.2.1. Sei X eine Menge, $d : X \times X \rightarrow \mathbb{R}$ eine Funktion. d heißt eine *Metrik* auf X und (X, d) ein *metrischer Raum*, wenn folgende Bedingungen erfüllt sind:

- (i) Für alle $x, y \in X$ ist $d(x, y) \geq 0$. Dabei gilt $d(x, y) = 0$ genau dann, wenn $x = y$.
- (ii) Für alle $x, y \in X$ gilt $d(x, y) = d(y, x)$.
- (iii) Sind $x, y, z \in X$, so gilt die *Dreiecksungleichung*:

$$d(x, z) \leq d(x, y) + d(y, z)$$

[Kal15, S. 69-70].

Proposition 2.2.2. *In einem metrischen Raum (X, d) gilt*

$$|d(x, z) - d(z, y)| \leq d(x, y) \quad \forall x, y, z \in X.$$

Beweis. Es gilt $d(x, z) \leq d(x, y) + \underbrace{d(y, z)}_{=d(z, y)} \Rightarrow d(x, z) - d(z, y) \leq d(x, y)$. In gleicher Weise erhält man $d(z, y) \leq \underbrace{d(z, x)}_{=d(x, z)} + d(x, y) \Rightarrow d(z, y) - d(x, z) \leq d(x, y)$. Insgesamt gilt daher $|d(x, z) - d(z, y)| \leq d(x, y)$. □

Definition 2.2.3. Seien (X, d_X) und (Y, d_Y) metrische Räume, $D \subseteq X$ und $f : D \rightarrow Y$ eine Funktion. f heißt *stetig an einer Stelle $x \in D$* , wenn gilt

$$\forall \varepsilon > 0 \exists \delta > 0, \text{ so dass für alle } t \in D \text{ mit } d_X(t, x) < \delta : d_Y(f(t), f(x)) < \varepsilon$$

[Kal15, S. 165].

Definition 2.2.4. Sei (X, d) ein metrischer Raum, $x \in X$ und $r \in \mathbb{R}$, $r > 0$. Dann wird die Menge

$$B(x, r) := \{y \in X : d(y, x) < r\}$$

als *offene Kugel um den Punkt x mit Radius r* bezeichnet [Kal15, S. 127].

Definition 2.2.5. Sei (X, d) ein metrischer Raum. Eine Teilmenge U von X heißt *offen*, wenn es zu jedem Punkt $x \in U$ ein $r > 0$ gibt, so dass $B(x, r) \subseteq U$ [Kal15, S. 127].

Unter einer *offenen Überdeckung* von $A \subseteq X$ versteht man eine Familie $(U_i)_{i \in I}$ von offenen Teilmengen $U_i \subseteq X$ mit $A \subseteq \bigcup_{i \in I} U_i$ [For13, S. 28].

Beispiel. Offene Kugeln $B(x, r) = \{y \in X : d(y, x) < r\}$ in metrischen Räumen sind offen [Kal15, S. 129].

Definition 2.2.6. Sei (X, d) ein metrischer Raum. Eine Menge $A \subseteq X$ heißt *kompakt*, wenn jede offene Überdeckung $(U_i)_{i \in I}$ von A eine endliche Teilüberdeckung besitzt, d.h. endlich viele Indizes $i_1, i_2, \dots, i_k \in I$ existieren, so dass $A \subseteq U_{i_1} \cup U_{i_2} \cup \dots \cup U_{i_k}$ [For13, S. 28] ist.

Wie der Begriff „Dynamisches System“ bereits andeutet, werden in der Theorie der dynamischen Systeme „Dynamiken“ in Abhängigkeit von der Zeit untersucht. Für nachfolgende Definition vergleiche [ES14, S. 1] und [Rai09, S. 3]:

Definition 2.2.7. Sei X ein nichtleerer kompakter metrischer Raum und $T : X \rightarrow X$ eine stetige Abbildung. Dann heißt (X, T) ein *diskretes topologisches dynamisches System*, wobei in dieser Arbeit meist nur kurz *topologisches dynamisches System* oder *dynamisches System* geschrieben wird.

Zur Veranschaulichung kann die Menge X für ein gegebenes physikalisches System als der Raum aller möglichen Zustände interpretiert werden und die Abbildung T (T für Transformation) kann als die diskrete zeitliche Entwicklung des Systems aufgefasst werden. Es wird $T(x) := Tx$ gesetzt. Für ein $x \in X$ ist $T^0x := x$ der Anfangszustand des Systems und für $n \in \mathbb{N}$ ist $T^n x$ der Zustand des Systems zum Zeitpunkt n . Bei der Analyse eines solchen Systems ist insbesondere das Langzeitverhalten des Systems und damit die Folgen $(T^n x)_{n \in \mathbb{N}_0}$ für $x \in X$ von Interesse [ES14, S. 1].

Definition 2.2.8. Sei (X, T) ein diskretes topologisches dynamisches System und $x \in X$. Die Folge $(T^n x)_{n \in \mathbb{N}_0} = (x, Tx, T^2x, \dots)$ wird der *Orbit* von x genannt. Der Vektor $(x, Tx, \dots, T^{n-1}x)$ wird als *Orbit der Länge n* bezeichnet [Rai09, S. 4].

Folgende Proposition, welche zwei allgemeine Eigenschaften des Urbilds einer Funktion angibt, wird in der Betrachtung dynamischer Systeme Anwendung finden. Für die Proposition und die Vorgehensweise des Beweises vergleiche [SS09, S. 162, Bsp. 4.3.15]

Proposition 2.2.9. Sei $f : A \rightarrow B$ eine Funktion, I eine beliebige Indexmenge und seien für alle $i \in I$ $B_i \subseteq B$ und $B_1, B_2 \subseteq B$. Dann gelten:

- (i) $f^{-1}(\bigcup_{i \in I} B_i) = \bigcup_{i \in I} f^{-1}B_i$
- (ii) $f^{-1}(B_1 \setminus B_2) = (f^{-1}B_1) \setminus (f^{-1}B_2)$.

Beweis. Diese beiden Eigenschaften lassen sich beziehentlich auf elementare Definitionen der Mengenlehre beweisen. Zu (i):

$$\begin{aligned} a \in f^{-1}\left(\bigcup_{i \in I} B_i\right) &\Leftrightarrow f(a) \in \bigcup_{i \in I} B_i \\ &\Leftrightarrow \exists i \in I \text{ mit } f(a) \in B_i \Leftrightarrow \exists i \in I \text{ mit } a \in f^{-1}B_i \\ &\Leftrightarrow a \in \bigcup_{i \in I} f^{-1}B_i. \end{aligned}$$

Beim Beweis von (ii) wird analog zu (i) vorgegangen:

$$\begin{aligned} a \in f^{-1}(B_1 \setminus B_2) &\Leftrightarrow f(a) \in B_1 \setminus B_2 \\ &\Leftrightarrow f(a) \in B_1 \wedge f(a) \notin B_2 \Leftrightarrow a \in f^{-1}B_1 \wedge a \notin f^{-1}B_2 \\ &\Leftrightarrow a \in (f^{-1}B_1) \setminus (f^{-1}B_2). \end{aligned}$$

□

3. Entropiebegriff der Informationstheorie

3.1. Informationsgehalt

Wie schon beim Begriff der Wahrscheinlichkeit gilt es auch beim Begriff der Information einen intuitiven Begriff mathematisch exakt zu definieren. Dabei wird folgende Ausgangssituation betrachtet [EM81, S. 51]: Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, der ein Zufallsexperiment beschreibt und das Ereignis A tritt beim Experiment ein. Die Aufgabe ist eine Abbildung I zu finden, so dass $I(A)$ ein quantitatives Maß für den Informationsgehalt der Information darstellt, dass das Ereignis A eingetreten ist. Dabei soll der Informationsgehalt nur von der Wahrscheinlichkeit des Ereignisses A abhängen. Für die folgenden Eigenschaften vergleiche [Sch15, S. 13] und [EM81, S. 52]:

Die erste Eigenschaft des Informationsgehalts soll sein, dass ein Ereignis, dessen Eintreten mehr „überrascht“ als ein anderes, einen größeren Informationsgehalt hat als das andere Ereignis. Dies lässt sich wie folgt mit dem Begriff der Wahrscheinlichkeit in Verbindung bringen: Je unwahrscheinlicher ein Ereignis ist, desto mehr überrascht dessen Eintreffen und umso größer ist der Informationsgehalt dieses Ereignisses.

Sind zwei Ereignisse unabhängig, so beeinflusst das Eintreten des einen Ereignisses nicht die Wahrscheinlichkeit des anderen Ereignisses und umgekehrt. Wird die Information erhalten, dass zwei unabhängige Ereignisse eingetreten sind, soll der Gehalt dieser Information der Summe der Informationsgehalte der beiden Ereignisse entsprechen.

Wie auch bei der Messung von Masse oder Zeit wird eine Einheit benötigt. Daher wird die Informationsmenge eines Ereignisses mit Wahrscheinlichkeit $\frac{1}{\xi}$, $\xi \in \mathbb{R}$, $\xi > 1$, gleich 1 gesetzt.

Die Information, dass ein sicheres Ereignis eingetreten ist, soll den Informationsgehalt 0 und die Information, dass ein unmögliches Ereignis eingetreten ist den Informationsgehalt $+\infty$ erhalten.

Dies führt zu folgender Definition [EM81, S. 52] [Sch15, S. 13 und S. 17]:

Definition 3.1.1. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Die *Information* I ist eine Abbildung $I : \mathcal{A} \rightarrow [0, +\infty]$ mit $I(A) := \Lambda(P(A))$, wobei $\Lambda : [0, 1] \rightarrow [0, +\infty]$ die Abbildung mit folgenden Eigenschaften ist:

- (i) Λ ist streng monoton fallend: Für $x < y \in [0, 1]$ folgt $\Lambda(x) > \Lambda(y)$.
- (ii) Für alle $x, y \in [0, 1] : \Lambda(x \cdot y) = \Lambda(x) + \Lambda(y)$
- (iii) Für $\frac{1}{\xi}, \xi \in \mathbb{R}, \xi > 1$, gilt $\Lambda(\frac{1}{\xi}) = 1$
- (iv) $\Lambda(1) = 0$ und $\Lambda(0) = +\infty$.

Es wird nun gezeigt, dass Λ durch diese Eigenschaften eindeutig bestimmt ist. Für Proposition 3.1.2. und Satz 3.1.3. und deren Beweise vergleiche [Sch15, S. 14-15].

Proposition 3.1.2. Sei $f : (0, +\infty) \rightarrow \mathbb{R}$ mit $f(x \cdot y) = f(x) + f(y)$ und $a \in \mathbb{R}, a > 0$ beliebig. Dann gelten:

- (i) $f(1) = 0$
- (ii) $n \in \mathbb{Z} : f(a^n) = n \cdot f(a)$
- (iii) $q \in \mathbb{Q} : f(a^q) = q \cdot f(a)$

Beweis.

- (i) $f(1) = f(1 \cdot 1) = f(1) + f(1) \Rightarrow f(1) = 0$.
- (ii) Sei zuerst $n \in \mathbb{N}_0$. Es wird ein Beweis durch Induktion nach n durchgeführt:
Für $n = 0$ gilt $f(a^0) = f(1) = 0 = 0 \cdot f(a)$.
Die Induktionsvoraussetzung ist: Für $n > 0$ gilt $f(a^n) = n \cdot f(a)$.
Sei $n > 0$. Dann ist

$$\begin{aligned} f(a^n) &= f(a^{n-1} \cdot a) = f(a^{n-1}) + f(a) \\ &\stackrel{\text{IV}}{=} (n-1) \cdot f(a) + f(a) = n \cdot f(a). \end{aligned}$$

Damit ist $f(a^n) = n \cdot f(a)$ für $n \in \mathbb{N}_0$ bewiesen.

Sei nun $n \in \mathbb{Z}$, $n < 0 \Rightarrow -n > 0$. Daraus ergibt sich

$$\begin{aligned} 0 &= f(1) = f(a^n \cdot a^{-n}) = f(a^n) + f(a^{-n}) \\ &= f(a^n) + (-n) \cdot f(a) \Rightarrow f(a^n) = n \cdot f(a). \end{aligned}$$

(iii) Sei $q \in \mathbb{Q}$, also $\exists m \in \mathbb{Z}$, $n \in \mathbb{N}$ mit $q = \frac{m}{n}$.

$$\begin{aligned} f(a) &= f(a^{\frac{n}{n}}) = n \cdot f(a^{\frac{1}{n}}) \Rightarrow f(a^{\frac{1}{n}}) = \frac{1}{n} \cdot f(a) \\ \Rightarrow f(a^q) &= f(a^{\frac{m}{n}}) = m \cdot f(a^{\frac{1}{n}}) = \frac{m}{n} \cdot f(a) = q \cdot f(a). \end{aligned}$$

□

Folgender Satz lässt sich mit dieser Proposition beweisen:

Satz 3.1.3. *Sei $f : (0, 1] \rightarrow [0, +\infty)$. Dann $\exists a \leq 0$ so dass $f(x) = a \cdot \log(x)$ genau dann, wenn $f(x \cdot y) = f(x) + f(y) \forall x, y \in (0, 1]$ und f monoton fallend ist.*

Beweis. (\Rightarrow) Da e^x bijektiv und streng monoton steigend ist, ist auch die Umkehrfunktion $\log : (0, +\infty) \rightarrow \mathbb{R}$ streng monoton steigend. Damit ist \log ebenfalls auf $(0, 1]$ streng monoton steigend.

Ist $a < 0$, ist hiermit $f(x) = a \log(x)$ streng monoton fallend. Ist $a = 0$, so ist $f(x) = 0$ und daher monoton fallend.

Seien $x, y \in (0, 1]$ beliebig:

$$\begin{aligned} e^{\log x + \log y} &= \underbrace{e^{\log x}}_{=x} \cdot \underbrace{e^{\log y}}_{=y} = x \cdot y \\ \Rightarrow \log(x \cdot y) &= \log x + \log y \\ \Rightarrow f(x \cdot y) &= a \log(x \cdot y) = a \log x + a \log y = f(x) + f(y). \end{aligned}$$

(\Leftarrow) Für ein $x \in (0, 1]$ und $\xi \in \mathbb{R}$, $\xi > 1$ bel.: $-\frac{\log x}{\log \xi} \in \mathbb{R}$. Da \mathbb{Q} dicht in \mathbb{R} liegt, existieren zwei Folgen $(q_n)_{n \in \mathbb{N}}$, $(r_n)_{n \in \mathbb{N}}$ in \mathbb{Q} mit $q_n \leq -\frac{\log x}{\log \xi} \leq r_n$ und $\lim_{n \rightarrow \infty} q_n = \lim_{n \rightarrow \infty} r_n = -\frac{\log x}{\log \xi}$. Da $(\frac{1}{\xi})^x$ monoton fallend ist, folgt

$$\begin{aligned} \left(\frac{1}{\xi}\right)^{q_n} &\geq \underbrace{\left(\frac{1}{\xi}\right)^{-\frac{\log x}{\log \xi}}}_{=\xi^{\frac{\log x}{\log \xi}=x}} \geq \left(\frac{1}{\xi}\right)^{r_n} \quad f \text{ m.f.} \\ \underbrace{f\left(\left(\frac{1}{\xi}\right)^{q_n}\right)}_{=q_n \cdot f\left(\frac{1}{\xi}\right) \rightarrow -f\left(\frac{1}{\xi}\right) \frac{\log x}{\log \xi}} &\leq f(x) \leq \underbrace{f\left(\left(\frac{1}{\xi}\right)^{r_n}\right)}_{=-r_n \cdot f\left(\frac{1}{\xi}\right) \rightarrow -f\left(\frac{1}{\xi}\right) \frac{\log x}{\log \xi}} \quad \Rightarrow \\ -f\left(\frac{1}{\xi}\right) \frac{\log x}{\log \xi} &\leq f(x) \leq -f\left(\frac{1}{\xi}\right) \frac{\log x}{\log \xi} \quad \Rightarrow f(x) = -f\left(\frac{1}{\xi}\right) \frac{\log x}{\log \xi}. \end{aligned}$$

Mit $a := -\frac{f\left(\frac{1}{\xi}\right)}{\log \xi} \leq 0$ erhält man $f(x) = a \log x$. □

Aus dem Beweis des Satzes ergibt sich, dass eine Funktion f mit $f(x \cdot y) = f(x) + f(y) \forall x, y \in (0, 1]$ und f monoton fallend genau dann streng monoton fallend ist, wenn $a < 0$ ist. Aus obigem Satz und den Regeln für die erweiterten reellen Zahlen ergeben sich:

Korollar 3.1.3.1. Für die in Definition 3.1.1. beschriebene Abbildung $\Lambda : [0, 1] \rightarrow [0, +\infty]$ gilt:

$$\Lambda(x) = \begin{cases} -\frac{\log x}{\log \xi} = -\log_{\xi}(x) & \text{falls } x \in (0, 1], \\ +\infty & \text{falls } x = 0. \end{cases}$$

Anmerkung. Mit der Definition $\log 0 := -\infty$ und den Rechenregeln in den erweiterten reellen Zahlen lässt sich der oben beschriebene Ausdruck kürzer $\Lambda(x) = -\log_{\xi}(x)$ schreiben.

Korollar 3.1.3.2. Die in Definition 3.1.1. beschriebene Abbildung $I : \mathcal{A} \rightarrow [0, +\infty]$ hat die Abbildungsvorschrift $I(A) = -\log_{\xi}(P(A))$.

Da $\log x$ ein Vielfaches von $\log_{\xi} x$ ist ($\log x = \log \xi \cdot \log_{\xi} x$), wird wie in [EM81, S. 55] im weiteren Verlauf der Arbeit für ξ die Eulersche Zahl e gewählt. Damit ist $I(A) = -\log(P(A))$.

3.2. Shannon-Entropie

Aufbauend auf dem Begriff der Information wird nun die (*Shannon-*) *Entropie* für endliche Wahrscheinlichkeitsräume definiert.

Zunächst wird gezeigt, dass sich ein gegebener Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , wo-

bei Ω endlich mit $|\Omega| = n$ und $\mathcal{A} := \mathfrak{P}(\Omega)$ ist, eindeutig durch einen Vektor $\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$

$\in \mathbb{R}^n$ beschreiben lässt. Dies geht hervor aus:

Proposition 3.2.1. *Sei $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ eine endliche Menge (mit $|\Omega| = n$) und $\mathfrak{P}(\Omega)$ die Potenzmenge von Ω . Ist P ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}(\Omega)$, dann gilt für den Vektor $(p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ mit $p_j := P(\alpha_j) \forall j \in \{1, 2, \dots, n\}: p_j \geq 0 \forall j \in \{1, 2, \dots, n\}$ und $\sum_{j=1}^n p_j = 1$.*

Falls umgekehrt $(p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ die Eigenschaften $p_j \geq 0 \forall j \in \{1, 2, \dots, n\}$ und $\sum_{j=1}^n p_j = 1$ erfüllt, dann ist durch

$$P(A) := \sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A}} p_j \text{ für } A \subseteq \Omega$$

ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}(\Omega)$ gegeben.

Beweis. (\Rightarrow) $p_j = P(\{\alpha_j\}) \geq 0$ gilt, da für ein Wahrscheinlichkeitsmaß P gilt: $P(A) \geq 0$. Da die Mengen $\{\alpha_j\}$ paarweise disjunkt sind, folgt

$$\sum_{j=1}^n p_j = \sum_{j=1}^n P(\{\alpha_j\}) = P(\underbrace{\bigcup_{j=1}^n \{\alpha_j\}}_{=\Omega}) = P(\Omega) = 1.$$

(\Leftarrow) Es werden die Eigenschaften eines Wahrscheinlichkeitsmaßes überprüft:

$$(i) \quad P(A) = \sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A}} \underbrace{p_j}_{\geq 0} \geq 0$$

$$(ii) \quad P(\Omega) = \sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in \Omega}} p_j = \sum_{j=1}^n p_j = 1$$

(iii) Sei $(A_k)_{k=1}^m$, wobei $1 \leq m \leq n$, eine Familie von paarweisen disjunkten Mengen $A_k \subseteq \Omega$. Aus der Definition von P ergibt sich:

$$\begin{aligned} P\left(\bigcup_{k=1}^m A_k\right) &= \sum_{\substack{j \in \{1,2,\dots,n\} \\ \alpha_j \in \bigcup_{k=1}^m A_k}} p_j \stackrel{A_k \text{ paarw. disj.}}{=} \\ &= \underbrace{\sum_{\substack{j \in \{1,2,\dots,n\} \\ \alpha_j \in A_1}} p_j}_{=P(A_1)} + \underbrace{\sum_{\substack{j \in \{1,2,\dots,n\} \\ \alpha_j \in A_2}} p_j}_{=P(A_2)} + \cdots + \underbrace{\sum_{\substack{j \in \{1,2,\dots,n\} \\ \alpha_j \in A_m}} p_j}_{=P(A_m)} = \sum_{k=1}^m P(A_k). \end{aligned}$$

Damit ist P ein Wahrscheinlichkeitsmaß. □

Definition 3.2.2. Ein Vektor $(p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ mit $p_j \geq 0 \forall j \in \{1, 2, \dots, n\}$ und $\sum_{j=1}^n p_j = 1$ heißt *Wahrscheinlichkeitsverteilung* oder *Wahrscheinlichkeitsvektor* auf $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$.

Bei gegebener endlicher Ergebnismenge $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ und Verteilung p auf Ω stellt sich die Frage, wie viel Information die Durchführung des durch Ω beschriebenen Zufallsgeschehens durchschnittlich liefert [Sch15, S. 26]. Wird das Experiment s mal, $s \in \mathbb{N}$, durchgeführt, und tritt das Ereignis $\{\alpha_j\}$ in den s Versuchen $N_s(\{\alpha_j\})$ mal ein, so liefert dies den Informationsgehalt $N_s(\{\alpha_j\}) \cdot I(\{\alpha_j\})$. Der durchschnittliche Informationsgehalt bei s Durchführungen ist daher

$$\begin{aligned} & \frac{N_s(\{\alpha_1\}) \cdot I(\{\alpha_1\}) + N_s(\{\alpha_2\}) \cdot I(\{\alpha_2\}) + \cdots + N_s(\{\alpha_n\}) \cdot I(\{\alpha_n\})}{s} \\ &= \frac{N_s(\{\alpha_1\})}{s} \cdot I(\{\alpha_1\}) + \frac{N_s(\{\alpha_2\})}{s} \cdot I(\{\alpha_2\}) + \cdots + \frac{N_s(\{\alpha_n\})}{s} \cdot I(\{\alpha_n\}). \end{aligned}$$

Werden die relativen Häufigkeiten $\frac{N_s(\{\alpha_j\})}{s}$ mit der in Kapitel 2.1 angeführten Idee, dass die relativen Häufigkeiten „fast sicher“ gegen die Wahrscheinlichkeiten $P(\{\alpha_j\}) = p_j$ konvergieren, durch p_j ersetzt und wird der Informationsgehalt durch den in Korollar 3.1.3.2. gefundenen Ausdruck $I(\{\alpha_j\}) = -\log(P(\{\alpha_j\})) = -\log p_j$ dargestellt, so

führt dies zum Ausdruck

$$-p_1 \log p_1 + (-p_2 \log p_2) + \cdots + (-p_n \log p_n) = - \sum_{j=1}^n p_j \log p_j.$$

Der soeben abgeleitete Ausdruck wird die *Shannon-Entropie* des Wahrscheinlichkeitsvektors (p_1, p_2, \dots, p_n) genannt [Sha48, S. 393-394]:

Definition 3.2.3. Für eine Wahrscheinlichkeitsverteilung (p_1, p_2, \dots, p_n) sei

$$h \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} := - \sum_{j=1}^n p_j \log p_j$$

die *Shannon-Entropie* von (p_1, p_2, \dots, p_n) , wobei $-0 \log 0 := 0$ gesetzt wird.

Wie vorhergehend dargestellt, lässt sich dieser Ausdruck als der „fast sichere“ Grenzwert des durchschnittlichen Informationsgehalts eines endlichen Wahrscheinlichkeitsraums bei unendlich vielen Versuchsdurchführungen interpretieren. Dies lässt sich auch als die „zu erwartende“ Informationsmenge dieses Wahrscheinlichkeitsraums interpretieren.

Die Festlegung $-0 \log 0 := 0$ wird wie folgt begründet:

Proposition 3.2.4. $\lim_{x \rightarrow 0^+} (-x \log x) = 0.$

Beweis. Dies lässt sich mit der Regel von de l'Hospital [For16, S. 190, Satz 10] beweisen:

$$\lim_{x \rightarrow 0^+} (-x \log x) = \lim_{x \rightarrow 0^+} \left(\frac{\log x}{-\frac{1}{x}} \right) = \lim_{x \rightarrow 0^+} \frac{\frac{1}{x}}{\frac{1}{x^2}} = \lim_{x \rightarrow 0^+} x = 0.$$

□

Anmerkung. Mit dieser Festlegung ist die Funktion $h : [0, +\infty)^n \rightarrow [0, +\infty)$, $h(p) = - \sum_{j=1}^n p_j \log p_j$ stetig.

3.3. Shannon-Entropie eines Produktraums

Wie in Kapitel 1. angeführt, wird in dieser Arbeit ein Teilchensystem aus unabhängigen und nicht unterscheidbaren Teilchen betrachtet, um Gemeinsamkeiten und formale Zusammenhänge zwischen den behandelten Entropieausdrücken zu ermitteln. Dabei wird wie in [Geo09, S. 35] angenommen, dass die Ununterscheidbarkeit der Teilchen nur an den mangelhaften experimentellen Möglichkeiten liegt, die Teilchen aber „in Wirklichkeit“ mit $1, \dots, N$ durchnummeriert werden können. Die Menge $\Omega = \{\alpha_1, \dots, \alpha_n\}$ wird als die Menge der (Energie-) Zustände eines Teilchens interpretiert. Für eine beliebige Wahrscheinlichkeitsverteilung $p = (p_1, \dots, p_n)$ wird $p_j = P(\alpha_j)$ als die Wahrscheinlichkeit, dass sich ein Teilchen im Zustand α_j befindet, angesehen.

Damit die Entropiebegriffe nach Boltzmann und Shannon am Beispiel eines solchen Teilchensystems verglichen werden können, wird ein geeigneter Wahrscheinlichkeitsraum benötigt. Dieser lässt sich analog zu den Überlegungen in Kapitel 2.1. bezüglich $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ finden. Dass mehrere unabhängige und ununterscheidbare Teilchen gleichzeitig betrachtet werden, lässt sich als ein Experiment interpretieren, bei dem ein Zufallsgeschehen öfters nebeneinander ausgeführt wird (wie z.B. beim gleichzeitigen Werfen von N Würfeln [Bau68, S. 112]). Als Ergebnismenge für ein solches Experiment wird Ω^N betrachtet und für die σ -Algebra auf Ω^N wird $\mathfrak{P}(\Omega)^{\otimes N} = \mathfrak{P}(\Omega^N)$ gewählt. Für die Wahrscheinlichkeitsverteilung auf $\mathfrak{P}(\Omega^N)$ wird $p^{\otimes N}$ herangezogen [Bau68, S. 112] (für die Bezeichnung $p^{\otimes N}$ in der nachfolgenden Proposition vergleiche [Kle13, S. 284, Satz 14.14]):

Proposition 3.3.1. *Sei $\Omega = \{\alpha_1, \dots, \alpha_n\}$ eine endliche Menge mit Wahrscheinlichkeitsverteilung $p = (p_1, \dots, p_n)$. Für die ebenfalls endliche Menge Ω^N , $N \in \mathbb{N}$, erfüllt $p^{\otimes N} : \{1, \dots, n\}^N \rightarrow [0, 1]$ gegeben durch $p_J^{\otimes N} := \prod_{k=1}^N p_{J_k}$ für $J = (J_1, \dots, J_N) \in \{1, \dots, n\}^N$ die Eigenschaften einer Wahrscheinlichkeitsverteilung auf Ω^N .*

3.3. SHANNON-ENTROPIE EINES PRODUKTRAUMS

Beweis. Für $N \in \mathbb{N}$ beliebig gilt: Für alle $J \in \{1, \dots, n\}^N$ ist $p_J^{\otimes N} = \prod_{k=1}^N \underbrace{p_{J_k}}_{\geq 0} \geq 0$.

Um $\sum_{J \in \{1, \dots, n\}^N} p_J^{\otimes N} = 1$ zu zeigen, wird eine Induktion nach $N \in \mathbb{N}$ durchgeführt. Für $N = 1$ ist $p^{\otimes 1} = p$ und daher eine Wahrscheinlichkeitsverteilung.

Sei nun $N > 1$.

$$\sum_{J \in \{1, \dots, n\}^N} p_J^{\otimes N} = \sum_{J \in \{1, \dots, n\}^N} \prod_{k=1}^N p_{J_k} = \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \quad (1)$$

$$= \sum_{J_N=1}^N \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_{N-1}=1}^N \underbrace{\prod_{k=1}^N p_{J_k}}_{= p_{J_N} \prod_{k=1}^{N-1} p_{J_k}} \quad (2)$$

$$= \sum_{J_N=1}^N p_{J_N} \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_{N-1}=1}^N \prod_{k=1}^{N-1} p_{J_k} \quad (3)$$

$$= \sum_{J_1=1}^N p_{J_N} \underbrace{\sum_{L \in \{1, \dots, n\}^{N-1}} \prod_{k=1}^{N-1} p_{L_k}}_{\stackrel{\text{IV}}{=} 1} = \sum_{J_N=1}^N p_{J_N} = 1, \quad (4)$$

wobei in Zeile (2) die Summationsreihenfolge vertauscht wurde. □

Wie folgender Satz besagt, lässt sich die Entropie von $p^{\otimes N}$ einfach berechnen:

Satz 3.3.2. *Seien Ω^N und $p^{\otimes N}$ definiert wie in Proposition 3.3.1. Dann gilt:*

$$h(p^{\otimes N}) = N \cdot h(p),$$

wobei bei der linken Seite der Gleichung h auf $[0, +\infty)^{n^N}$ und bei der rechten Seite auf $[0, +\infty)^n$ definiert betrachtet wird.

3.3. SHANNON-ENTROPIE EINES PRODUKTRAUMS

Beweis. Für $N = 1$ ist $p^{\otimes 1} = p$ und somit $h(p^{\otimes 1}) = h(p) = 1 \cdot h(p)$.

Für $N > 1$ gilt

$$\begin{aligned}
 h(p^{\otimes N}) &= - \sum_{J \in \{1, \dots, n\}^N} p_J^{\otimes N} \log p_J^{\otimes N} \\
 &= - \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \underbrace{\log \prod_{k=1}^N p_{J_k}}_{\sum_{i=1}^N \log p_{J_i}} \\
 &= - \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \sum_{i=1}^N \log p_{J_i} \\
 &= - \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \sum_{i=1}^N \prod_{k=1}^N p_{J_k} \log p_{J_i} \\
 &= \sum_{i=1}^N \left(- \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \log p_{J_i} \right).
 \end{aligned}$$

Für $i \in \{1, \dots, N\}$ wird nun $-\sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \log p_{J_i}$ betrachtet. Ähnlich wie im Beweis von Proposition 3.3.1. wird der Index J_i an den Anfang der Summe gebracht und danach wird Proposition 3.3.1. angewendet:

$$\begin{aligned}
 - \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \log p_{J_i} &= - \sum_{J_i=1}^N \sum_{J_1=1}^N \cdots \sum_{J_{i-1}=1}^N \sum_{J_{i+1}=1}^N \cdots \sum_{J_N=1}^N p_{J_i} \log p_{J_i} \prod_{\substack{k=1 \\ k \neq i}}^N p_{J_k} \\
 &= - \sum_{J_i=1}^N p_{J_i} \log p_{J_i} \underbrace{\sum_{J_1=1}^N \cdots \sum_{J_{i-1}=1}^N \sum_{J_{i+1}=1}^N \cdots \sum_{J_N=1}^N \prod_{\substack{k=1 \\ k \neq i}}^N p_{J_k}}_{\stackrel{1}{=} \text{Proposition 3.3.1}} \\
 &= - \sum_{J_i=1}^N p_{J_i} \log p_{J_i} = h(p).
 \end{aligned}$$

3.4. SHANNON-ENTROPIE FÜR KONTINUIERLICHE ERGEBNISMENGEN

Damit erhält man weiter

$$\sum_{i=1}^N \left(- \sum_{J_1=1}^N \sum_{J_2=1}^N \cdots \sum_{J_N=1}^N \prod_{k=1}^N p_{J_k} \log p_{J_k} \right) = \sum_{i=1}^N h(p) = N \cdot h(p).$$

□

Für ein Teilchensystem aus N unabhängigen und nicht unterscheidbaren Teilchen ist nach Satz 3.3.2 die Shannon-Entropie durch $h(p^{\otimes N}) = N \cdot h(p)$ bestimmt. Daher ist für ein derartiges System die Shannon-Entropie *pro Teilchen* durch $\frac{h(p^{\otimes N})}{N} = \frac{N \cdot h(p)}{N} = h(p)$ gegeben. Die Shannon-Entropie pro Teilchen entspricht also der Shannon-Entropie eines Teilchens für ein System aus unabhängigen und nicht unterscheidbaren Teilchen.

3.4. Shannon-Entropie für kontinuierliche Ergebnismengen

Als Beispiel wird ein einzelnes Teilchen eines Systems aus N unabhängigen und nicht unterscheidbaren Teilchen in einem kubischem Raum betrachtet. Die Entropie bezüglich der Verteilung dieses Teilchens in diesem Raum soll bestimmt werden. Die folgende Betrachtung erfolgt analog zur Betrachtung der Gleichverteilung in einem Intervall $[0, a]$ mit $a > 0$ aus [Sch15, S. 129]:

Mit Kenntnis von kontinuierlichen Zufallsvariablen kann als Ergebnismenge die Menge $[0, a]^3$, $a > 0 \in \mathbb{R}$, in Betracht gezogen werden. Wie in [Geo09, S. 13] wird für diese Menge die Borel'sche σ -Algebra $\mathcal{B}([0, a]^3)$ als σ -Algebra gewählt und P sei ein beliebiges Maß auf $\mathcal{B}([0, a]^3)$.

Definition 3.4.1. Es seien X_1, \dots, X_d reelle Zufallsvariablen. Falls sich die Verteilung von $X := (X_1, \dots, X_d)$ ausdrücken lässt durch

$$P_X(A) = \int_A \cdots \int p(x_1, \dots, x_d) dx_1 \dots dx_d,$$

für beliebige d -dimensionale Borelmengen $A \subseteq \mathcal{B}(\mathbb{R})$, dann hat X eine d -dimensionale kontinuierliche Verteilung und X ist eine kontinuierliche Zufallsvariable. Die Funktion $p(x_1, \dots, x_d)$ heißt *Wahrscheinlichkeitsdichte* [Iha93, S. 13].

3.4. SHANNON-ENTROPIE FÜR KONTINUIERLICHE ERGEBNISMENGEN

Bisher wurde der Entropiebegriff nur für Wahrscheinlichkeitsräume bezüglich endlicher Ergebnismengen definiert. Hier liegt jedoch eine zwar beschränkte, aber überabzählbar unendliche Ergebnismenge vor. Deshalb wird Shannons Entropiebegriff für solche Räume übernommen [Iha93, S. 13]:

Definition 3.4.2. Die *kontinuierliche* oder *differentielle Entropie* einer d -dimensionalen kontinuierlichen Zufallsvariablen $X = (X_1, \dots, X_d)$ mit der Wahrscheinlichkeitsdichte $p(x) = p(x_1, \dots, x_d)$ ist definiert als

$$h(X) := h(p) := - \int_{\Omega} p(x) \log p(x) dx,$$

sofern das Integral existiert.

Für die Verteilung des Teilchens im kubischen Raum kommt die Gleichverteilung auf $[0, a]^3$ in Frage. Die Dichtefunktion für eine gleichverteilte Zufallsvariable X auf $([0, a]^3, \mathcal{B}([0, a]^3))$ ist $p(x_1, x_2, x_3) = \frac{1}{a^3}$ [Iha93, S. 13].

Anmerkung. Die Dichtefunktion der Gleichverteilung ergibt sich wie folgt: Als Maß auf $\mathcal{B}([0, a]^3)$ wird $\frac{1}{a^3} \lambda^3$ betrachtet und es wird $X(\omega) := \omega$ gesetzt. Dadurch gilt für $A \in \mathcal{B}([0, a]^3)$ [For17, S. 45, Satz 6], dass

$$P(X \in A) = \frac{1}{a^3} \lambda^3(A) = \frac{1}{a^3} \int_A 1 dx = \int_A \frac{1}{a^3} dx.$$

Nach dem Satz von Fubini [For17, S. 88, Satz 2] und dem Satz, dass eine Riemannintegrierbare Funktion f auf einem abgeschlossenen Intervall $I \subseteq \mathbb{R}$ auch Lebesgueintegrierbar ist und die Integrale übereinstimmen [For17, S. 56, Satz 2], folgt

$$\begin{aligned} h(p) &= - \iiint_{[0, a]^3} \frac{1}{a^3} \log \frac{1}{a^3} d(x_1, x_2, x_3) \\ &= - \int_0^a \int_0^a \int_0^a \frac{1}{a^3} \log \frac{1}{a^3} dx_1 dx_2 dx_3 = -a^3 \frac{1}{a^3} \log \frac{1}{a^3} = 3 \log a. \end{aligned}$$

Wird $0 < a < 1$ gewählt, so ist $\log a < 0$ und daher $h(p) = 3 \log a < 0$. Übernimmt man die Interpretation der Shannon-Entropie für endliche Wahrscheinlichkeitsräume als Maß für die durchschnittliche Informationsmenge, so ist hier die durchschnittliche

3.4. SHANNON-ENTROPIE FÜR KONTINUIERLICHE ERGEBNISMENGEN

Informationsmenge negativ [Sch15, S. 129]. Dies schlägt sich mit der intuitiven Vorstellung, dass die geringste Informationsmenge, die übermittelt werden kann, gleich 0 ist.

Daher wird nun versucht, trotz der überabzählbar unendlichen Ergebnismenge die Shannon-Entropie für endliche Ergebnismengen anzuwenden. Dafür wird die Ergebnismenge durch eine Zerlegung „diskretisiert“ [ES14, S. 91] (Solche Zerlegungen von Mengen werden auch in Kapitel 5 zur Anwendung kommen):

Definition 3.4.3. Sei \mathcal{A} eine σ -Algebra auf einer Menge Ω . Eine endliche oder abzählbar unendliche Teilmenge \mathcal{Z} von $\mathcal{A} \setminus \{\emptyset\}$ heißt Zerlegung von Ω , wenn $\bigcup_{A \in \mathcal{Z}} A = \Omega$ gilt und die Mengen in \mathcal{Z} paarweise disjunkt sind.

In diesem Beispiel bietet sich für den Würfel eine Zerlegung in Würfel mit einer Seitenlänge $\frac{a}{m}$, $m \in \mathbb{N}$, an. Der Einfachheit halber wird nun für Ω der halboffene Würfel $[0, a)^3$ betrachtet. $[0, a)^3$ lässt sich in die Würfel

$$\left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right) \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right) \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right)$$

mit $j, k, l \in \{1, \dots, m\}$ zerlegen. Da $\left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right) \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right) \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right) \in \mathcal{B}([0, a)^3)$ und alle diese Teilwürfel disjunkt sind, ist daher $\mathcal{W} := \left\{ \left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right) \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right) \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right) : j, k, l \in \{1, \dots, m\} \right\}$ eine Zerlegung von $[0, a)^3$ mit $|\mathcal{W}| = m^3$. Folgende Proposition zeigt, wie man aus dem Wahrscheinlichkeitsmaß P durch eine Zerlegung einen Wahrscheinlichkeitsvektor p erhält.

Proposition 3.4.4. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und \mathcal{Z} eine Zerlegung von Ω . Dann wird durch $p_i := P(A_i)$ für $A_i \in \mathcal{Z}$, $i \in \mathbb{N}$ (oder $i \in \{1, \dots, n\}$, falls \mathcal{Z} endlich mit $|\mathcal{Z}| = n \in \mathbb{N}$ ist) eine Wahrscheinlichkeitsverteilung auf \mathcal{Z} definiert.

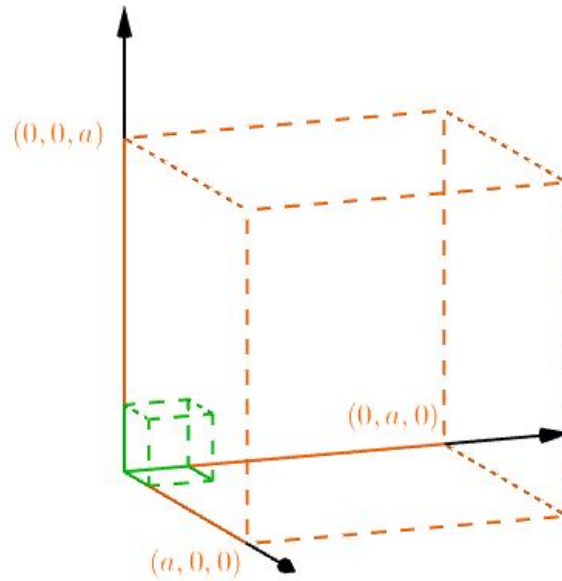


Abbildung 3.1.: Darstellung von $[0, a]^3$ (orange) und dem Würfel $[0, \frac{a}{m}]^3$ aus \mathcal{W} (grün).
Quelle: Eigene Darstellung.

Beweis. Da P ein Wahrscheinlichkeitsmaß ist, gilt $p_i = P(A_i) \geq 0$ und da zusätzlich alle A_i paarweise disjunkt sind, gilt

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P(A_i) = P\left(\underbrace{\bigcup_{i=1}^{\infty} A_i}_{=\Omega}\right) = P(\Omega) = 1.$$

Daher wird durch $p : \mathbb{N} \rightarrow [0, 1]$, $p_i := P(A_i)$ eine Wahrscheinlichkeitsverteilung definiert. □

Definition 3.4.5. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{Z} := \{A_1, \dots, A_n\}$ eine endliche Zerlegung von Ω . Die Entropie der Zerlegung \mathcal{Z} wird mit dem durch Anwendung von Proposition 3.4.4. erhaltenen Wahrscheinlichkeitsvektor $p := (P(A_1), \dots, P(A_n))$ definiert:

$$h(\mathcal{Z}) := h(p) = - \sum_{i=1}^n P(A_i) \log P(A_i).$$

Nach obiger Proposition wird durch $p : \{1, \dots, m\} \rightarrow [0, 1]$ mit

$$p_{jkl} := P \left(\left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right] \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right] \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right] \right)$$

eine Wahrscheinlichkeitsverteilung definiert. Die Shannon-Entropie ist dann durch

$$h(\mathcal{W}) = - \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m p_{jkl} \log p_{jkl}$$

gegeben.

Wird wieder $P = \frac{1}{a^3} \lambda^3$ gesetzt, so gilt

$$\begin{aligned} p_{jkl} &= P \left(\left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right] \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right] \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right] \right) \\ &= \frac{1}{a^3} \lambda^3 \left(\left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right] \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right] \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right] \right) \\ &= \frac{1}{a^3} \cdot \left(\frac{a \cdot j}{m} - \frac{a \cdot (j-1)}{m} \right) \cdot \left(\frac{a \cdot k}{m} - \frac{a \cdot (k-1)}{m} \right) \cdot \left(\frac{a \cdot l}{m} - \frac{a \cdot (l-1)}{m} \right) \\ &= \frac{1}{a^3} \cdot \frac{a^3}{m^3} = \frac{1}{m^3}. \end{aligned}$$

Für h bezüglich dieser Zerlegung ergibt sich

$$\begin{aligned} h(\mathcal{W}) &= - \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m p_{jkl} \log p_{jkl} = - \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m \frac{1}{m^3} \log \frac{1}{m^3} \\ &= -m^3 \cdot \frac{1}{m^3} \log \frac{1}{m^3} = \log m^3. \end{aligned}$$

Wie sich in nachfolgendem Unterkapitel herausstellen wird, ist dies das Maximum von h bezüglich dieser Zerlegung von $[0, a]^3$.

3.5. Maximale Shannon-Entropie

Um das Maximum von h auf der Menge $\{p \in [0, +\infty)^n : \sum_{j=1}^n p_j = 1\}$ zu bestimmen, wird folgende Proposition verwendet:

Proposition 3.5.1. Für $n \in \mathbb{N}$ gilt: die Menge $M_n := \{p \in [0, +\infty)^n : \sum_{j=1}^n p_j = 1\}$ ist kompakt.

Beweis. Sei $(p_k)_{k \in \mathbb{N}} = (p_k^{(1)}, p_k^{(2)}, \dots, p_k^{(n)})_{k \in \mathbb{N}}$ eine gegen $p = (p^{(1)}, p^{(2)}, \dots, p^{(n)}) \in \mathbb{R}^n$ konvergente Folge in M_n . Also gilt $p_k^{(j)} \geq 0 \forall j \in \{1, 2, \dots, n\}$ und $\sum_{j=1}^n p_k^{(j)} = 1$. Da $p_k \rightarrow p \Rightarrow p_k^{(j)} \rightarrow p^{(j)} \forall j \in \{1, 2, \dots, n\} \Rightarrow p^{(j)} \geq 0 \forall j \in \{1, 2, \dots, n\}$. Außerdem ist $1 = \sum_{j=1}^n p_k^{(j)} \rightarrow \sum_{j=1}^n p^{(j)} \Rightarrow \sum_{j=1}^n p^{(j)} = 1$. Daher ist $p \in M_n$ und somit M_n abgeschlossen.

Sei $x \in M_n$. Dann ist $0 \leq x_k \leq \sum_{j=1}^n x_j = 1 \Rightarrow x_k \in [0, 1] \Rightarrow M_n \subseteq [0, 1]^n$. Als Teilmenge einer beschränkten Menge ist M_n selbst beschränkt. Als abgeschlossene und beschränkte Teilmenge von \mathbb{R}^n ist M_n kompakt. \square

Nun wird das Maximum auf $\{p \in [0, +\infty)^n : \sum_{j=1}^n p_j = 1\}$ bestimmt:

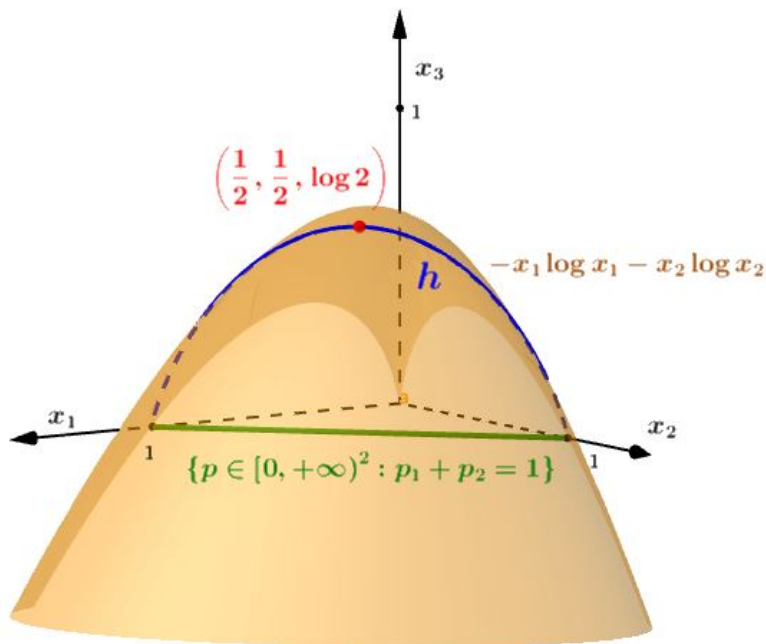


Abbildung 3.2.: Darstellung von h in zwei Variablen: In **braun**: Funktionsgraph von h auf $[0, +\infty)^2$. In **blau**: Funktionsgraph von h auf $\{p \in [0, +\infty)^2 : p_1 + p_2 = 1\}$.

Quelle: Eigene Darstellung.

Satz 3.5.2. *Ist (p_1, p_2, \dots, p_n) eine Wahrscheinlichkeitsverteilung, dann gilt $h(p_1, p_2, \dots, p_n) \leq \log n$. Dabei gilt $h(p_1, p_2, \dots, p_n) = \log n$ genau dann, wenn $p_j = \frac{1}{n}$ für alle $j \in \{1, 2, \dots, n\}$.*

Beweis. Es wird ein Beweis mittels Induktion nach n geführt.

Für $n = 1$ ist der Vektor $(p) = (1)$ der einzige Vektor, der die Eigenschaften aus Definition 3.2.2. erfüllt. Damit ist $h(p) = -1 \log 1 = 0 = \log 1$ und $p = \frac{1}{1}$.

Die Induktionsvoraussetzung ist:

$$\text{Für } n > 1 : h(p_1, \dots, p_n) \leq \log n.$$

Sei $n > 1$. Das Maximum von h auf $[0, +\infty)^n$ unter der Nebenbedingung $\sum_{j=1}^n p_j = 1$ soll gefunden werden. Um die Methode der Lagrange'schen Multiplikatoren [For13, S. 114, Satz 4] anwenden zu können, wird zunächst h auf der offenen Menge $(0, +\infty)^n$ betrachtet und anschließend mit Hilfe der Induktion gezeigt, dass der gefundene Kandidat für ein lokales Extremum auf $(0, +\infty)^n$ unter der Nebenbedingung $\sum_{j=1}^n p_j = 1$ das Maximum auf $[0, +\infty)^n$ unter dieser Nebenbedingung ist.

Betrachte dafür $h : (0, +\infty)^n \rightarrow \mathbb{R}$ unter der Nebenbedingung $\sum_{j=1}^n p_j = 1$, welche gleichbedeutend ist mit der Nebenbedingung $g(p_1, p_2, \dots, p_n) := \sum_{j=1}^n p_j - 1 = 0$. Definiere $F := h - \lambda g = -\sum_{j=1}^n p_j \log p_j - \lambda(\sum_{j=1}^n p_j - 1)$ mit $\lambda \in \mathbb{R}$.

Sei $k \in \{1, 2, \dots, n\}$. Dann ist

$$\begin{aligned} \frac{\partial F}{\partial p_k} &= \frac{\partial}{\partial p_k} \left(-\sum_{j=1}^n p_j \log p_j - \lambda \left(\sum_{j=1}^n p_j - 1 \right) \right) \\ &= -\sum_{j=1}^n \underbrace{\frac{\partial}{\partial p_k} p_j \log p_j}_{=\delta_{kj} \frac{\partial}{\partial p_k} p_j \log p_j} - \lambda \left(\sum_{j=1}^n \underbrace{\frac{\partial}{\partial p_k} p_j}_{=\delta_{kj}} - \underbrace{\frac{\partial}{\partial p_k} 1}_{=0} \right) \\ &= -\underbrace{\frac{\partial}{\partial p_k} p_k \log p_k}_{=-\log p_k - 1} - \lambda = -\log p_k - 1 - \lambda = 0 \\ &\Rightarrow \log p_k = -\lambda - 1 \Rightarrow p_k = e^{-\lambda - 1} \end{aligned}$$

Durch Einsetzen in die Nebenbedingung erhält man

$$\begin{aligned} 1 &= \sum_{j=1}^n p_j = \sum_{j=1}^n e^{-\lambda-1} = ne^{-\lambda-1} \\ \Rightarrow \frac{1}{n} &= e^{-\lambda-1} = p_j \quad \forall j \in \{1, 2, \dots, n\}. \end{aligned}$$

Die Entropie ist dann

$$h\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = - \sum_{j=1}^n \frac{1}{n} \underbrace{\log \frac{1}{n}}_{= -\log n} = \frac{1}{n} \sum_{j=1}^n \log n = \log n.$$

Betrachte nun $[0, +\infty)$ unter der Nebenbedingung $\sum_{j=1}^n p_j = 1$. Die Vektoren, die bisher nicht betrachtet wurden, sind $(p_1, p_2, \dots, p_n) \in [0, +\infty)$ mit $\sum_{j=1}^n p_j = 1$ und $\exists k \in \{1, \dots, n\}$ mit $p_k = 0$.

Sei daher $(p_1, p_2, \dots, p_n) \in [0, +\infty)$ mit $\sum_{j=1}^n p_j = 1$ und $\exists k \in \{1, \dots, n\}$ mit $p_k = 0$. Es gilt dann, dass $(p_1, p_2, \dots, p_{k-1}, p_{k+1}, \dots, p_n)$ ein Wahrscheinlichkeitsvektor ist und

$$h_{n-1}(p_1, p_2, \dots, p_{k-1}, p_{k+1}, \dots, p_n) = - \sum_{\substack{j=1 \\ j \neq k}}^n p_j \log p_j = - \sum_{j=1}^n p_j \log p_j = h_n(p_1, \dots, p_n),$$

wobei h_{n-1} bzw. h_n die Entropie eines $(n-1)$ -dimensionalen bzw. n -dimensionalen Wahrscheinlichkeitsvektors ist. Mit der Induktionsvoraussetzung gilt:

$$\begin{aligned} h_n(p_1, \dots, p_n) &= h_{n-1}(p_1, p_2, \dots, p_{k-1}, p_{k+1}, \dots, p_n) \\ &\stackrel{\text{IV}}{\leq} \log(n-1) \underset{\log \text{ str. m. st.}}{<} \log n = h_n\left(\frac{1}{n}, \dots, \frac{1}{n}\right). \end{aligned}$$

Da die Menge $M_n = \{p \in [0, +\infty)^n : \sum_{j=1}^n p_j = 1\}$ kompakt und h stetig ist $\Rightarrow \exists p_0 \in M_n$ mit $h(p_0) \geq h(p) \quad \forall p \in M_n$.

Außerdem gilt

$$M_n = \left\{ p \in (0, +\infty)^n : \sum_{j=1}^n p_j = 1 \right\} \\ \cup \left\{ p \in [0, +\infty)^n : \sum_{j=1}^n p_j = 1 \text{ und } \exists k \in \{1, \dots, n\} \text{ mit } p_k = 0 \right\}.$$

wobei die erste Menge mit $M_n^{(1)}$ und die zweite Menge mit $M_n^{(2)}$ bezeichnet werden soll und diese disjunkt sind.

Oben wurde gezeigt, dass aus der Induktionsvoraussetzung folgt, dass p_0 nicht in $M_n^{(2)}$ liegt. Daraus folgt weiter, dass $p_0 \in M_n^{(1)}$ und daher p_0 eine globale Maximumstelle in $M_n^{(1)}$ ist. Für eine beliebige offene Menge $V \subseteq (0, +\infty)^n$ mit $p_0 \in V$ ist infolgedessen erfüllt, dass $h(p) \leq h(p_0) \forall p \in V \cap M_n^{(1)}$ gilt. Also ist p_0 auch eine lokale Maximumstelle von h in $M_n^{(1)}$. Mit dem durch Anwendung der Lagrange'schen Multiplikatoren gezeigten Resultat, dass für eine lokale Extremstelle von h in $M_n^{(1)}$ gilt, dass $\forall j \in \{1, \dots, n\} p_j = \frac{1}{n}$ ist, folgt schließlich, dass der Vektor $p_0 = (\frac{1}{n}, \dots, \frac{1}{n})$ eine globale Maximumstelle in M_n ist und daher $h(p) \leq h(\frac{1}{n}, \dots, \frac{1}{n}) = \log n \forall p \in M_n$ gilt. Aus dem bisher gezeigten geht sogar hervor, dass die Maximumstelle durch $(\frac{1}{n}, \dots, \frac{1}{n})$ eindeutig bestimmt ist. \square

4. Entropiebegriff der statistischen Thermodynamik

4.1. Boltzmann-Entropie

Ludwig Boltzmann war der Erste, der eine wahrscheinlichkeitstheoretische Interpretation der thermodynamischen Entropie gab und die berühmte Formel

$$S = k \log W$$

prägte. Dabei bezeichnet S die (Boltzmann-) Entropie eines beobachteten makroskopischen Zustands, W die Anzahl der mikroskopischen Zustände, die zum beobachteten Makrozustand führen und k die *Boltzmannkonstante*, die zwar physikalische Relevanz besitzt, für mathematische Betrachtungen aber ignoriert werden kann [Geo03, S. 37].

Sei $\Omega = \{\alpha_1, \dots, \alpha_n\}$ und $q = (q_1, \dots, q_n)$ ein Wahrscheinlichkeitsvektor, der $N \cdot q_i \in \mathbb{Z} \forall i \in \{1, \dots, n\}$ erfüllt. Im Kontext zur Boltzmann Entropie werden N nicht unterscheidbare Teilchen untersucht, die auf n unterschiedliche Energieniveaus oder Energiezustände verteilt werden [Geo09, S.34-35]. Dabei wird Ω als die Menge der möglichen Energieniveaus für ein solches Teilchensystem interpretiert und q entspricht hierbei einer relativen Häufigkeitsverteilung der Teilchen auf diese Energieniveaus. Durch jedes q wird ein Makrozustand beschrieben. Wie in Kapitel 3.3 wird erneut angenommen, dass die Ununterscheidbarkeit der Teilchen nur an den mangelhaften experimentellen Möglichkeiten liegt, die Teilchen aber „in Wirklichkeit“ mit $1, \dots, N$ durchnummeriert werden können [Geo09, S. 35]. Der Vektor $(\omega_1, \dots, \omega_N) \in \Omega^N$ wird als ein Mikrozustand interpretiert [Geo03, S. 37], wobei ω_j als der Energiezustand des Teilchens mit der Nummer $j \in \{1, \dots, N\}$ ist.

Wie in Abbildung 4.1 dargestellt, können mehrere Mikrozustände zur selben relativen Häufigkeitsverteilung dieser N Teilchen auf die n Energiezustände führen. Da die N Teilchen für den Betrachtenden nicht unterscheidbar sind, kann nur die relative Häufigkeitsverteilung festgestellt werden. Können mehrere Mikrozustände zu dieser Verteilung führen, herrscht somit eine Ungewissheit über den Mikrozustand des Teilchensystems. Für ein Teilchensystem aus $N \in \mathbb{N}$ ununterscheidbaren Teilchen und einer

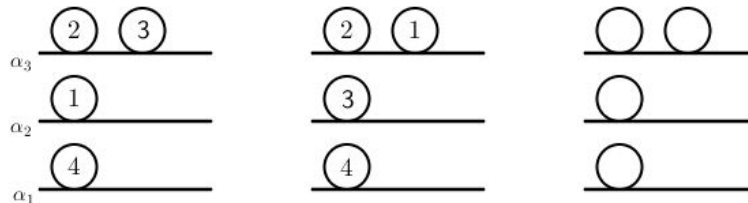


Abbildung 4.1.: Verteilung von $N = 4$ Teilchen auf $n = 3$ Energieniveaus bei einer Verteilung $q = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. Die vier Teilchen sind durchnummeriert.

Links: Einer der möglichen Mikrozustände.

Mitte: Die Teilchen mit der Nr. 1 und Nr. 3 sind vertauscht, dennoch ist die Verteilung dieselbe.

Rechts: Für den/die BeobachterIn des Systems, der/die die Nummerierung nicht sehen kann, sehen beide Mikrozustände gleich aus. Daher rührt die Ungewissheit über den Mikrozustand bei gegebener Verteilung.

Quelle: Eigene Darstellung.

relativen Häufigkeitsverteilung $q = (q_1, \dots, q_n)$ ist die Anzahl $W_N(q)$ der Möglichkeiten, N Teilchen so auf die n Energieniveaus zu verteilen, dass das k -te Energieniveau genau $q_k \cdot N$ Teilchen enthält, gleich

$$W_N(q) := \binom{N}{q_1 \cdot N, q_2 \cdot N, \dots, q_n \cdot N} = \frac{N!}{(q_1 \cdot N)! \cdot (q_2 \cdot N)! \cdot \dots \cdot (q_n \cdot N)!}$$

(für die Definition von $W_N(q) := \binom{N}{q_1 \cdot N, q_2 \cdot N, \dots, q_n \cdot N}$ vergleiche [ES14, S. 90]).

Boltzmanns Vorstellung über seinen Entropiebegriff war Folgende [Geo03, S.38]:

Die Entropie eines, durch eine Verteilung q der Teilchen auf die Energieniveaus hervorgerufenen, Makrozustands entspricht dem Grad an Ungewissheit

über den Mikrozustand $(\omega_1, \dots, \omega_N) \in \Omega^N$, wenn nur q bekannt ist und wird gemessen als $\log W_N(q)$, dem Logarithmus der Anzahl an Mikrozuständen, die zu q führen.

4.2. Herleitung von Boltzmann's Entropieformel

Die Begründung, warum $S = k \log W$ als Maß für die Ungewissheit über den Mikrozustand eines Systems bei gegebener Verteilung gewählt wird, erfolgt in diesem Kapitel. In [EM81, S. 51-52] werden der Informationsgehalt des Eintreffens eines Ereignisses und die Ungewissheit über das Eintreffen des Ereignisses vor der Versuchsdurchführung gleichgesetzt und mit der Wahrscheinlichkeit des Ereignisses in Verbindung gebracht. Die Eigenschaften der Ungewissheit eines Ereignisses werden für die Beschreibung der Ungewissheit über den Mikrozustand eines Teilchensystems übernommen, wobei hier die Eigenschaften in Verbindung mit der Anzahl W an Mikrozuständen zu einer gegebenen Verteilung q gebracht werden.

Führen mehr Mikrozustände zu einer Verteilung q_1 als zu einer Verteilung q_2 , so soll die Ungewissheit über den Mikrozustand der Verteilung q_1 größer sein als die Ungewissheit über die Verteilung q_2 :

$$\text{Für } W(q_1) > W(q_2) \text{ soll } S(W(q_1)) > S(W(q_2)) \text{ gelten.}$$

Betrachte ein ideales Gas in einem kubischen Behälter, der wieder wie in Kapitel 3.4 in endlich viele gleich große Teilvolumina unterteilt ist. Die Mikrozustände bezüglich des Ortes sind unabhängig von den Mikrozuständen bezüglich des Impulses, denn an jedem Ort im Ortsraum kann ein ideales Gasteilchen jeden zulässigen Impuls im Impulsraum besitzen [Sti10, S. 64], wobei auch hier der Impulsraum in endlich viele gleich große Teilvolumina unterteilt ist (der Impulsraum ist beschränkt, da ein Teilchen maximal die Lichtgeschwindigkeit im Vakuum erreichen kann und im idealen Gasmodell sogar vorausgesetzt wird, dass die Geschwindigkeit eines Gasteilchens weitaus kleiner als die Lichtgeschwindigkeit ist [Sti10, S. 29]). Somit ist die gesamte Anzahl an Mikrozuständen bezüglich der Orts- und Impulsverteilung, W_{Gesamt} , gleich dem Produkt aus der Anzahl an Mikrozuständen bezüglich der Ortsverteilung, $W(q_{\text{Ort}})$, und der Anzahl an Mikrozuständen bezüglich der Impulsverteilung, $W(q_{\text{Impuls}})$. Da der Ortszustand

und der Impulszustand unabhängig voneinander sind, soll die Ungewissheit über den Mikrozustand bezüglich Ort *und* Impuls gleich der Summe der Ungewissheit bezüglich des Orts und der Ungewissheit bezüglich des Impulses sein:

$$\begin{aligned} \text{Für } W_{\text{Gesamt}} = W(q_{\text{Ort}}) \cdot W(q_{\text{Impuls}}) \text{ soll} \\ S(W_{\text{Gesamt}}) = S(W(q_{\text{Ort}}) \cdot W(q_{\text{Impuls}})) = S(W(q_{\text{Ort}})) + S(W(q_{\text{Impuls}})) \text{ gelten.} \end{aligned}$$

Gibt es nur einen Mikrozustand zu einer gegebenen Verteilung, so gibt es keine Ungewissheit über den Mikrozustand. Die Ungewissheitsfunktion soll für $W = 1$ also 0 sein:

$$\text{Es soll } S(1) = 0 \text{ gelten.}$$

Analog wie in Kapitel 3.1. führt dies zu folgender Definition:

Definition 4.2.1. Ein Maß für die Ungewissheit über den Mikrozustand bei gegebener Verteilung ist eine Funktion $S : [0, +\infty) \rightarrow [0, +\infty)$ mit

- (i) S ist streng monoton steigend: Für $W_1 > W_2 \in [0, +\infty)$ folgt $S(W_1) > S(W_2)$.
- (ii) Für alle $W_1, W_2 \in [0, +\infty)$: $S(W_1 \cdot W_2) = S(W_1) + S(W_2)$.
- (iii) $S(1) = 0$.

Ebenfalls analog wie in Kapitel 3.1. lässt sich zeigen, dass sich S eindeutig bis auf eine multiplikative Konstante $a > 0$ als $S = a \cdot \log W$ darstellen lässt. In der statistischen Thermodynamik wird a mit der Einheit $\frac{J}{K}$ belegt, wobei $a = k := 1,380\,649 \cdot 10^{-23} \frac{J}{K}$ [Nat] gesetzt und k als *Boltzmann-Konstante* bezeichnet wird. Demnach lässt sich die *Boltzmann-Entropie* schreiben als

$$S = k \cdot \log W.$$

4.3. Maximale Boltzmann-Entropie

Eine Gemeinsamkeit der Entropiebegriffe nach Boltzmann und Shannon sind ihre Maximumeigenschaften. Nach Satz 3.5.2. ist $h(p)$ genau dann maximal, wenn die Wahrscheinlichkeitsverteilung p gleich $(\frac{1}{n}, \dots, \frac{1}{n})$ ist. Ähnliches gilt für die Boltzmann-Entropie bezüglich der relativen Häufigkeitsverteilung $(\frac{m_1}{N}, \dots, \frac{m_n}{N})$ mit $(m_1, \dots, m_n) \in$

4.3. MAXIMALE BOLTZMANN-ENTROPIE

\mathbb{N}_0^n und $m_1 + \dots + m_n = N$. Dies geht aus folgender Proposition [CK92, S. 101-102] hervor:

Proposition 4.3.1. *Seien $n \leq N \in \mathbb{N}$ mit $N = nq + r$, wobei $q \in \mathbb{N}_0$ und $0 \leq r \leq n - 1$, $r \in \mathbb{N}_0$ und sei*

$$M(N, n) := \max \left\{ \binom{N}{m_1, \dots, m_n} : m_i \in \mathbb{N}_0, \sum_{i=1}^n m_i = N \right\}.$$

Dann gilt

$$M(N, n) = \binom{N}{\underbrace{q, \dots, q}_{n-r \text{ mal}}, \underbrace{q+1, \dots, q+1}_{r \text{ mal}}} = \frac{N!}{(q!)^{n-r} ((q+1)!)^r} = \frac{N!}{(q+1)^r (q!)^n}.$$

Dabei ist $\binom{N}{m_1, \dots, m_n}$ genau dann maximal, wenn für $n - r$ der m_i gilt, dass $m_i = q$ und für r der m_i gilt, dass $m_i = q + 1$ ist.

Da $S = k \log W$ streng monoton steigend ist, folgt aus obiger Proposition unmittelbar nachfolgender Satz:

Satz 4.3.2. *Seien $n \leq N \in \mathbb{N}$ mit $N = nq + r$, wobei $q \in \mathbb{N}_0$ und $0 \leq r \leq n - 1$, $r \in \mathbb{N}_0$. Dann gilt*

$$S\left(\binom{N}{m_1, \dots, m_n}\right) \leq S\left(\binom{N}{\underbrace{q, \dots, q}_{n-r \text{ mal}}, \underbrace{q+1, \dots, q+1}_{r \text{ mal}}}\right),$$

wobei $S\left(\binom{N}{m_1, \dots, m_n}\right)$ genau dann maximal ist, wenn für $n - r$ der m_i gilt, dass $m_i = q$ und für r der m_i gilt, dass $m_i = q + 1$ ist.

Gibt es ein $q > 0 \in \mathbb{N}$, so dass $N = q \cdot n$, so ist $q = \frac{N}{n}$. Nach obigem Satz gilt daher

$$\begin{aligned} S\left(\binom{N}{m_1, \dots, m_n}\right) &\leq S\left(\binom{N}{\underbrace{q, \dots, q}_{n-r \text{ mal}}, \underbrace{q+1, \dots, q+1}_{r \text{ mal}}}\right) \\ &= S\left(\binom{N}{\frac{N}{n}, \dots, \frac{N}{n}}\right). \end{aligned}$$

Aus $\frac{N}{n} \frac{1}{N} = \frac{1}{n}$ und obigem Satz folgt, dass die Boltzmann-Entropie genau dann maximal ist, wenn die relative Häufigkeitsverteilung gleich $(\frac{1}{n}, \dots, \frac{1}{n})$ ist.

4.4. Zusammenhang mit der Shannon-Entropie

Im Folgenden wird versucht einen Zusammenhang zwischen Boltzmanns und Shannons Entropiebegriffen zu ermitteln. Dazu folgende Überlegung: Ein System aus (abzählbar) unendlich vielen, bereits durchnummerierten Teilchen, die unabhängig und ununterscheidbar sind und endlich viele Energiezustände aus einer Menge $\Omega := \{\alpha_1, \dots, \alpha_n\}$ annehmen können, wird untersucht. Die Ununterscheidbarkeit wird so interpretiert, dass sich jedes Teilchen auch stochastisch nicht unterscheiden lässt: Jedes Teilchen wird durch denselben Wahrscheinlichkeitsraum bestehend aus der Menge Ω mit einem Wahrscheinlichkeitsvektor $p = (p_1, \dots, p_n)$ auf Ω beschrieben. p_k wird dabei als die Wahrscheinlichkeit interpretiert, dass ein Teilchen im Energiezustand α_k ist.

Dass unendlich viele unabhängige und nicht unterscheidbare Teilchen gleichzeitig betrachtet werden, lässt sich so interpretieren, dass ein Zufallsgeschehen unendlich oft unabhängig nebeneinander ausgeführt wird. Daher beschreibt der Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ das oben beschriebene unendliche Teilchensystem, wobei $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ wie in Kapitel 2.1. definiert ist. Die Folge $(\omega_1, \omega_2, \dots) \in \tilde{\Omega}$ wird dabei so interpretiert, dass ω_i der Energiezustand des Teilchens mit der Nummer $i \in \mathbb{N}$ ist.

Betrachte außerdem für $k \in \{1, 2, \dots, n\}$ und $N \in \mathbb{N}$ die Zufallsvariablen $X^{(k)} : \Omega \rightarrow \{0, 1\}$ mit

$$X^{(k)}(\omega) := \begin{cases} 1 & \text{für } \omega = \alpha_k, \\ 0 & \text{sonst} \end{cases}$$

und $\tilde{X}_j^{(k)} : \tilde{\Omega} \rightarrow \{0, 1\}$ mit $\tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots) := X^{(k)}(\omega_j)$. Die Zufallsvariable $\tilde{R}_N^{(k)} : \tilde{\Omega} \rightarrow \mathbb{R}$ definiert durch

$$\tilde{R}_N^{(k)}(\omega_1, \omega_2, \dots) := \frac{1}{N} \sum_{j=1}^N \tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots)$$

ist in dieser Interpretation die relative Häufigkeit des k -ten Energieniveaus unter den ersten N Teilchen. Es gelten folgende Propositionen:

Proposition 4.4.1. Seien $n, N \in \mathbb{N}$ und $k \in \{1, 2, \dots, n\}$. Sei $\tilde{R}_N^{(k)} : \tilde{\Omega} \rightarrow \mathbb{R}$ mit

$$\tilde{R}_N^{(k)}(\omega_1, \omega_2, \dots) := \frac{1}{N} \sum_{j=1}^N \tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots).$$

Es gilt: $\forall (\omega_1, \omega_2, \dots) \in \tilde{\Omega}$ ist $\tilde{R}_N(\omega_1, \omega_2, \dots) := (\tilde{R}_N^{(1)}(\omega_1, \omega_2, \dots), \dots, \tilde{R}_N^{(n)}(\omega_1, \omega_2, \dots))$ ein Wahrscheinlichkeitsvektor, der $N \cdot \tilde{R}_N^{(k)}(\omega_1, \omega_2, \dots) \in \mathbb{Z}$ erfüllt.

Beweis. Sei $\tilde{\omega} = (\omega_1, \omega_2, \dots) \in \tilde{\Omega}$ beliebig.

Es gelten $\tilde{R}_N^{(k)}(\omega_1, \omega_2, \dots) = \frac{1}{N} \sum_{j=1}^N \underbrace{\tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots)}_{\geq 0} \geq 0$ und

$$\begin{aligned} \sum_{k=1}^n \tilde{R}_N^{(k)}(\omega_1, \omega_2, \dots) &= \sum_{k=1}^n \left(\frac{1}{N} \sum_{j=1}^N \tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots) \right) = \frac{1}{N} \sum_{k=1}^n \sum_{j=1}^N \tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots) \\ &= \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^n \underbrace{\tilde{X}_j^{(k)}(\omega_1, \omega_2, \dots)}_{=\delta_{kj}} = \frac{1}{N} \sum_{j=1}^N 1 = \frac{1}{N} \cdot N = 1. \end{aligned}$$

Darüber hinaus ist $N \cdot \tilde{R}_N^{(k)}(\tilde{\omega}) = N \cdot \frac{1}{N} \sum_{j=1}^N \tilde{X}_j^{(k)}(\tilde{\omega}) = \sum_{j=1}^N \underbrace{\tilde{X}_j^{(k)}(\tilde{\omega})}_{\in \{0,1\}} \in \mathbb{Z}$. □

Proposition 4.4.2. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt: Für $n \in \mathbb{N}$ und $A_1, A_2, \dots, A_n \in \mathcal{A}$ mit $P(A_i) = 1 \forall i \in \{1, 2, \dots, n\}$ ist $P(\bigcap_{i=1}^n A_i) = 1$.

Beweis. Es wird ein Beweis durch Induktion nach n durchgeführt.

Für $n = 1$ ist $\bigcap_{i=1}^1 A_i = A_1$, also $P(\bigcap_{i=1}^1 A_i) = P(A_1) = 1$.

Sei nun $n > 1$. Wegen Proposition 2.1.4. (i) gilt

$$P\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cup A_n\right) = P\left(\bigcap_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\underbrace{\left(\bigcap_{i=1}^{n-1} A_i\right) \cap A_n}_{=\bigcap_{i=1}^n A_i}\right).$$

Aus $A_n \subseteq (\bigcap_{i=1}^{n-1} A_i) \cup A_n$ folgt mit Proposition 2.1.4. (ii):

$$\begin{aligned} 1 = P(A_n) &\leq P\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cup A_n\right) \leq 1 \\ &\Rightarrow P\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cup A_n\right) = 1. \end{aligned}$$

Daraus erhält man mit der Induktionsvoraussetzung

$$P\left(\bigcap_{i=1}^n A_i\right) = \underbrace{P\left(\bigcap_{i=1}^{n-1} A_i\right)}_{\stackrel{\text{IV}}{=}1} + \underbrace{P(A_n)}_{=1} - \underbrace{P\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cup A_n\right)}_{=1} = 1 + 1 - 1 = 1.$$

wodurch die Aussage bewiesen ist. □

Um die N -Abhängigkeit von $\log W_N(q)$ zu umgehen, kann für eine Verteilung p (hierfür muss p nicht $p_k \cdot N \in \mathbb{Z}$ erfüllen) der Grenzwert der Ungewissheit pro Teilchen für $N \rightarrow \infty$, also der Ausdruck

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N \left(\tilde{R}_N(\omega_1, \omega_2, \dots) \right)$$

betrachtet werden. Dazu folgender Satz [Geo03, S. 38]:

Satz 4.4.3. Seien $p = (p_1, \dots, p_n)$ und $q_N = (q_N^{(1)}, \dots, q_N^{(n)})$ Verteilungen auf $\Omega = \{\alpha_1, \dots, \alpha_n\}$, so dass $q_N \rightarrow p$ und $q_N^{(k)} \cdot N \in \mathbb{Z}$ für alle $k \in \{1, \dots, n\}$. Dann existiert der Limes

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N(q_N)$$

und es gilt

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N(q_N) = h(p) = - \sum_{k=1}^n p_k \log p_k.$$

Beweis. Dieser Satz lässt sich mit Lemma 2.3 aus [CK81, S. 30] beweisen [Geo03, S. 38]. Dieses Lemma besagt, dass für alle $N \in \mathbb{N}$ gilt, dass

$$(N + 1)^{-n} e^{N \cdot h(q_N)} \leq W_N(q_N) \leq e^{N \cdot h(q_N)}$$

erfüllt ist. Dies ist gleichbedeutend mit

$$\underbrace{\log \left((N + 1)^{-n} e^{N \cdot h(q_N)} \right)}_{=-n \log(N+1) + N \cdot h(q_N)} \leq \log W_N(q_N) \leq N \cdot h(q_N).$$

Durch Division durch N erhält man weiter

$$-n \cdot \frac{\log(N + 1)}{N} + h(q_N) \leq \frac{1}{N} \log W_N(q_N) \leq h(q_N).$$

Durch Anwendung der Regel von de l'Hospital erhält man

$$\lim_{x \rightarrow \infty} \frac{\log(x + 1)}{x} = \lim_{x \rightarrow \infty} \frac{1}{x + 1} = 0.$$

Daher gilt insbesondere $\lim_{N \rightarrow \infty} \frac{\log(N+1)}{N} = 0$. Da außerdem h stetig ist, gilt

$$\lim_{N \rightarrow \infty} h(q_N) = h(p).$$

Daraus folgt weiter, dass

$$\underbrace{-n \cdot \underbrace{\frac{\log(N + 1)}{N}}_{\rightarrow 0} + \underbrace{h(q_N)}_{\rightarrow h(p)}}_{\rightarrow h(p)} \leq \frac{1}{N} \log W_N(q_N) \leq \underbrace{h(q_N)}_{\rightarrow h(p)}$$

und daher folgt mit dem Einschluss-Satz [Kal15, S. 80, Satz 3.3.2], dass

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N(q_N) = h(p).$$

□

Nach Korollar 2.1.15.2. gilt für $\Omega = \{\alpha_1, \dots, \alpha_n\}$, $p = (p_1, \dots, p_n)$ und $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, dass für alle $k \in \{1, \dots, n\}$ ein $\tilde{A}_k \in \tilde{\mathcal{A}}$ mit $\tilde{P}(\tilde{A}_k) = 1$ existiert, so dass $\forall \tilde{\omega} \in \tilde{A}_k$: $\lim_{N \rightarrow \infty} \tilde{R}_N^{(k)}(\tilde{\omega}) = p_k$, wobei wieder $\tilde{R}_N^{(k)}(\tilde{\omega}) = \frac{1}{N} \sum_{j=1}^N \tilde{X}_j^{(k)}(\tilde{\omega})$ ist. Aufgrund von Proposition 4.4.2. gilt $\tilde{P}(\bigcap_{k=1}^n \tilde{A}_k) = 1$. Nach Proposition 4.4.1. gilt für beliebiges $\tilde{\omega} \in \tilde{\Omega}$ und somit auch für beliebiges $\tilde{\omega} \in \bigcap_{k=1}^n \tilde{A}_k$, dass $\tilde{R}_N(\tilde{\omega}) = \left(\tilde{R}_N^{(1)}(\tilde{\omega}), \dots, \tilde{R}_N^{(n)}(\tilde{\omega}) \right)$ eine Wahrscheinlichkeitsverteilung auf Ω ist mit $N \cdot \tilde{R}_N^{(k)}(\tilde{\omega}) \in \mathbb{Z}$. Für $\tilde{\omega} \in \bigcap_{k=1}^n \tilde{A}_k$ gilt $\tilde{R}_N(\tilde{\omega}) \rightarrow p$ für $N \rightarrow \infty$.

Mit Satz 4.4.3. erhält man somit

Korollar 4.4.3.1. *Es gilt*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N(\tilde{R}_N) = h(p) = - \sum_{k=1}^n p_k \log p_k \quad \tilde{P} - \text{fast sicher.}$$

Wird obige Gleichung mit der Boltzmann-Konstante k multipliziert, so folgt weiter

$$\lim_{N \rightarrow \infty} \frac{1}{N} S \left(W_N(\tilde{R}_N) \right) = k \cdot h(p) \quad \tilde{P} - \text{fast sicher.}$$

Anschaulich bedeutet das folgendes: Wird ein unendliches Teilchensystem, welches die eingangs dieses Kapitels beschriebenen Eigenschaften erfüllt, betrachtet, so konvergiert die Boltzmann-Entropie pro Teilchen „fast sicher“ gegen die Shannon-Entropie eines Teilchens eines solchen Teilchensystems, wenn diese mit der Boltzmann-Konstante k multipliziert wird.

Dies ähnelt folgendem Resultat bezüglich der Shannon-Entropie: Die Shannon-Entropie pro Teilchen eines Teilchensystems aus N unabhängigen und nicht unterscheidbaren Teilchen ist gleich der Shannon-Entropie eines einzelnen Teilchens. Daher konvergiert die Shannon-Entropie pro Teilchen für $N \rightarrow \infty$ gegen $h(p)$.

In der Praxis gibt es natürlich keine Teilchensysteme mit unendlich vielen Teilchen. Boltzmann wusste schon 1877, dass in der statistischen Thermodynamik betrachtete Teilchensysteme immer sehr groß sind (eine Stoffmenge von einem Mol entspricht etwa 10^{23} Teilchen) [Bol09, S. 185].

Betrachte als Gedankenexperiment die räumliche Verteilung eines idealen Gases im Gleichgewichtszustand in einem in $n \in \mathbb{N}$ gleichgroßen Teilwürfel aufgeteilten kubischen Raum zu einem festen Zeitpunkt. Stellt man sich diese räumliche Verteilung analog wie in [Mül03, S. 25-28] so vor, dass diese nicht durch die unregelmäßige Molekularbewegung der Gasteilchen, sondern durch eine gleichzeitige und unabhängige Verteilung der Gasteilchen gemäß der diskreten Gleichverteilung $p = (\frac{1}{n}, \dots, \frac{1}{n})$ auf diese Teilwürfel zustande kam und deutet die große Teilchenzahl so, dass sich dieses System näherungsweise wie ein unendlich großes System verhält, so gilt in diesem Gedankenexperiment für die Boltzmann-Entropie pro Teilchen „fast sicher“ näherungsweise $\frac{1}{N}S(W) \approx k \log n$.

5. Entropiebegriffe der dynamischen Systeme

5.1. Topologische Entropie

In Kapitel 2.2 wurde einführend erwähnt, dass in der Theorie dynamischer Systeme „Dynamiken“ in Abhängigkeit von der Zeit betrachtet werden. Wie in [Rai09, S. 1-6] gezeigt, gibt es deterministische Systeme, also vorhersagbare Systeme, sofern der Anfangszustand exakt bekannt ist, die sich dennoch „chaotisch“ verhalten. So können im Beispiel des logistischen Wachstums kleine Änderungen des Anfangszustands nach einiger Zeit große Auswirkungen haben [Rai09, S. 6].

In der Literatur gibt es verschiedene nicht äquivalente Definitionen von Chaos [Rai09, S. 1]. Die in diesem Unterkapitel behandelte *topologische Entropie* kann als Maß dafür aufgefasst werden, wie chaotisch ein dynamisches System ist [Rai09, S. 6].

Im Folgenden wird die Definition der topologischen Entropie ausgehend von folgender Idee abgeleitet: Die Chaozität eines Systems soll daran gemessen werden, wie „rasant“ sich ein System mit der Zeit entwickelt. Dementsprechend wird die topologische Entropie mit der exponentiellen Wachstumsrate der Anzahl der Orbits der Länge s in Verbindung gebracht [Vri14, S. 378]. Bezeichnet k_s die Anzahl der Orbits der Länge s , so führt die Ausführung der Idee zu folgender Äquivalenzkette:

$$\begin{aligned} k_s = c \cdot e^{h \cdot s} &\Leftrightarrow \log k_s = \log c + h \cdot s \\ &\Leftrightarrow \frac{1}{s} \log k_s = \frac{1}{s} \log c + h \Leftrightarrow h = \frac{1}{s} \log k_s - \frac{1}{s} \log c. \end{aligned}$$

Um die Abhängigkeit von s zu umgehen, wird der Grenzwert von s gegen unendlich betrachtet: $h = \lim_{s \rightarrow \infty} \frac{1}{s} \log k_s$. Dieser Ausdruck weist bereits eine erste Diskrepanz auf: Die Anzahl der Orbits der Länge s ist üblicherweise unendlich [You03, S. 314],

im Beispiel des logistischen Wachstums sogar überabzählbar unendlich. Daher wird die Anzahl der Orbits „in niedrigerer Auflösung“ betrachtet: Zwei Orbits der Länge s werden genau dann noch als unterschiedlich betrachtet, wenn sich die Zustände der Orbits durch Beobachtung mit einer „ ε -Lupe“ an mindestens einem Zeitpunkt zwischen 0 und $s - 1$ unterscheiden lassen, die Zustände der Orbits also im Laufe von 0 bis $s - 1$ mindestens einmal einen größeren Abstand als ein $\varepsilon > 0$ voneinander haben. Dies lässt sich wie folgt formalisieren [You03, S. 314]:

Definition 5.1.1. Sei $T : X \rightarrow X$ eine stetige Abbildung auf einem kompakten metrischen Raum X . Für $\varepsilon > 0$ und $s \in \mathbb{N}_0$ heißt $E \subseteq X$ eine (s, ε) -trennende Menge oder kurz (s, ε) -trennend, falls für alle $x \neq y \in E$ ein $j \in \{0, 1, \dots, s - 1\}$ existiert, so dass $d(T^j x, T^j y) > \varepsilon$ ist. Setze $k_s(\varepsilon) := \sup_{E \text{ } (s, \varepsilon)\text{-trennend}} |E|$.

$k_s(\varepsilon) := \sup_{E \text{ } (s, \varepsilon)\text{-trennend}} |E|$ lässt sich als die größte Anzahl an Orbits der Länge s interpretieren, die sich bei einer Auflösung von ε unterscheiden lassen [Vri14, S. 381]. Nachfolgend wird begründet, dass $k_s(\varepsilon)$ für alle $s \in \mathbb{N}_0$ und alle $\varepsilon > 0$ eine natürliche Zahl ist.

Proposition 5.1.2. Die Abbildung $d_s^T : X \rightarrow X$ definiert durch

$$d_s^T(x, y) := \max_{0 \leq i \leq s-1} d(T^i x, T^i y) \text{ für } x, y \in X.$$

ist eine Metrik.

Beweis. (i) Da d eine Metrik ist, ist aus der Definition von d_s^T ersichtlich, dass

$$d_s^T(x, y) \geq 0 \text{ und } d_s^T(x, x) = 0 \text{ erfüllt sind.}$$

$$\text{Sei } d_s^T(x, y) = 0 \Rightarrow \forall i \in \{0, 1, \dots, s - 1\} : d(T^i x, T^i y) = 0 \Rightarrow d(x, y) = 0 \Rightarrow x = y.$$

(ii) Die zweite Eigenschaft ist erfüllt, da d eine Metrik und damit alle $d(T^i x, T^i y)$ symmetrisch sind.

(iii) Seien $x, y, z \in X$. Da ein $j \in \{0, 1, \dots, s - 1\}$ existiert mit

$$d_s^T(x, z) = \max_{0 \leq i \leq s-1} d(T^i x, T^i z) = d(T^j x, T^j z),$$

folgt

$$d_s^T(x, z) = d(T^j x, T^j y) \leq \underbrace{d(T^j x, T^j y)}_{\leq d_s^T(x, y)} + \underbrace{d(T^j y, T^j z)}_{\leq d_s^T(y, z)} \leq d_s^T(x, y) + d_s^T(y, z).$$

□

Proposition 5.1.3. *Sei (X, T) ein topologisches dynamisches System. Eine Menge $E \subseteq X$ ist genau dann (s, ε) -trennend, wenn*

$$\forall x, y \in E : x \neq y \Rightarrow d_s^T(x, y) > \varepsilon.$$

Beweis. $(\Rightarrow) \forall x \neq y \in E : \exists j \in \{0, 1, \dots, s-1\}$, so dass $d(T^j x, T^j y) > \varepsilon$. Da $d_s^T(x, y) = \max_{0 \leq i \leq s-1} d(T^i x, T^i y) \geq d(T^j x, T^j y)$, folgt $d_s^T(x, y) > \varepsilon$.

$(\Leftarrow) \forall x \neq y \in E : d_s^T(x, y) = \max_{0 \leq i \leq s-1} d(T^i x, T^i y) > \varepsilon$. Da es ein $j \in \{0, 1, \dots, s-1\}$ mit $d_s^T(x, y) = d(T^j x, T^j y)$ gibt, gilt für dieses $j \in \{0, 1, \dots, s-1\}$, dass $d(T^j x, T^j y) > \varepsilon$. □

Für die nachfolgende Begründung, dass $k_s(\varepsilon) \in \mathbb{N}$ ist, vergleiche [Vri14, S. 381]: In [Vri14, S. 380] wird gezeigt, dass d_s^T die gleiche Topologie auf X erzeugt wie d . Daher gilt insbesondere, dass X auch bezüglich $d_s^T \forall s \in \mathbb{N}$ kompakt ist. Deshalb gibt es zu der offenen Überdeckung $\bigcup_{x \in X} B_{d_s^T}(x, \frac{\varepsilon}{2})$ von X eine endliche Teilüberdeckung $\bigcup_{k=1}^m B_{d_s^T}(x_k, \frac{\varepsilon}{2})$. Ist $E \subseteq X$ (s, ε) -trennend, so gilt: Sei $x \in E$ beliebig. Wegen $E \subseteq X \subseteq \bigcup_{k=1}^m B_{d_s^T}(x_k, \frac{\varepsilon}{2})$ gibt es ein $i \in \{1, \dots, m\}$, so dass $x \in B_{d_s^T}(x_i, \frac{\varepsilon}{2}) = \{z \in X : d_s^T(z, x_i) < \frac{\varepsilon}{2}\}$ ist. Betrachte nun ein $y \neq x \in E$. Da E (s, ε) -trennend ist, gilt wegen Proposition 5.1.3., dass $d_s^T(y, x) > \varepsilon$. Aus Proposition 2.2.2. folgt daher, dass

$$\begin{aligned} d_s^T(y, x_i) &\geq |d_s^T(y, x) - d_s^T(x, x_i)| = \underbrace{d_s^T(y, x)}_{> \varepsilon} - d_s^T(x, x_i) \\ &> \varepsilon - \underbrace{d_s^T(x, x_i)}_{< \frac{\varepsilon}{2}} > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}. \end{aligned}$$

Also ist $y \notin B_{d_s^T}(x_i, \frac{\varepsilon}{2})$. Da aber $\bigcup_{k=1}^m B_{d_s^T}(x_k, \frac{\varepsilon}{2})$ die Menge E überdeckt, muss y in einer der anderen Kugeln der endlichen Überdeckung sein. Somit gilt: jedes Element in E muss in einer der endlich vielen Kugeln sein, aber in jeder Kugel ist höchstens ein Element. Daher hat E höchstens m viele Elemente und ist damit endlich. Da für ein $x \in X$

die Menge $\{x\}$ (s, ε) -trennend ist, ist daher $\{|E| : E \text{ ist } (s, \varepsilon)\text{-trennend}\} \subseteq \{1, \dots, m\}$. Als endliche Teilmenge von \mathbb{N} hat daher $\{|E| : E \text{ ist } (s, \varepsilon)\text{-trennend}\}$ ein Maximum [Kal15, S. 32, Lemma 2.3.17]. Somit ist $k_s(\varepsilon) = \max\{|E| : E \text{ ist } (s, \varepsilon)\text{-trennend}\} \in \mathbb{N}$.

Wird nun in der vorläufigen Definition $h = \lim_{s \rightarrow \infty} \frac{1}{s} \log k_s$ der Ausdruck k_s durch $k_s(\varepsilon)$ für ein $\varepsilon > 0$, ersetzt, also $h = \lim_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon)$ betrachtet, so treten dennoch zwei Probleme auf:

1. $\frac{1}{s} \log k_s(\varepsilon)$ muss nicht konvergieren.
2. $\frac{1}{s} \log k_s(\varepsilon)$ hängt von ε ab.

Diese Probleme lassen sich aber durch folgende Abänderungen in der Definition von h umgehen:

Zu 1.: Es gilt $\frac{1}{s} \log k_s(\varepsilon) \geq 0$, daher ist die Folge $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ nach unten beschränkt. Daher gilt: $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ ist entweder nach oben beschränkt und damit beschränkt, oder nach oben unbeschränkt. Ist $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ nach oben beschränkt und damit beschränkt, so existiert nach dem Satz von Bolzano-Weierstraß [For16, S. 55, Satz 6] eine konvergente Teilfolge von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$. Daher ist die Menge aller Häufungswerte von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ nicht leer und da $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ beschränkt ist, ist auch die Menge aller Häufungswerte beschränkt. Daher existiert das Supremum der Menge aller Häufungswerte, welches gleich dem Limes superior von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ ist, wobei $\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) \in [0, +\infty)$ gilt. Ist $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ nach oben unbeschränkt, so ist $\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) = +\infty$. Insgesamt existiert der Limes superior daher immer und es gilt $\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) \in [0, +\infty]$. Daher wird der Limes durch den Limes superior ersetzt.

Zu 2.: Es gilt folgende Proposition [Vri14, S. 381, 382]:

Proposition 5.1.4. *Für $0 < \varepsilon' < \varepsilon$ gelten*

- (i) $k_s(\varepsilon') \geq k_s(\varepsilon)$
- (ii) $\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon') \geq \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon)$.

Beweis. (i) Ist eine Menge E (s, ε) -trennend, so ist diese Menge auch (s, ε') -trennend [Vri14, S. 381], denn: Ist E (s, ε) -trennend, so gilt $\forall x, y \in E : d_s^T(x, y) > \varepsilon > \varepsilon'$, daher ist E auch (s, ε') -trennend. Daher ist $\{E \subseteq X : E \text{ ist } (s, \varepsilon)\text{-trennend}\} \subseteq \{E \subseteq X : E \text{ ist } (s, \varepsilon')\text{-trennend}\} \Rightarrow \{|E| : E \text{ ist } (s, \varepsilon)\text{-trennend}\} \subseteq \{|E| : E \text{ ist } (s, \varepsilon')\text{-trennend}\}$

$$\begin{aligned} \Rightarrow k_s(\varepsilon) &= \max\{|E| : E \text{ ist } (s, \varepsilon)\text{-trennend}\} \\ &\leq k_s(\varepsilon') = \max\{|E| : E \text{ ist } (s, \varepsilon')\text{-trennend}\} \end{aligned}$$

Für die letzte Implikation wurde Lemma 2.2.8 (ii) in [Kal15, S. 19] angewandt.

(ii) Es wird gezeigt, dass es zu jedem Häufungswert von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ einen Häufungswert b' von $(\frac{1}{s} \log k_s(\varepsilon'))_{s \in \mathbb{N}}$ gibt mit $b' \geq b$.

Sei daher $b \in [0, +\infty]$ ein Häufungswert von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ (wie oben bei der Behandlung von Problem 1. gezeigt, hat die Folge $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ immer einen Häufungswert in $[0, +\infty]$). Es existiert also eine Teilfolge $(\frac{1}{s_k} \log k_{s_k}(\varepsilon))_{k \in \mathbb{N}}$ von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ mit $(\frac{1}{s_u} \log k_{s_u}(\varepsilon))_{u \rightarrow \infty} \rightarrow b$ (eigentliche oder uneigentliche Konvergenz). Betrachte die Teilfolge $(\frac{1}{s_u} \log k_{s_u}(\varepsilon'))_{u \in \mathbb{N}}$ von $(\frac{1}{s} \log k_s(\varepsilon'))_{s \in \mathbb{N}}$. Auch die Folge $(\frac{1}{s_u} \log k_{s_u}(\varepsilon'))_{u \in \mathbb{N}}$ besitzt eine (eigentlich oder uneigentlich) konvergente Teilfolge $(\frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon'))_{r \in \mathbb{N}}$. Es gilt also $(\frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon'))_{r \rightarrow \infty} \rightarrow b'$. Da Teilfolgen von (eigentlich oder uneigentlich) konvergenten Folgen gegen denselben Grenzwert wie die Folge selbst konvergieren und $\frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon') \geq \frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon)$ erfüllt ist (wegen (i) und da der Logarithmus streng monoton steigend ist), folgt:

$$b' = \lim_{r \rightarrow \infty} \frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon') \geq \lim_{r \rightarrow \infty} \frac{1}{s_{u_r}} \log k_{s_{u_r}}(\varepsilon) = b.$$

Betrachte nun die Mengen der Häufungswerte von $(\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}$ bzw. $(\frac{1}{s} \log k_s(\varepsilon'))_{s \in \mathbb{N}}$, also die Mengen $B := \{b \in [0, +\infty] : b \text{ ist Häufungswert von } (\frac{1}{s} \log k_s(\varepsilon))_{s \in \mathbb{N}}\}$ bzw. $B' := \{b' \in [0, +\infty] : b' \text{ ist Häufungswert von } (\frac{1}{s} \log k_s(\varepsilon'))_{s \in \mathbb{N}}\}$ und für ein $b \in B$ die Menge $\mathfrak{B}_b := \{b' \in B' : b' \geq b\}$. Für die Familie von Mengen $(\mathfrak{B}_b)_{b \in B}$ gibt es nach dem Auswahlaxiom [Neu17, S. 10] eine Funktion $f : B \rightarrow \bigcup_{b \in B} \mathfrak{B}_b$, die für alle $b \in B$ erfüllt, dass $f(b) \in \mathfrak{B}_b$. Das heißt, dass für alle $b \in B$ gilt, dass $f(b)$ ein Häufungswert von $(\frac{1}{s} \log k_s(\varepsilon'))_{s \in \mathbb{N}}$ mit $f(b) \geq b$ ist. Sowohl B als auch $f(B)$ haben ein Supremum in $[0, +\infty]$.

Sei nun $b \in B$ beliebig. Es gilt $b \leq f(b) \leq \sup f(B)$, daher ist $\sup f(B)$ eine obere Schranke von $B \Rightarrow \sup B \leq \sup f(B)$. Da $f(B) \subseteq B'$, gilt $\sup f(B) \leq \sup B'$. Insgesamt gilt demnach

$$\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) = \sup B \leq \sup B' = \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon').$$

□

Das zweite Problem kann folgendermaßen gelöst werden: Die Folge

$$\left(\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s\left(\frac{1}{m}\right) \right)_{m \in \mathbb{N}}$$

ist monoton steigend. Daher hat diese Folge einen Grenzwert in $[0, +\infty]$. Dies ist gleichbedeutend mit $\lim_{\varepsilon \rightarrow 0^+} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) \in [0, +\infty]$, wobei

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) = \sup_{\varepsilon > 0} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon).$$

Daher wird der feste und daher von $\varepsilon > 0$ unabhängige Grenzwert

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon)$$

anstelle des von $\varepsilon > 0$ abhängigen Werts $\limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon)$ gewählt. Dies führt letztendlich zu folgender Definition [Bow71, S. 402-403]:

Definition 5.1.5. Sei (X, T) ein topologisches dynamisches System. Dann ist die *topologische Entropie* definiert durch

$$h_{top}(T) := \lim_{\varepsilon \rightarrow 0^+} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) = \sup_{\varepsilon > 0} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon).$$

Um zu kennzeichnen, welcher Zustandsraum betrachtet wird, wird auch $h_{top}(X, T)$ geschrieben.

5.2. Metrische Entropie

Ein weiterer Entropiebegriff der Theorie dynamischer Systeme ist die *metrische Entropie*. Hier wird Shannons Entropiebegriff in die Theorie dynamischer Systeme einbezogen.

Folgende Situation wird untersucht: Sei (X, T) ein topologisches dynamisches System, \mathcal{A} eine σ -Algebra auf X , P ein Wahrscheinlichkeitsmaß auf \mathcal{A} und $\mathcal{Z} = \{A_1, \dots, A_n\}$ eine endliche messbare Zerlegung von X . Dadurch wird die Theorie der dynamischen Systeme mit der Wahrscheinlichkeitstheorie verbunden: Es wird ein Zustandsraum betrachtet, bei dem ein Zustand bzw. ein Zustandsbereich mit einer gewissen Wahrscheinlichkeit eintreten kann. In Anbetracht der durch die Transformation T beschriebenen Dynamik des Systems und den bereits behandelten Kapiteln stellt sich die Frage nach einem Ausdruck, der die durchschnittliche Information angibt, die man erhält, wenn für jedes $x \in X$ verfolgt wird, in welchen $A \in \mathcal{Z}$ die Komponenten des Orbits $(x, Tx, \dots, T^{s-1}x)$ liegen [ES14, S. 96].

Dafür werden die folgende Definition und Proposition verwendet:

Definition 5.2.1. Sei (X, T) ein topologisches dynamisches System, \mathcal{A} eine σ -Algebra auf X und P ein Wahrscheinlichkeitsmaß auf \mathcal{A} . Die Transformation T wird *maßerhaltend* genannt, wenn für alle $A \in \mathcal{A}$ die Gleichung $P(A) = P(T^{-1}A)$ gilt [You03, S. 315]. (X, \mathcal{A}, P, T) wird dann ein *maßerhaltendes topologisches dynamisches System* oder kürzer *maßerhaltendes dynamisches System* genannt [ES14, S. 64].

Anmerkung. Da T stetig ist, ist T messbar, es gilt also $T^{-1}A \in \mathcal{A}$ für alle $A \in \mathcal{A}$. Insbesondere gilt damit auch $T^{-i}A \in \mathcal{A}$ für alle $A \in \mathcal{A}$ und alle $i \in \mathbb{N}_0$.

Für die Festlegungen in folgender Definition vergleiche [You03, S. 315]:

Proposition 5.2.2. Sei (X, \mathcal{A}, P, T) ein maßerhaltendes dynamisches System. Für zwei endliche Zerlegungen $\mathcal{Z} = \{A_1, \dots, A_k\}$, $\mathcal{Q} = \{B_1, \dots, B_m\}$ setze

$$\mathcal{Z} \vee \mathcal{Q} := \{A \cap B : A \in \mathcal{Z}, B \in \mathcal{Q}\} \setminus \emptyset.$$

und $T^{-i}\mathcal{Z} := \{T^{-i}A_1, \dots, T^{-i}A_k\}$. Dann gelten:

- (i) $\mathcal{Z} \vee \mathcal{Q}$ ist eine Zerlegung von X .
- (ii) $\bigvee_{i=0}^{s-1} T^{-i}\mathcal{Z}$ ist eine Zerlegung von X .

Beweis. Es werden die Eigenschaften einer Zerlegung überprüft:

(i) Seien $C \neq D \in \mathcal{Z} \vee \mathcal{Q}$ beliebig. Dann $\exists j, i \in \{1, \dots, k\}$, $u, r \in \{1, \dots, m\}$ mit $C = A_j \cap B_u$, $D = A_i \cap B_r$ und $j \neq i$ oder $u \neq r$. Daher ist

$$C \cap D = (A_j \cap B_u) \cap (A_i \cap B_r) = (A_j \cap A_i) \cap (B_u \cap B_r) = \emptyset.$$

Außerdem ist

$$\bigcup_{C \in \mathcal{Z} \vee \mathcal{Q}} C = \bigcup_{j=1}^k \bigcup_{i=1}^m (A_j \cap B_i) = \bigcup_{j=1}^k \left(\underbrace{\left(\bigcup_{i=1}^m A_j \right)}_{=A_j} \cap \underbrace{\left(\bigcup_{i=1}^m B_i \right)}_{=X} \right) = \bigcup_{j=1}^k \underbrace{(A_j \cap X)}_{=A_j} = X.$$

(ii) Für $j \neq i \in \{1, \dots, k\}$ gilt

$$\underbrace{T^{-1}(A_j) \cap T^{-1}(A_i)}_{=T^{-1}(A_j \cap A_i)} = T^{-1} \underbrace{(A_j \cap A_i)}_{=\emptyset} = T^{-1}\emptyset = \{x \in X : Tx \in \emptyset\} = \emptyset.$$

Des Weiteren gilt

$$\bigcup_{j=1}^k T^{-1}A_j = T^{-1} \left(\bigcup_{j=1}^k A_j \right) = T^{-1}X = X.$$

Daher ist $T^{-1}\mathcal{Z}$ eine Zerlegung von X . Daraus folgt, dass $T^{-i}\mathcal{Z}$ für alle $i \in \{0, \dots, s-1\}$ eine Zerlegung ist. Mit Induktion lässt sich leicht zeigen, dass $\bigvee_{i=0}^{s-1} T^{-i}\mathcal{Z}$ eine Zerlegung von X ist. \square

Für eine Zerlegung $\mathcal{Z} = \{A_1, \dots, A_k\}$ von X betrachte nun die Mengen in

$$\bigvee_{i=0}^{s-1} T^{-i}\mathcal{Z} = \left\{ \bigcap_{i=0}^{s-1} T^{-i}A_{j_i} : j_0, j_1, \dots, j_{s-1} \in \{1, \dots, k\} \right\} \setminus \emptyset :$$

$$\begin{aligned} \bigcap_{i=0}^{s-1} T^{-i}A_{j_i} &= \{x \in X : x \in A_{j_0}, x \in T^{-1}A_{j_1}, \dots, x \in T^{-(s-1)}A_{j_{s-1}}\} \\ &= \{x \in X : x \in A_{j_0}, Tx \in A_{j_1}, \dots, T^{s-1}x \in A_{j_{s-1}}\}, \end{aligned}$$

vergleiche [You03, S. 315]. Die Zustandsmenge X wird daher in alle möglichen Orbits

der Länge s aufgeteilt. Daher ist die durchschnittliche Informationsmenge, die man erhält, wenn man für ein $x \in X$ überprüft, in welchen $A \in \mathcal{Z}$ die Komponenten des Orbits bezüglich dieses x liegen, gleich $h(\bigvee_{i=0}^{s-1} T^{-i} \mathcal{Z})$ [ES14, S. 96].

Dieser Ausdruck ist von s und von der Zerlegung \mathcal{Z} abhängig. Analog zur Untersuchung der Boltzmann-Entropie in Kapitel 4.4. könnte die Abhängigkeit von s umgangen werden, indem der Grenzwert der Entropie pro Zeit für $s \rightarrow \infty$ betrachtet wird. Nach [Sin15, S. 12] existiert dieser Grenzwert für jede endliche Zerlegung von X .

Definition 5.2.3. Sei (X, \mathcal{A}, P, T) ein maßerhaltendes dynamisches System und \mathcal{Z} eine endliche Zerlegung von X . Die *Entropie von T bezüglich der Zerlegung \mathcal{Z}* ist durch

$$h_P(T, \mathcal{Z}) := \lim_{s \rightarrow \infty} \frac{1}{s} h\left(\bigvee_{i=0}^{s-1} T^{-i} \mathcal{Z}\right)$$

gegeben [ES14, S. 96]

Wie in der Definition der topologischen Entropie lässt sich die Abhängigkeit von \mathcal{Z} umgehen, indem das Supremum aller Zerlegungen gewählt wird (dieses existiert in den erweiterten reellen Zahlen, da $\{h_P(T, \mathcal{Z}) : \mathcal{Z} \text{ ist Zerlegung von } X\}$ nicht leer ist, da $X \neq \emptyset$). Dies führt zu folgender Definition:

Definition 5.2.4. Sei (X, \mathcal{A}, P, T) ein maßerhaltendes topologisches dynamisches System. Der Ausdruck

$$h_P(T) := \sup_{\mathcal{Z} \text{ endl. Zerl. von } X} h_P(T, \mathcal{Z})$$

heißt die *metrische Entropie* der Transformation T . Zur Kennzeichnung, welcher Zustandsraum betrachtet wird, wird auch $h_P(X, T, \mathcal{Z})$ und $h_P(X, T)$ geschrieben. ([Sin15, S. 13] und [You03, S. 315]).

Um die bisher genannten Entropiebegriffe der dynamischen Systeme und die Shannon-Entropie (und damit auch die Boltzmann-Entropie) bezüglich des Teilchensystems zu verbinden werden folgende Definitionen und Sätze verwendet:

Definition 5.2.5. Sei (X, \mathcal{A}, P, T) ein maßerhaltendes topologisches dynamisches System. Eine Zerlegung \mathcal{Z} von X heißt (*einseitiger*) *Erzeuger*, wenn die σ -Algebra \mathcal{A} von der Menge $\bigcup_{s=1}^{\infty} (\bigvee_{i=0}^{s-1} T^{-i} \mathcal{Z})$ erzeugt wird [ES14, S. 103].

Für die folgenden Sätze vergleiche [ES14, S. 105] und [You03, S. 316]:

Satz 5.2.6. *Sei (X, \mathcal{A}, P, T) ein maßerhaltendes topologisches dynamisches System und \mathcal{Z} eine endliche Zerlegung von X . Ist \mathcal{Z} ein Erzeuger, dann gilt $h_P(T) = h_P(T, \mathcal{Z})$.*

Satz 5.2.7 (Variationsprinzip). *Sei (X, T) ein dynamisches System. Dann gilt, dass $h_{top}(T) = \sup_P h_P(T)$, wobei das Supremum bezüglich aller Wahrscheinlichkeitsmaße auf der Borel- σ -Algebra $\mathcal{B}(X)$ gemeint ist, bezüglich derer T maßerhaltend ist.*

5.3. Verbindung mit der Shannon-Entropie

Bei der Betrachtung der Teilchensysteme aus unabhängigen und ununterscheidbaren Teilchen wurde bisher folgende statische Situation betrachtet: Die Verteilung von N bzw. abzählbar unendlich vieler solcher Teilchen auf n Niveaus wurde zu einem festen Zeitpunkt untersucht. Für diese Situation wurden Resultate bezüglich der Shannon- und der Boltzmann-Entropie hergeleitet.

In Kapitel 3.4. wurde die Verteilung eines Teilchens eines Systems aus N unabhängigen und nicht unterscheidbaren Teilchen in einem kubischen Raum untersucht. Stellt man sich diese Teilchen als Gasteilchen eines idealen Gases vor, so wird folgendes ersichtlich: Die Teilchen eines Gases sind (bei einer Temperatur größer 0) ständig in Bewegung, wodurch sich der Ortszustand eines Teilchens mit der Zeit ändern kann. Deshalb werden in diesem Unterkapitel Teilchen, deren Zustände sich mit der Zeit ändern können, betrachtet und mit Hilfe der Theorie dynamischer Systeme untersucht.

Hierfür wird abermals der Wahrscheinlichkeitsraum $([0, a]^3, \mathcal{B}([0, a]^3), \frac{1}{a^3}\lambda)$ betrachtet. In Kapitel 3.4. wurde ein Wahrscheinlichkeitsvektor bezüglich der Zerlegung

$$\mathcal{W} := \left\{ \left[\frac{a \cdot (j-1)}{m}, \frac{a \cdot j}{m} \right) \times \left[\frac{a \cdot (k-1)}{m}, \frac{a \cdot k}{m} \right) \times \left[\frac{a \cdot (l-1)}{m}, \frac{a \cdot l}{m} \right) : \right. \\ \left. j, k, l \in \{1, \dots, m\} \right\},$$

wobei $|\mathcal{W}| = m^3$ ist, durch Anwendung von Proposition 3.4.4. gefunden. Wird jeder Teilwürfel $A \in \mathcal{W}$ von 1 bis m^3 durchnummeriert, lässt sich die Verteilung eines Teilchens im Raum bezüglich dieser Zerlegung auch durch den Wahrscheinlichkeitsraum $(\Omega = \{\alpha_1, \dots, \alpha_n\}, \mathfrak{P}(\Omega), p)$ mit $n = m^3$ beschreiben, wobei $p_j := \frac{1}{a^3}\lambda(A_j)$ gesetzt wird.

Allgemeiner lässt sich die Verteilung eines Teilchens im Raum bezüglich einer beliebigen Zerlegung $\mathcal{Q} = \{Q_1, \dots, Q_n\}$ durch $(\Omega = \{\alpha_1, \dots, \alpha_n\}, \mathfrak{P}(\Omega), p)$ mit $n = |\mathcal{Q}|$ und $p_j = \frac{1}{a^3} \lambda(Q_j)$ beschreiben. Entsprechend beschreibt Ω^N mit $p^{\otimes N}$ die Verteilung von N Teilchen im Raum bezüglich einer Zerlegung \mathcal{Q} .

Daher wird nun die Menge $\Omega = \{\alpha_1, \dots, \alpha_n\}$ mit einem Wahrscheinlichkeitsvektor p auf Ω betrachtet und als Verteilung eines Teilchens im Raum bezüglich einer beliebigen Zerlegung interpretiert. Für ein $j \in \{1, \dots, n\}$ wird α_j als der j -te Ortszustand angesehen. Wie schon bei der Behandlung der dynamischen Systeme wird die Zeit diskret betrachtet. Die zeitliche Entwicklung des Aufenthaltsorts eines Teilchens im kubischen Raum wird daher durch eine Folge $(\omega_s)_{s \in \mathbb{N}_0} \in \Omega^{\mathbb{N}_0}$ beschrieben, wobei ω_j den Aufenthaltsort des Teilchens zum Zeitpunkt j beschreibt. Die Verteilung p des Orts soll in dieser Betrachtung unabhängig vom (diskreten) Zeitpunkt, an dem das System betrachtet wird, sein (In der Zeitspanne zwischen zwei Messungen sind die „Karten wieder neu gemischt“: Dass ein Teilchen an einem Messzeitpunkt im Zustand α_i war, beeinflusst die Wahrscheinlichkeit nicht, dass ein Teilchen zu einem späteren Messzeitpunkt im Zustand α_k ist). Daher soll die Wahrscheinlichkeit der Zylindermenge $[A_0, A_1, \dots, A_{s-1}]$ mit $A_0, A_1, \dots, A_{s-1} \subseteq \Omega$ gleich

$$\left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_0}} p_j \right) \cdot \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_1}} p_j \right) \cdots \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_{s-1}}} p_j \right)$$

sein. Diese Überlegung führt analog wie in Kapitel 2.1. zum Wahrscheinlichkeitsraum $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}) := (\Omega^{\mathbb{N}_0}, \mathfrak{P}(\Omega)^{\otimes \mathbb{N}_0}, P^{\otimes \mathbb{N}_0})$, wobei wieder $\mathfrak{P}(\Omega)^{\otimes \mathbb{N}_0}$ gleich der kleinsten σ -Algebra ist, die alle Zylindermengen enthält und $P^{\otimes \mathbb{N}_0}$ das eindeutige Maß auf $\mathfrak{P}(\Omega)^{\otimes \mathbb{N}_0}$ ist, welches für alle Zylindermengen $[A_0, A_1, \dots, A_{s-1}]$

$$P^{\otimes \mathbb{N}_0}([A_0, A_1, \dots, A_{s-1}]) = \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_0}} p_j \right) \cdot \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_1}} p_j \right) \cdots \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_{s-1}}} p_j \right)$$

erfüllt. \hat{P} wird das *Bernoulli-Maß* bezüglich des Wahrscheinlichkeitsvektors p genannt [BS02, S. 219].

Damit die Überlegungen und Resultate aus den beiden vorangegangenen Unterkapitel für die weitere Betrachtung des Teilchensystems angewendet werden können, wird ein maßerhaltendes dynamisches System, also ein kompakter Raum und eine maßerhaltende, stetige Abbildung auf diesem Raum benötigt:

Nach [Vri14, S. 221-223] ist $\Omega^{\mathbb{N}_0}$ mit der Metrik $d : \Omega^{\mathbb{N}_0} \times \Omega^{\mathbb{N}_0} \rightarrow [0, +\infty)$ definiert durch

$$d((\omega_s)_{s \in \mathbb{N}_0}, (\theta_s)_{s \in \mathbb{N}_0}) := \begin{cases} \frac{1}{1 + \min\{s \in \mathbb{N}_0 : \omega_s \neq \theta_s\}} & \text{falls } (\omega_s)_{s \in \mathbb{N}_0} \neq (\theta_s)_{s \in \mathbb{N}_0}, \\ 0 & \text{falls } (\omega_s)_{s \in \mathbb{N}_0} = (\theta_s)_{s \in \mathbb{N}_0} \end{cases}$$

ein kompakter metrischer Raum. Die *Shift-Abbildung* $\tau : \Omega^{\mathbb{N}_0} \rightarrow \Omega^{\mathbb{N}_0}$, die durch

$$\tau((\omega_s)_{s \in \mathbb{N}_0}) := (\omega_{s+1})_{s \in \mathbb{N}_0}$$

definiert ist, ist bezüglich d eine stetige Abbildung [Vri14, S. 223] (in den nachfolgenden Überlegungen wird klar werden, warum hier τ als Transformation gewählt wird). Daher ist das Paar $(\Omega^{\mathbb{N}_0}, \tau)$ ein topologisches dynamisches System.

Proposition 5.3.1. $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}, \tau)$ ist ein maßerhaltendes topologisches dynamisches System.

Beweis. Es muss nur noch gezeigt werden, dass τ eine maßerhaltende Transformation ist. Dafür wird das *Prinzip der guten Mengen* [Wer09, S. 213] angewandt [AB19, S. 11]: Sei $\hat{\mathcal{A}}_0$ die Familie der Mengen $\hat{A} \in \hat{\mathcal{A}}$, die $\hat{P}(\tau^{-1}\hat{A}) = \hat{P}(\hat{A})$ erfüllen. Daher gilt $\hat{\mathcal{A}}_0 \subseteq \hat{\mathcal{A}}$. Es wird gezeigt, dass die beiden Aussagen

1. $\hat{\mathcal{A}}_0$ enthält die Menge aller Zylindermengen
2. $\hat{\mathcal{A}}_0$ ist eine σ -Algebra

gelten.

Zu (1): Für eine beliebige Zylindermenge $[A_0, A_1, \dots, A_{s-1}]$ gilt $\tau^{-1}[A_0, A_1, \dots, A_{s-1}] = [\Omega, A_0, A_1, \dots, A_{s-1}]$ [AB19, S. 11], denn

$$\begin{aligned} \hat{\omega} \in \tau^{-1}[A_0, A_1, \dots, A_{s-1}] &\Leftrightarrow \tau(\hat{\omega}) \in [A_0, A_1, \dots, A_{s-1}] \\ &\Leftrightarrow \tau(\hat{\omega}) = (\omega_0, \omega_1, \dots, \omega_{s-1}, \dots) \text{ für } \omega_0 \in A_0, \omega_1 \in A_1, \dots, \omega_{s-1} \in A_{s-1} \\ &\Leftrightarrow \hat{\omega} = (\omega, \omega_0, \omega_1, \dots, \omega_{s-1}, \dots) \text{ für } \omega \in \Omega, \omega_0 \in A_0, \omega_1 \in A_1, \dots, \omega_{s-1} \in A_{s-1} \\ &\Leftrightarrow \hat{\omega} \in [\Omega, A_0, A_1, \dots, A_{s-1}]. \end{aligned}$$

Daraus folgt, dass

$$\begin{aligned}
 \hat{P}(\tau^{-1}[A_0, A_1, \dots, A_{s-1}]) &= \hat{P}([\Omega, A_0, A_1, \dots, A_{s-1}]) \\
 &= \underbrace{\left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in \Omega}} p_j \right)}_{=1} \cdot \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_0}} p_j \right) \cdot \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_1}} p_j \right) \cdots \left(\sum_{\substack{j \in \{1, 2, \dots, n\} \\ \alpha_j \in A_{s-1}}} p_j \right) \\
 &= \hat{P}([A_0, A_1, \dots, A_{s-1}]).
 \end{aligned}$$

Daher sind alle Zylindermengen in $\hat{\mathcal{A}}_0$ enthalten.

Zu (2): Die Eigenschaften einer σ -Algebra werden überprüft.

- (i) Da $\tau^{-1}\Omega = \Omega$, ist $\hat{P}(\tau^{-1}\Omega) = \hat{P}(\Omega)$.
- (ii) Sei $\hat{A} \in \hat{\mathcal{A}}_0$ beliebig. Aus Proposition 2.1.4.(iv) und Proposition 2.2.9.(ii) folgt

$$\begin{aligned}
 \hat{P}(\hat{A}^c) &= \hat{P}(\underbrace{\tau^{-1}(\hat{\Omega} \setminus \hat{A})}_{=(\tau^{-1}\hat{\Omega}) \setminus (\tau^{-1}\hat{A})}) = \hat{P}(\hat{\Omega} \setminus \tau^{-1}\hat{A}) \\
 &= 1 - \hat{P}(\tau^{-1}\hat{A}) = 1 - \hat{P}(\hat{A}) = \hat{P}(\hat{A}^c).
 \end{aligned}$$

Daher gilt auch $\hat{A}^c \in \hat{\mathcal{A}}_0$.

- (iii) Sei $(\hat{A}_m)_{m \in \mathbb{N}}$ eine Folge in $\hat{\mathcal{A}}_0$. Aus Proposition 2.2.9.(i) folgt

$$\begin{aligned}
 \hat{P}(\tau^{-1}(\bigcup_{m=1}^{\infty} \hat{A}_m)) &= \hat{P}(\bigcup_{m=1}^{\infty} \tau^{-1}\hat{A}_m) = \sum_{m=1}^{\infty} \hat{P}(\tau^{-1}\hat{A}_m) \\
 &= \sum_{m=1}^{\infty} \hat{P}(\hat{A}_m) = \hat{P}(\bigcup_{m=1}^{\infty} \hat{A}_m).
 \end{aligned}$$

Somit ist $\bigcup_{m=1}^{\infty} \hat{A}_m \in \hat{\mathcal{A}}_0$.

Insgesamt ist damit gezeigt, dass $\hat{\mathcal{A}}_0$ eine σ -Algebra ist.

$\hat{\mathcal{A}}$ ist die kleinste σ -Algebra, die alle Zylindermengen enthält. Da $\hat{\mathcal{A}}_0$ eine σ -Algebra ist, die alle Zylindermengen enthält, gilt $\hat{\mathcal{A}} \subseteq \hat{\mathcal{A}}_0$. Daher gilt $\hat{\mathcal{A}} = \hat{\mathcal{A}}_0$. Somit ist τ eine ma erhaltende Transformation auf $\hat{\Omega}$. \square

Ein Zusammenhang zwischen topologischer-, metrischer- und Shannon-Entropie lässt sich, wie in [AB19, S. 10-11] oder in [ES14, S. 104, Bsp. 3.26] gezeigt wird, wobei hier im Folgenden die Vorgehensweise von [ES14, S. 104, Bsp. 3.26] übernommen und für das N Teilchensystem interpretiert wird, wie folgt finden:

Betrachte für $s \in \mathbb{N}$ die Zerlegung

$$\mathcal{Z}_s := \{[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] : j_0, j_1, \dots, j_{s-1} \in \{1, \dots, n\}\}$$

von $\hat{\Omega}$, wobei $[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] := [\{\alpha_{j_0}\}, \{\alpha_{j_1}\}, \dots, \{\alpha_{j_{s-1}}\}]$. Es gilt folgende Proposition:

Proposition 5.3.2. *Ist $\mathcal{Z} := \mathcal{Z}_1$, dann gilt $\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z} = \mathcal{Z}_s$.*

Beweis. Wie im Beweis von Proposition 5.3.1. bereits gezeigt wurde, gilt

$$\tau^{-1}[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] = [\Omega, \alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}].$$

Daraus folgt weiter

$$\tau^{-i}[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] = \underbrace{[\Omega, \dots, \Omega]}_{i\text{-mal}}, \alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}].$$

Daher gilt

$$\begin{aligned} \tau^{-i} \mathcal{Z} &= \{\tau^{-i}[\alpha_1], \tau^{-i}[\alpha_2], \dots, \tau^{-i}[\alpha_n]\} \\ &= \left\{ \underbrace{[\Omega, \dots, \Omega]}_{i\text{-mal}}, \alpha_1 \right\}, \left\{ \underbrace{[\Omega, \dots, \Omega]}_{i\text{-mal}}, \alpha_2 \right\}, \dots, \left\{ \underbrace{[\Omega, \dots, \Omega]}_{i\text{-mal}}, \alpha_n \right\}. \end{aligned}$$

Seien $\hat{A}_i \in \tau^{-i} \mathcal{Z}$ für $i \in \{0, 1, \dots, s-1\}$, das heißt $\hat{A}_i = \underbrace{[\Omega, \dots, \Omega]}_{i\text{-mal}}, \alpha_{j_i}$, $j_i \in \{1, \dots, n\}$.

Es gilt $\bigcap_{i=0}^{s-1} \hat{A}_i = [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$, denn:

IA: $\bigcap_{i=0}^{s-1} \hat{A}_i = [\alpha_{j_0}]$.

IV: Für $s > 1$: $\bigcap_{i=0}^{s-1} \hat{A}_i = [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$.

IS: Sei $s > 1$.

$$\begin{aligned}
 \bigcap_{i=0}^{s-1} \hat{A}_i &= \left(\bigcap_{i=0}^{s-2} \hat{A}_i \right) \cap \hat{A}_{s-1} \\
 &= [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-2}}] \cap \underbrace{[\Omega, \dots, \Omega, \alpha_{j_{s-1}}]}_{(s-1)\text{-mal}} \\
 &= [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}],
 \end{aligned}$$

denn es gilt

$$\begin{aligned}
 \hat{\omega} &= (\omega_0, \omega_1, \dots, \omega_{s-2}, \omega_{s-1}, \dots) \in [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-2}}] \cap \underbrace{[\Omega, \dots, \Omega, \alpha_{j_{s-1}}]}_{(s-1)\text{-mal}} \\
 &\Leftrightarrow (\omega_0 = \alpha_{j_0}, \omega_1 = \alpha_{j_1}, \dots, \omega_{s-2} = \alpha_{j_{s-2}}) \wedge (\omega_0, \omega_1, \dots, \omega_{s-2} \in \Omega \wedge \omega_{s-1} = \alpha_{j_{s-1}}) \\
 &\Leftrightarrow \hat{\omega} \in [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}].
 \end{aligned}$$

Umgekehrt lässt sich analog zeigen, dass sich ein beliebiges $[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$ aus \mathcal{Z}_s darstellen lässt durch

$$[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] = \bigcap_{i=0}^{s-1} \underbrace{[\Omega, \dots, \Omega, \alpha_{j_i}]}_{i\text{-mal}}$$

Daraus folgt letztendlich, dass

$$\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z} = \left\{ \bigcap_{i=0}^{s-1} \hat{A}_i : \hat{A}_i \in \tau^{-i} \mathcal{Z} \right\} = \mathcal{Z}_s$$

erfüllt ist. □

Wie in Kapitel 5.2. schon allgemein gezeigt, haben die Mengen in $\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}$ mit $\mathcal{Z} = \{[\alpha_1], [\alpha_2], \dots, [\alpha_n]\}$ die Form

$$\{\hat{\omega} \in \hat{\Omega} : \hat{\omega} \in [\alpha_{j_0}], \tau \hat{\omega} \in [\alpha_{j_1}], \dots, \tau^{s-1} \hat{\omega} \in [\alpha_{j_{s-1}}]\}$$

Da für ein $\hat{\omega} = (\omega_0, \omega_1, \dots) \in \hat{\Omega}$ gilt, dass $\hat{\omega} \in [\alpha_{j_0}]$ gleichbedeutend ist mit $\omega_0 = \alpha_{j_0}$ und $\tau \hat{\omega} = (\omega_1, \omega_2, \dots)$ ist, folgt, dass $\tau^i \hat{\omega} \in [\alpha_{j_i}]$ gleichbedeutend ist mit $\omega_i = \alpha_{j_i}$.

Daher sind die Mengen in $\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}$ sogar von der Form

$$\{\hat{\omega} = (\omega_0, \omega_1, \dots) \in \hat{\Omega} : \omega_0 = \alpha_{j_0}, \omega_1 = \alpha_{j_1}, \dots, \omega_{s-1} = \alpha_{j_{s-1}}\}$$

Dementsprechend ist $\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}$ die Zerlegung von $\hat{\Omega}$ in alle möglichen Orbits der Länge s . Der Ausdruck $h(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z})$ ist daher der durchschnittliche Informationsgehalt, der erhalten wird, wenn für ein Teilchen überprüft wird, in welchem Zustand dieses Teilchen im Zeitraum zwischen 0 und $s - 1$ liegt. Der Grenzwert des durchschnittlichen Informationsgehalts pro Zeiteinheit ist damit

$$\lim_{s \rightarrow \infty} \frac{1}{s} h\left(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}\right).$$

Wie sich nachfolgend herausstellen wird, entspricht dieser Ausdruck der metrischen Entropie dieses maßerhaltenden dynamischen Systems.

Proposition 5.3.3. *Die Zerlegung \mathcal{Z} ist ein Erzeuger, erfüllt also*

$$\sigma\left(\bigcup_{s=1}^{\infty} \left(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}\right)\right) = \hat{\mathcal{A}}.$$

Beweis. (\subseteq) Es gilt

$$\begin{aligned} \bigcup_{s=1}^{\infty} \underbrace{\left(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}\right)}_{=\mathcal{Z}_s} &= \bigcup_{s=1}^{\infty} \mathcal{Z}_s = \{[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] : s \in \mathbb{N}, j_0, j_1, \dots, j_{s-1} \in \{1, \dots, n\}\} \\ &\subseteq \{[A_0, A_1, \dots, A_{s-1}] : s \in \mathbb{N}, A = 0, A = 1, \dots, A_{s-1}\} \\ &\Rightarrow \sigma\left(\bigcup_{s=1}^{\infty} \mathcal{Z}_s\right) \subseteq \hat{\mathcal{A}}. \end{aligned}$$

(\supseteq) Es wird analog wie im Beweis von Satz 2.1.12. vorgegangen: Es gilt

$$[A_0, A_1, \dots, A_{s-1}] = \bigcup_{\substack{j_0=1, j_1=1, \dots, j_{s-1}=1 \\ \alpha_{j_0} \in A_0, \alpha_{j_1} \in A_1, \dots, \alpha_{j_{s-1}} \in A_{s-1}}}^n [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}].$$

Ist somit \mathcal{F} eine σ -Algebra mit $\bigcup_{s=1}^{\infty} \mathcal{Z}_s \subseteq \mathcal{F}$, so ist

$$[A_0, A_1, \dots, A_{s-1}] = \bigcup_{\substack{j_0=1, j_1=1, \dots, j_{s-1}=1 \\ \alpha_{j_0} \in A_0, \alpha_{j_1} \in A_1, \dots, \alpha_{j_{s-1}} \in A_{s-1}}}^n [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}] \in \mathcal{F}.$$

Demzufolge ist $\{[A_0, A_1, \dots, A_{s-1}] : s \in \mathbb{N}, A_0, A_1, \dots, A_{s-1} \in \mathcal{A}\} \subseteq \mathcal{F}$. Daher gilt

$$\begin{aligned} & \{\mathcal{F} \text{ ist } \sigma\text{-Algebra} : \bigcup_{s=1}^{\infty} \mathcal{Z}_s \subseteq \mathcal{F}\} \\ & \subseteq \{\mathcal{F} \text{ ist } \sigma\text{-Algebra} : \{[A_0, A_1, \dots, A_{s-1}] : s \in \mathbb{N}, A_0, A_1, \dots, A_{s-1} \in \mathcal{A}\} \subseteq \mathcal{F}\} \end{aligned}$$

Damit erhalt man

$$\hat{\mathcal{A}} = \bigcap_{\substack{\{[A_0, A_1, \dots, A_{s-1}] : s \in \mathbb{N}, A_0, A_1, \dots, A_{s-1} \in \mathcal{A}\} \subseteq \mathcal{F} \\ \mathcal{F} \text{ ist } \sigma\text{-Algebra}}} \mathcal{F} \subseteq \bigcap_{\substack{\bigcup_{s=1}^{\infty} \mathcal{Z}_s \subseteq \mathcal{F} \\ \mathcal{F} \text{ ist } \sigma\text{-Algebra}}} \mathcal{F} = \sigma\left(\bigcup_{s=1}^{\infty} \mathcal{Z}_s\right),$$

da im linken Durchschnitt mindestens so viele Mengen geschnitten werden wie im rechten Durchschnitt. \square

Da \mathcal{Z} ein Erzeuger ist, gilt nach Satz 5.2.6. fur ein beliebiges Wahrscheinlichkeitsma μ auf $\hat{\mathcal{A}}$, fur welches τ maerhaltend ist, die Gleichung

$$h_{\mu}(\tau) = h_{\mu}(\tau, \mathcal{Z}) = \lim_{s \rightarrow \infty} -\frac{1}{s} \sum_{\hat{A} \in \mathcal{Z}_s} \mu(\hat{A}) \log(\mu(\hat{A})).$$

Mit Hilfe folgender Proposition [AB19, S. 11] kann ein Zusammenhang zwischen der metrischen- und der Shannon-Entropie hergestellt werden .

Proposition 5.3.4. *Fur \hat{P} gilt $-\sum_{\hat{A} \in \mathcal{Z}_s} \hat{P}(\hat{A}) \log \hat{P}(\hat{A}) = s \cdot h(p)$ fur alle $s \in \mathbb{N}$.*

Beweis. Der Beweis wird mittels Induktion nach $s \in \mathbb{N}$ gefuhrt.

IA: Sei $s = 1$. Dann gilt $-\sum_{\hat{A} \in \mathcal{Z}_1} \hat{P}(\hat{A}) \log \hat{P}(\hat{A}) = -\sum_{j=1}^n p_j \log p_j = 1 \cdot h(p)$.

IV: Fur $s > 1$ gilt $-\sum_{\hat{A} \in \mathcal{Z}_s} \hat{P}(\hat{A}) \log \hat{P}(\hat{A}) = s \cdot h(p)$.

IS: Sei $s > 1$. Es gilt

$$-\sum_{\hat{A} \in \mathcal{Z}_s} \hat{P}(\hat{A}) \log \hat{P}(\hat{A}) = -\sum_{j_0=1}^n \sum_{j_1=1}^n \cdots \sum_{j_{s-2}=1}^n \sum_{j_{s-1}=1}^n \prod_{k=0}^{s-1} p_{j_k} \log \left(\prod_{k=0}^{s-1} p_{j_k} \right).$$

Durch analoges Vorgehen wie im Beweis von Satz 3.3.2. erhält man daher durch Anwendung der Induktionsvoraussetzung

$$-\sum_{\hat{A} \in \mathcal{Z}_s} \hat{P}(\hat{A}) \log \hat{P}(\hat{A}) = s \cdot h(p).$$

□

Aus dieser Proposition folgt, dass der durchschnittliche Informationsgehalt, den man erhält, wenn man für ein Teilchen überprüft, in welchem Zustand dieses Teilchen im Zeitraum zwischen 0 und $s-1$ liegt, pro Zeiteinheit gleich der Shannon-Entropie eines Teilchens im Teilchensystem ist: $\frac{1}{s} h(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}) = h(p)$.

Daraus folgt weiter, dass auch die metrische Entropie bezüglich τ gleich der Shannon-Entropie ist:

Satz 5.3.5. *Es gilt*

$$h_{\hat{P}}(\tau) = h(p).$$

Beweis. Durch Anwendung von Proposition 5.3.4. ergibt sich

$$\begin{aligned} h_{\hat{P}}(\tau) &= h_{\hat{P}}(\tau, \mathcal{Z}) = \lim_{s \rightarrow \infty} \frac{1}{s} \cdot \underbrace{\left(-\sum_{\hat{A} \in \mathcal{Z}_s} \hat{P}(\hat{A}) \log(\hat{P}(\hat{A})) \right)}_{=s \cdot h(p)} \\ &= \lim_{s \rightarrow \infty} \frac{1}{s} s h(p) = h(p). \end{aligned}$$

□

5.4. Maximale metrische Entropie der Shift-Abbildung

Aufgrund von $|\mathcal{Z}_s| = n^s$ und Satz 3.5.2 gilt für ein beliebiges Maß μ auf $\hat{\mathcal{A}}$, für welches τ maßerhaltend ist, die Ungleichung $-\sum_{\hat{A} \in \mathcal{Z}_s} \mu(\hat{A}) \log(\mu(\hat{A})) \leq \log(n^s) = s \cdot \log n$. Daraus folgt weiter, dass

$$h_\mu(\tau) = h_\mu(\tau, \mathcal{Z}) = \lim_{s \rightarrow \infty} \frac{1}{s} \cdot \underbrace{\left(- \sum_{\hat{A} \in \mathcal{Z}_s} \mu(\hat{A}) \log(\mu(\hat{A})) \right)}_{\leq s \cdot \log n} \leq \lim_{s \rightarrow \infty} \log n = \log n.$$

Um den Satz 5.2.7. anwenden zu können, wird gezeigt, dass $\hat{\mathcal{A}}$ die Borel- σ -Algebra bezüglich $\hat{\Omega}$ ist:

Proposition 5.4.1. *Es gilt $\hat{\mathcal{A}} = \mathcal{B}(\hat{\Omega})$.*

Beweis. (\subseteq) Betrachte für ein $s \in \mathbb{N}$ die Zylindermenge $[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$. Sei $\hat{\omega} \in [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$ beliebig. Wähle $r = \frac{1}{s}$. Die offene Kugel mit Radius $r = \frac{1}{s}$ bezüglich der in Kapitel 5.3. eingeführten Metrik auf $\hat{\Omega}$ ist dann $B(\hat{\omega}, \frac{1}{s}) = \{\hat{x} \in \hat{\Omega} : d(\hat{x}, \hat{\omega}) < \frac{1}{s}\}$. Sei $\hat{x} \in B(\hat{\omega}, \frac{1}{s})$ beliebig. Ist $\hat{x} = \hat{\omega}$, so ist $\hat{x} \in [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$. Ist $\hat{x} \neq \hat{\omega}$, dann gilt

$$d(\hat{x}, \hat{\omega}) = \frac{1}{1 + \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}} < \frac{1}{s} \Leftrightarrow s - 1 < \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}.$$

Daraus folgt, dass für alle $k \in \{0, 1, \dots, s-1\}$ gilt, dass $x_k = \omega_k = \alpha_{j_k}$. Daher ist $\hat{x} \in [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$. Insgesamt gilt demnach $B(\hat{\omega}, \frac{1}{s}) \subseteq [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$. Daher ist $[\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}]$ offen.

Da für eine Zylindermenge $[A_0, A_1, \dots, A_{s-1}]$ mit $A_0, A_1, \dots, A_{s-1} \subseteq \Omega$ gilt, dass

$$[A_0, A_1, \dots, A_{s-1}] = \bigcup_{\substack{j_0=1, j_1=1, \dots, j_{s-1}=1 \\ \alpha_{j_0} \in A_0, \alpha_{j_1} \in A_1, \dots, \alpha_{j_{s-1}} \in A_{s-1}}} [\alpha_{j_0}, \alpha_{j_1}, \dots, \alpha_{j_{s-1}}].$$

und Vereinigungen offener Mengen offen sind [Kal15, S. 129], ist daher $[A_0, A_1, \dots, A_{s-1}]$ offen. Die Menge der Zylindermengen ist aus diesem Grund in der Menge der offenen Mengen enthalten und deshalb folgt $\hat{\mathcal{A}} \subseteq \mathcal{B}(\hat{\Omega})$.

(\supseteq) Sei $\hat{U} \subseteq \hat{\Omega}$ offen und sei $\hat{\omega} \in \hat{U}$. Es existiert ein $r = \frac{1}{t} > 0$, so dass $B(\hat{\omega}, \frac{1}{t}) = \{\hat{x} \in \hat{\Omega} : d(\hat{x}, \hat{\omega}) < \frac{1}{t}\} \subseteq \hat{U}$. Betrachte $B(\hat{\omega}, \frac{1}{t}) \setminus \{\hat{\omega}\} = \{\hat{x} \in \hat{\Omega} : \frac{1}{1 + \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}} < \frac{1}{t}\}$. Es

gilt folgende Äquivalenzkette:

$$\begin{aligned} \hat{x} \in B(\hat{\omega}, \frac{1}{t}) \setminus \{\hat{\omega}\} &\Leftrightarrow \frac{1}{1 + \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}} < \frac{1}{t} \\ &\Leftrightarrow t - 1 < \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\} \Leftrightarrow \lfloor t - 1 \rfloor < \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}. \end{aligned}$$

Daher ist $B(\hat{\omega}, \frac{1}{t}) \setminus \{\hat{\omega}\} = \{\hat{x} \in \hat{\Omega} : \lfloor t - 1 \rfloor < \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}\}$.

1. Fall: $t \in (0, 1)$. Dann ist $\lfloor t - 1 \rfloor = -1$. Demnach gilt $B(\hat{\omega}, \frac{1}{t}) \setminus \{\hat{\omega}\} = \{\hat{x} \in \hat{\Omega} : 0 \leq \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}\} = \hat{\Omega} \setminus \{\hat{\omega}\}$. Daraus ergibt sich, dass $B(\hat{\omega}, \frac{1}{t}) = \{\hat{\omega}\} \cup \hat{\Omega} \setminus \{\hat{\omega}\} = \hat{\Omega} \subseteq \hat{U}$, also $\hat{U} = \hat{\Omega}$.

2. Fall: $t \geq 1 \Rightarrow \lfloor t - 1 \rfloor \geq 0$. Daher ist $B(\hat{\omega}, \frac{1}{t}) \setminus \{\hat{\omega}\} = \{\hat{x} \in \hat{\Omega} : 0 \leq \lfloor t - 1 \rfloor < \min\{j \in \mathbb{N}_0 : x_j \neq \omega_j\}\} = [\omega_0, \dots, \omega_{\lfloor t-1 \rfloor}] \setminus \{\hat{\omega}\}$. Daher ist $B(\hat{\omega}, \frac{1}{t}) = \{\hat{\omega}\} \cup [\omega_0, \dots, \omega_{\lfloor t-1 \rfloor}] \setminus \{\hat{\omega}\} = [\omega_0, \dots, \omega_{\lfloor t-1 \rfloor}] \subseteq \hat{U}$.

Es gilt $\hat{U} = \bigcup_{\hat{\omega} \in \hat{U}} B(\hat{\omega}, \frac{1}{t_{\hat{\omega}}})$. Ist eines der $t_{\hat{\omega}} \in (0, 1)$, so ist $\hat{U} = \hat{\Omega} \in \hat{\mathcal{A}}$. Ist keines der $t_{\hat{\omega}} \in (0, 1)$, so ist \hat{U} eine Vereinigung von Zylindermengen der Form $[\alpha_0, \alpha_1, \dots, \alpha_{s-1}]$, $s \in \mathbb{N}$. Es gibt jedoch nur abzählbar unendlich viele solcher Mengen [ES14, S. 31]. Durch folgende Abbildungsvorschrift erhält man eine Bijektion zwischen \mathbb{N} und der Menge der Zylindermengen der Form $[\alpha_0, \alpha_1, \dots, \alpha_{s-1}]$, $s \in \mathbb{N}$: Die ersten n natürlichen Zahlen werden auf die Mengen $[\alpha_1], \dots, [\alpha_n]$ abgebildet, die nächsten n^2 natürlichen Zahlen auf die Mengen $[\alpha_1, \alpha_1], [\alpha_1, \alpha_2], \dots, [\alpha_n, \alpha_n]$, die nächsten n^3 natürlichen Zahlen auf die Mengen $[\alpha_1, \alpha_1, \alpha_1], [\alpha_1, \alpha_1, \alpha_2], \dots, [\alpha_n, \alpha_n, \alpha_n]$ und so weiter.

Deshalb ist \hat{U} , als Vereinigung solcher Zylindermengen, eine abzählbar unendliche Vereinigung solcher Zylindermengen und daher gilt $\hat{U} \in \hat{\mathcal{A}}$. Da somit $\hat{\mathcal{A}}$ eine σ -Algebra ist, die alle offenen Mengen enthält, folgt $\mathcal{B}(\hat{\Omega}) \subseteq \hat{\mathcal{A}}$. \square

Aus Satz 5.2.7. und obiger Ungleichung folgt daher, dass für ein beliebiges Maß μ auf $\hat{\mathcal{A}} = \mathcal{B}(\hat{\Omega})$, für welches τ maßerhaltend ist, gilt:

$$h_{top}(\tau) = \sup_{\mu} h_{\mu}(\tau) \leq \log n.$$

Sei \hat{P}_{GI} das Bernoulli-Maß bezüglich des Wahrscheinlichkeitsvektors $(\frac{1}{n}, \dots, \frac{1}{n})$. Durch Anwendung von Satz 5.3.5. erhält man $h_{\hat{P}_{GI}}(\tau) = h(\frac{1}{n}, \dots, \frac{1}{n}) = \log n$. Daher ist $\log n \leq \sup_{\mu} h_{\mu}(\tau) \leq \log n$. Daher gilt folgender Satz:

Satz 5.4.2. *Es gilt*

$$h_{top}(\tau) = \sup_{\mu} h_{\mu}(\tau) = h_{\hat{P}_{GI}}(\tau) = \log n.$$

Insbesondere gilt, dass die metrische Entropie $h_{\hat{P}}(\tau)$ der Shift-Abbildung bezüglich des Bernoulli-Maßes zum Wahrscheinlichkeitsvektor p genau dann maximal ist, wenn $p = (\frac{1}{n}, \dots, \frac{1}{n})$ [BS02, S. 219].

Werden statt einem Teilchen N unabhängige und nicht unterscheidbare Teilchen untersucht, so erhält man mit $(\Omega^N)^{\mathbb{N}_0}$ gemeinsam mit dem Bernoulli-Maß \hat{P} zum Wahrscheinlichkeitsvektor $p^{\otimes N}$ analog zu den bisher angestellten Überlegungen bezüglich eines Teilchens durch Anwendung von Satz 3.3.2. folgende Resultate: Der durchschnittliche Informationsgehalt, den man erhält, wenn für alle N Teilchen überprüft wird, in welchen Zuständen diese Teilchen im Zeitraum zwischen 0 und $s-1$ liegen, pro Zeiteinheit ist gleich der Shannon-Entropie des Teilchensystems: $\frac{1}{s} h(\bigvee_{i=0}^{s-1} \tau^{-i} \mathcal{Z}^{(N)}) = N \cdot h(p)$, wobei $\mathcal{Z}^{(N)}$ als Zerlegung von $(\Omega^N)^{\mathbb{N}_0}$ analog wie \mathcal{Z} als Zerlegung von $\Omega^{\mathbb{N}_0}$ definiert ist. Außerdem lässt sich analog folgender Satz zeigen:

Satz 5.4.3. *Es gelten*

$$(i) \quad h_{\hat{P}}((\Omega^N)^{\mathbb{N}_0}, \tau) = h_{\hat{P}}((\Omega^N)^{\mathbb{N}_0}, \tau, \mathcal{Z}^{(N)}) = h(p^{\otimes N}) = N \cdot h(p).$$

$$(ii) \quad h_{top}((\Omega^N)^{\mathbb{N}_0}, \tau) = \sup_{\mu} h_{\mu}((\Omega^N)^{\mathbb{N}_0}, \tau) = N \cdot \log n.$$

Die metrische Entropie des gesamten Teilchenssystems entspricht daher der N -fachen Shannon-Entropie eines einzelnen Teilchens. Die topologische Entropie entspricht der maximalen Entropie eines Systems aus N unabhängigen und ununterscheidbaren Teilchen mit zeitunabhängiger Verteilung p , also $N \cdot h(\frac{1}{n}, \dots, \frac{1}{n}) = N \cdot \log n$.

6. Conclusio

Nach der Einführung des Begriffs *Entropie* durch Clausius im Jahre 1865 [ES14, S. 89] wurden in den darauf folgenden Jahrzehnten unterschiedliche Ausdrücke, die sich aus diversen Fragestellungen der Mathematik sowie der Physik heraus ableiteten, als Entropie bezeichnet. Das Interesse an Entropie und die Erforschung neuer und bestehender Entropiebegriffe hält bis heute an.

Die Entropiebegriffe, die in dieser Arbeit betrachtet werden, sind die Shannon-Entropie, die Boltzmann-Entropie, die topologische Entropie und die metrische Entropie. Diese Ausdrücke grenzen sich sowohl in ihrer mathematischen Definition als auch in ihrer kontextbezogenen Interpretation ab.

Ziel dieser Arbeit ist, durch die Betrachtung eines Teilchensystems aus unabhängigen und nicht unterscheidbaren Teilchen Gemeinsamkeiten und formale Zusammenhänge zwischen diesen Entropiebegriffen zu finden.

Folgende Gemeinsamkeiten und Zusammenhänge zwischen der Shannon-Entropie, der Boltzmann-Entropie, der metrischen Entropie und der topologischen Entropie wurden festgestellt:

6.1. Gemeinsamkeiten

Die Herleitungen der Definitionen führten bei allen vier Begriffen zu einem mathematischen Ausdruck, in dem der Logarithmus \log eine zentrale Rolle einnimmt:

$$h(p_1, \dots, p_n) := - \sum_{i=1}^n p_i \log p_i \quad S(W) := k \log W \quad (6.1)$$

$$h_{top}(T) := \lim_{\varepsilon \rightarrow 0^+} \limsup_{s \rightarrow \infty} \frac{1}{s} \log k_s(\varepsilon) \quad h_P(T) := \sup_{\mathcal{Z} \text{ endl. Zerl. von } X} h_P(T, \mathcal{Z}), \quad (6.2)$$

wobei $h_P(T, \mathcal{Z}) := \lim_{s \rightarrow \infty} \frac{1}{s} h(\bigvee_{i=0}^{s-1} T^{-i} \mathcal{Z})$. In (6.1) sind die Definitionen der Shannon- bzw. der Boltzmann-Entropie angeführt, in (6.2) die Definitionen der topologischen bzw. der metrischen Entropie.

Dabei wird der Logarithmus in den Definitionen der Shannon- und der Boltzmann-Entropie aufgrund der Funktionalgleichung und der Monotonieeigenschaft des Logarithmus verwendet. In der Definition der topologischen Entropie findet der Logarithmus Anwendung, da diese Funktion die Umkehrfunktion der Exponentialfunktion ist. Bei der Definition der metrischen Entropie steckt der Logarithmus implizit im Ausdruck $h(\bigvee_{i=0}^{s-1} T^{-i} \mathcal{Z})$.

Eine Gemeinsamkeit der Shannon-Entropie, der Boltzmann-Entropie und der metrischen Entropie, alle bezogen auf eine endliche Grundmenge $\Omega = \{\alpha_1, \dots, \alpha_n\}$, ist ihre Maximumseigenschaft: Die Shannon-Entropie $h(p)$ ist nach Satz 3.5.2. genau dann maximal, wenn die Wahrscheinlichkeitsverteilung $p = (\frac{1}{n}, \dots, \frac{1}{n})$ ist. Boltzmanns Entropieausdruck $S(W)$ mit $W = \binom{N}{m_1, \dots, m_n}$ bezüglich $N \in \mathbb{N}$ unabhängigen und nicht unterscheidbaren Teilchen, die auf n unterschiedliche Energieniveaus verteilt sind, mit $N = q$, ist nach Satz 4.3.2. genau dann maximal, wenn die relative Häufigkeitsverteilung $(\frac{m_1}{N}, \dots, \frac{m_n}{N})$ gleich $(\frac{1}{n}, \dots, \frac{1}{n})$ ist. Bezüglich des maßerhaltenden dynamischen Systems $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}, \tau)$ aus Kapitel 5.3. ist die metrische Entropie bezüglich eines Bernoulli-Maßes $h_{\hat{P}}(\tau)$ zu einem Wahrscheinlichkeitsvektor p und der Shift-Abbildung τ genau dann maximal, wenn $p = (\frac{1}{n}, \dots, \frac{1}{n})$ ist. Die topologische Entropie $h_{top}(\tau)$ bezüglich dieses dynamischen Systems entspricht dieser maximalen metrischen Entropie, also der metrischen Entropie bezüglich des Bernoulli-Maßes zur Wahrscheinlichkeitsverteilung $(\frac{1}{n}, \dots, \frac{1}{n})$.

6.2. Zusammenhänge

Durch Betrachtung eines Teilchensystem aus unabhängigen und nicht unterscheidbaren Teilchen ergaben sich folgende Zusammenhänge zwischen den vier Entropiebegriffen: Nach Korollar 4.4.3.1. ist

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log W_N(\tilde{R}_N) = h(p) = - \sum_{k=1}^n p_k \log p_k \quad \tilde{P} - \text{fast sicher},$$

wobei die Zufallsvariable \tilde{R} wie in Kapitel 4.4. definiert ist. Für ein unendliches Teilchensystem aus unabhängigen und nicht unterscheidbaren Teilchen, welches zu einem festen Zeitpunkt betrachtet wird, lässt sich dieses Resultat wie folgt interpretieren: Die Boltzmann-Entropie pro Teilchen konvergiert „fast sicher“ (im Sinne von Definition 2.1.14.) gegen die Shannon-Entropie eines Teilchens eines solchen Teilchensystems, wenn diese mit der Boltzmann-Konstante k multipliziert wird.

Wird das maßerhaltende dynamische System $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}, \tau)$ aus Kapitel 5.3. betrachtet, so gelten nach Satz 5.3.5., Satz 5.4.2. und Satz 5.4.3. folgende Resultate:

$$h_{\hat{P}}(\tau) = h(p) \tag{6.3}$$

$$h_{top}(\tau) = \sup_{\mu} h_{\mu}(\tau) = h_{\hat{P}_{GI}}(\tau) = h\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = \log n \tag{6.4}$$

$$h_{\hat{P}}((\Omega^N)^{N_0}, \tau) = h(p^{\otimes N}) = N \cdot h(p) \tag{6.5}$$

$$h_{top}((\Omega^N)^{N_0}, \tau) = \sup_{\mu} h_{\mu}((\Omega^N)^{N_0}, \tau) = N \cdot h\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = N \cdot \log n. \tag{6.6}$$

Die Gleichung (6.3) besagt, dass die metrische Entropie bezüglich des maßerhaltenden dynamischen Systems $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}, \tau)$ gleich der Shannon-Entropie bezüglich des Wahrscheinlichkeitsvektors p ist. Die metrische Entropie lässt sich in dieser Betrachtung als der Grenzwert bezüglich der Zeit des durchschnittlichen Informationsgehalts, den man erhält, wenn man für ein Teilchen überprüft, in welchem Zustand dieses Teilchen im Zeitraum zwischen 0 und $s - 1$ liegt, pro Zeiteinheit interpretieren.

Die Gleichung (6.4) verbindet die Shannon-, die topologische- und die metrische Entropie. Die metrische Entropie ist genau dann maximal, wenn im System $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}, \tau)$ \hat{P} das Bernoulli-Maß zur Verteilung $p = (\frac{1}{n}, \dots, \frac{1}{n})$ ist. Die topologische Entropie entspricht der maximalen metrischen Entropie dieses Systems und ist gleich der maximalen Shannon-Entropie eines Teilchens.

Die Gleichungen (6.5) und (6.6) entsprechen analogen Resultaten für eine Betrachtung von allen N Teilchen des Systems anstatt nur eines Teilchens.

6.3. Ausblick

Neben den in dieser Arbeit betrachteten Entropieausdrücken, wurden und werden nach wie vor in vielen weiteren wissenschaftlichen Disziplinen Entropieausdrücke definiert und untersucht.

So wurde der Mathematiker Grigori Perelman 2006 unter anderem für seine Arbeit „The entropy formula for the Ricci flow and its geometric applications“ [Per08] mit der Fields-Medaille ausgezeichnet, die er jedoch nicht annahm.

Ein aktuelles Beispiel der Erforschung eines weiteren Entropiebegriffs in der Physik liefert die Forschungsgruppe der TU Wien rund um Daniel Grumiller, die unter anderem die Entropie schwarzer Löcher untersucht [GHM17].

Literaturverzeichnis

- [AB19] Raymond Addabbo und Denis Blackmore. „A Dynamical Systems-Based Hierarchy for Shannon, Metric and Topological Entropy“. In: *Entropy* 21. 938 (10 2019). URL: <https://www.mdpi.com/1099-4300/21/10/938>.
- [Bau68] Heinz Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Walter de Gruyter & Co · Berlin, 1968.
- [Bol09] Ludwig Boltzmann. „Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht“. In: *Wissenschaftliche Abhandlungen von Ludwig Boltzmann*. Hrsg. von Fritz Hasenöhr. Johann Ambrosius Barth Verlag, 1909, S. 164–223.
- [Bow71] Rufus Bowen. „Entropy for Group Endomorphisms and Homogeneous Spaces“. In: *Transactions of the American Mathematical Society* 153 (1971), S. 401–414.
- [BS02] Michael Brin und Garrett Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, 2002.
- [CK92] Chuan-Chong Chen und Khee-Meng Koh. *Principles and Techniques in Combinatorics*. World Scientific Publishing Co. Pte. Ltd., 1992.
- [CK81] Imre Csiszár und János Körner. *Information Theory. Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, Budapest, 1981.
- [ES14] Manfred Einsiedler und Klaus Schmidt. *Dynamische Systeme. Ergodentheorie und topologische Dynamik*. Springer Basel, 2014.
- [Els18] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer-Verlag, 2018.
- [EM81] James W. England und Nathaniel F.G. Martin. *Mathematical Theory of Entropy*. Addison-Wesley Publishing Company, 1981.

- [For13] Otto Forster. *Analysis 2. Differentialrechnung im \mathbb{R}^n , gewöhnliche Differentialgleichungen*. Springer Spektrum, 2013.
- [For16] Otto Forster. *Analysis 1. Differential- und Integralrechnung einer Veränderlichen*. Springer Spektrum, 2016.
- [For17] Otto Forster. *Analysis 3. Maß- und Integrationstheorie, Integralsätze im \mathbb{R}^n und Anwendungen*. Springer Spektrum, 2017.
- [Geo03] Hans-Otto Georgii. „Probabilistic Aspects of Entropy“. In: *Entropy*. Hrsg. von Andreas Greven, Gerhard Keller und Gerald Warnecke. Princeton University Press, 2003. Kap. 3.
- [Geo09] Hans-Otto Georgii. *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter Berlin · New York, 2009.
- [GHM17] Daniel Grumiller, Philip Hacker und Wout Merbis. *Soft hairy warped black hole entropy*. arxiv.org. 2017. URL: <https://arxiv.org/pdf/1711.07975.pdf>.
- [Hen08] Norbert Henze. *Stochastik für Einsteiger*. Vieweg+Teubner Verlag, 2008.
- [Iha93] Shunsuke Ihara. *Information Theory for Continuous Systems*. World Scientific Publishing Co. Pte. Ltd., 1993.
- [Kal15] Michael Kaltenbäck. *Fundament Analysis*. Heldermann Verlag, 2015.
- [KW08] Götz Kersting und Anton Wakolbinger. *Elementare Stochastik*. Birkhäuser Verlag AG, 2008.
- [Kle13] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer-Verlag Berlin Heidelberg, 2013.
- [Mül03] Ingo Müller. „Entropy: a Subtle Concept in Thermodynamics“. In: *Entropy*. Hrsg. von Andreas Greven, Gerhard Keller und Gerald Warnecke. Princeton University Press, 2003. Kap. 2.
- [Nat] National Institute of Standards and Technology. URL: <https://physics.nist.gov/cgi-bin/cuu/Value?k>. (zuletzt aufgerufen: 25.11.2019).
- [Neu17] Jörg Neunhäuserer. *Schöne Sätze der Mathematik. Ein Überblick mit kurzen Beweisen*. Springer Spektrum, 2017.

- [Per08] Grigori Perelman. *The entropy formula for the Ricci flow and its geometric applications*. arxiv.org. 2008. URL: <https://arxiv.org/pdf/math/0211159.pdf>.
- [Rai09] Peter Raith. „Chaos und Fraktale“. In: *Didaktikhefte der Österreichischen Mathematischen Gesellschaft* 41 (2009), S. 85–99.
- [Sch15] Stefan Schäffler. *Mathematik der Information. Theorie und Anwendungen der Shannon-Wiener Information*. Springer-Verlag Berlin Heidelberg, 2015.
- [SS09] Hermann Schichl und Roland Steinbauer. *Einführung in das mathematische Arbeiten*. Springer-Verlag Berlin Heidelberg, 2009.
- [Sha48] Claude Shannon. „A Mathematical Theory of Communication“. In: *The Bell System technical journal* 27 (1948), S. 379–423, 623–656.
- [Sin15] Yakov Grigorevich Sinai. „On the notion of entropy of a dynamical system“. In: *A. N. Kolmogorov’s and Y. G. Sinai’s papers introducing entropy of dynamical systems*. Hrsg. von Nikolai V. Ivanov. 2015, S. 12–14. URL: <https://nikolaivivanov.files.wordpress.com/2015/05/definitionentropy2014-20151.pdf>.
- [Sti10] Klaus Stierstadt. *Thermodynamik. Von der Mikrophysik zur Makrophysik*. Springer-Verlag Berlin Heidelberg, 2010.
- [Tap13] Stefan Tappe. *Einführung in die Wahrscheinlichkeitstheorie*. Springer-Verlag Berlin Heidelberg, 2013.
- [Vri14] Jan de Vries. *Topological Dynamical Systems. An Introduction to the Dynamics of Continuous Mappings*. de Gruyter Verlag, 2014.
- [Wer09] Dirk Werner. *Einführung in die höhere Analysis. Topologische Räume, Funktionentheorie, Gewöhnliche Differentialgleichungen, Maß- und Integrations-theorie, Funktionalanalysis*. Springer-Verlag Berlin Heidelberg, 2009.
- [You03] Lai-Sang Young. „Entropy in Dynamical Systems“. In: *Entropy*. Hrsg. von Andreas Greven, Gerhard Keller und Gerald Warnecke. Princeton University Press, 2003. Kap. 16.

Abbildungsverzeichnis

3.1. Darstellung von $[0, a)^3$ und einem der Würfel aus \mathcal{W} ; selbst mit GeoGebra Version 5 erstellt	30
3.2. Darstellung von h ; selbst mit GeoGebra Version 5 erstellt	32
4.1. Beispiel einer Verteilung von Teilchen; selbst mit GeoGebra Version 5 erstellt	37